EDUCATıONAL PARAMETERS REVEALED FROM VLE LOGGıNG DATA

A.H.W. (Piet) van der Zanden, W. (Wim) Veen
Delft University of Technology, The Netherlands
a.h.w.vanderzanden@tudelft.nl, w.veen@tbm.tudelft.nl

ABSTRACT

Educational management wants to comprehend the uses of ICT in Education to get a grip on its effects due to the multiple annual investments in the Virtual Learning Environment. In the search to define educational parameters a vast amount of datasets is examined from 289 institutes using Blackboard. The focus is on the three dimensions growth, diffusion and uses. For such huge amounts of data the pre-processing approach is essential to choose the right subsections of the original abound rough data. This paper describes the approach of the pre-processing, of the data collection activities, and of the ongoing data analysis.

Keywords: Educational parameters, growth, diffusion, uses, VLE, Logging data, Blackboard

Logging Data as Research Source

Virtual Learning Environments (VLE) made their entrance within higher education since the end of the 1990s. The underlying technology pushed the organisation to centralise the techniques and support, and after a certain period the VLE became a mission critical application (Zanden & Veen, 2004a, 2004b). At Delft University of Technology (DUT) the mission critical assignment was caused by the demand of the students, much more than the awareness of the educational management, and merely indicated as education critical by a handful of teachers (Zanden & Jonker, 2002). Still today the usage of the VLE within the institutions is not completely accepted. However, educational management wants to comprehend the uses of ICT in Education to get a grip on its effects (Moonen, 2003). For many years large sums of money were invested in the growing machineries, in the supporting personnel, in training materials, and in the educational staff. For such reason the management wants to know the own institution's efforts and progress (Deinum, 2003a, 2003b, 2003c; Groot-Kormelink & Bos, 2002; Huizer, 2002).

Data Acquisition

The research aims at revealing patterns and dependencies out of logging data from the VLE. A vast amount of datasets is examined from 293 institutes using Blackboard. The relevant collected data exceeds 24 gigabyte of textual data, which is derived from logging datasets with much greater sizes.
For such huge amounts of data the pre-processing approach is essential to choose the right subsections of the original abound rough data. In May 2003, a dataset from the DUT's Blackboard VLE was used as a first attempt to do research on the thus far unapproachable logging data. The aim of the test case was to determine if relevant data was present within the logging data and to structure the relevant logged items. The usage of the VLE was analysed and halfway 2004 an informal request was made to the DUT information manager to use the logging data for PhD research. In several months a formal request concerning privacy rules followed which led to official access in May 2005. The DUT legal officer demanded that all publications had to be anonymous with no possibility to back track the data.

A following in depth study learned that most of the data were independent system handles and user data. However, some relations could be made between system handling and users actions. The first pre-processing phase was to determine the wanted data. After determination special collection queries were designed to obtain interesting but anonymous data from other institutes' databases.

In the data collection phase also the Application Service Providing (ASP) unit of the Blackboard company was asked to be interested in the PhD research, which was followed by a formal request in March 2006. In October 2006 the confidentially agreement was signed and anonymously logging data of 289 institutes were made available. Next to the ASP data 9 Dutch universities made their data available for the PhD research, which analysis is conducted in this period and will continue in the coming months.

Data Observations

During the research period several upgrades of the Blackboard VLE were operational and a huge change in the database structure occurred after the upgrade from version 5.x to version 6.x of the Blackboard VLE. Where in the earlier Blackboard versions a main tracking table for approximately 150 different tracking areas was available (e.g. 'Announcements', 'Check Grade', 'Send Email', 'cp_send_email', etc.) (Buelens, Roosels, Wils, & Rentergem, 2002), in Blackboard versions 6.x and higher categories or separate databases were introduced such as BBADMIN, BB_BB60, BB_BB60_REPORT, BB_BB60_STATS, DBSNMP, SYSTEM. From these databases only the BB_BB60 and BB_BB60_STATS databases contain data related to educational uses of the VLE. However, the data in BB60 is made volatile since the amounts of data

grew too rapidly and exceeded the 1 GB per day. For such reason a sweep function was built in to regularly decrease the amount of saved data. Once every month the oldest available logged data is deleted. With the sweep function in mind the Blackboard company developed the so called Advance System Reporting (ASR) database, which is named BB60_STATS on the database server. A great advantage of the ASR is that the VLE administrator has access to the database without legal permission needed from the database administrator, who normally operate at different departments. In the ASR an extract of cumulated and associated data of the BB60 data is available. Where BB_BB60 contains data similar to the main tracking table from earlier versions, the BB_BB60_STATS or ASR holds tables with abstracted and accumulated data from BB_BB60. In Table 1 a brief description is given for the BB_BB60_STATS database, which contains 12 tables; the table USER_ROLES is not used.

Table 1: Specifications of the BB_BB60_STATS Tables

BB_BB60_STATS Table Names	Rows	Description
ACTIVITY_ ACCUMULATOR	13	Blackboard logged actions related to 402 session handles
APPLICATION	19	30 Tools and features that appear in Blackboard
COURSE_MAIN	41	Course labels and date of creation and modification
COURSE_ROLES	6	Six different roles of Blackboard course users
COURSE_USERS	26	Users and enrollment date
DATA_SOURCE	5	Source of system used for execution
INSTITUTION_ ROLES	9	Twenty different roles for the institute
NAVIGATION_ ITEM	18	Internal handles per application, tool or feature
SYSTEM_ROLES	4	Eight different roles for system maintenance
SYSTEM_ TRACKING	51	Collection table of Blackboard uses on daily basis
USERS	50	User information classified for privacy reasons
USER_ROLES	6	Not used, empty records

When a user is connected to the Blackboard VLE all activities during a session are logged as system handles with corresponding timestamps. The activities can be derived from ACTIVITY_ACCUMULATOR. INTERNAL_HANDLE where every handle corresponds to one out of 402 unique ID's in the BB_BB60_STATS.NAVIGATION_ITEM. Such an INTERNAL_HANDLE is the link to the undertaken activity and the called application listed in NAVIGATION_ITEM. From the 402 different handles of the NAVIGATION_ITEM table 155 handles are directly related to one of the 30 predefined Blackboard applications for uses by the student, instructor, or other users.

Data Analysis

Although the research is still running an overview is given concerning the vast amount of logged data.

Data Description

The logging data is collected from 293 institutes, from which 284 datasets are ASP and of which the owners are not known. The other 9 datasets are from Dutch universities. From the 284 anonymous institutes 214 has the complete datasets asked for, 66 datasets missed the activity logs and 4 datasets were damaged and could not be retrieved. Of the 293 institutes 27 are K12, 14 are Professional Education, 236 are Higher Education, and 12 are Corporations.

In total 1.279.166 course-IDs are declared of which 1.254.622 are courses, 24.258 are communities, and 286 are tests for administrators. The courses and communities are subdivided into 238 subclasses or professions such as biology, chemistry, aerospace engineering, calculus, etc. These 238 professions are assigned to 5 science classes, i.e. 18 professions to General Sciences, 47 to Arts & Humanities, 44 to Natural sciences, 28 to Engineering, and 101 to Social Sciences. The courses are subdivided respectively into 2.940 to General Sciences, 12.815 to Arts & Humanities, 22.674 to Natural sciences, 15.201 to Engineering, and 1.225.536 to Social Sciences. The enormous assignment of courses to Social Sciences is due to the large numbers of courses with the CourseClass "Higher Education" assigned to it. The CourseClass indicates the profession of what the corresponding course is about, but we assume that the higher education institutes took it as a default value for the higher education ownership instead of a course which treats higher education itself.

6.895.092 users have made 88.637.021 different sessions which hold 123.430.515 activities or handles. From the 30 asked applications only 23 were used. From the 20 present roles we will only focus on the roles which represent a substantial part

of the collection. This is because sometimes institutes assigned additional roles, which we will ignore for reliability reasons. In figure 1 the classification of the logged data is depicted.

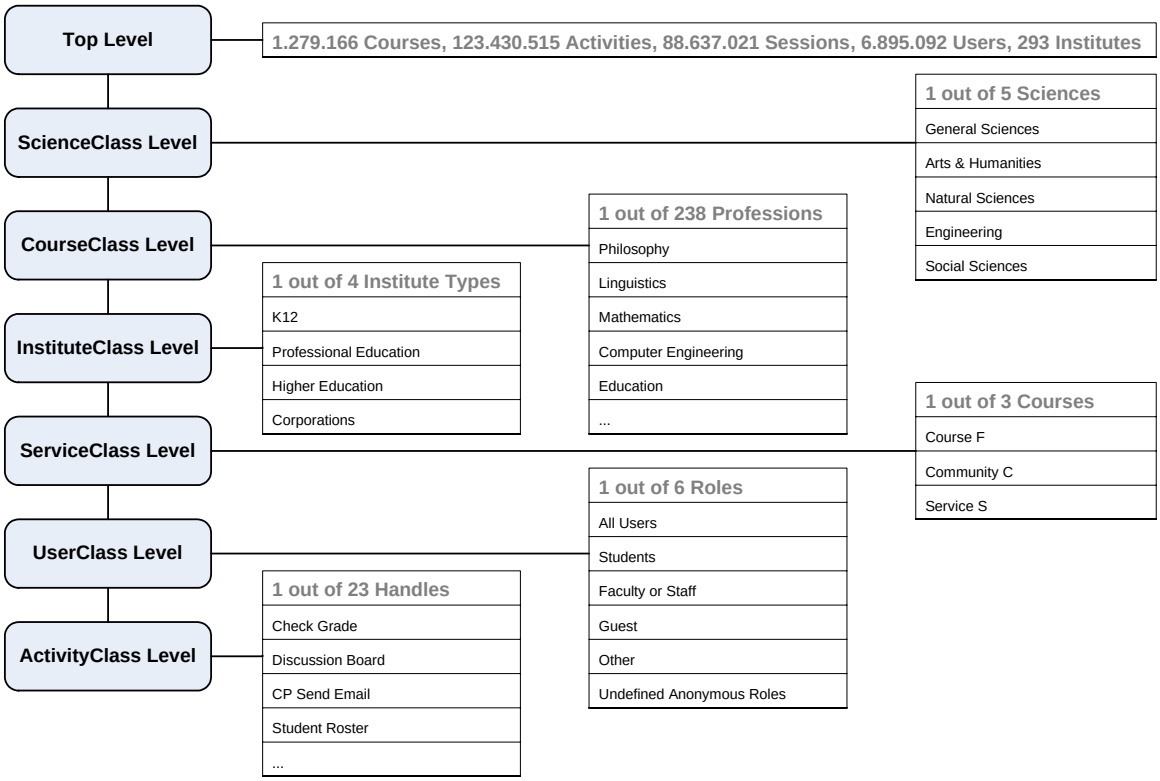


Figure 1: Classification Schema for the obtained Logging Data

Growth

Nolan and others has been carried out extensive research considering the IT and ICT uses in the corporate sector (Mutsaers, Zee, & Giertz, 1998; Nolan, 2000; Nolan & Croson, 1995; Nolan, Croson, & Seger, 1992; Nolan & Gibson, 1974; Zee & Koot, 1989). Every IT application used for the automation of business processes followed a certain path of growth; a natural S-shaped curve from scratch to maturity as depicted in Figure 2.

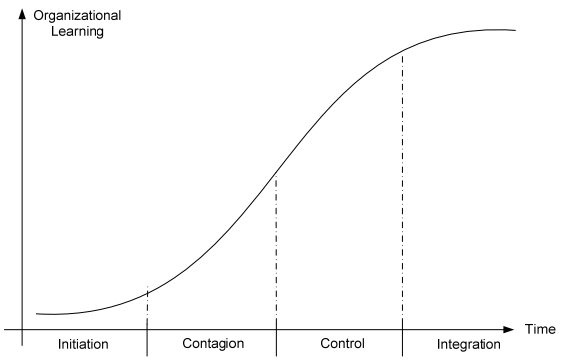


Figure 2: Four Stages of S-shaped curve according to R.L. Nolan

The first flat stage stands for a small increase of growth and is called initiation, which represents experiments, automation of simple and isolated tasks and limited investment. The second steep part, where growth is increasing rapidly, stands for contagion, for spreading of success, rapid expansion and little control. The third stage stands for gaining control on the expansion and corresponding costs as well as policy making. The growth in this third part is already slowing and finally the fourth stage, where growth is stopping, stands for integration or mainstream, for complete implementation and exploitation.

Nolan investigated S-shaped patterns based on the growth of applications, on the growth of personnel specialization, on the growth of the budget, and the related management techniques at each stage. We asked ourselves; will this growth pattern also count for educational applications and in particular for the VLE?

Out of our first findings it seems plausible that the VLE follows the same path. In figure 3 the growth of the number of users is presented over a period of 77 months, which follows more or less an S-shaped curve. However, further investigation is needed to corroborate with this hypothesis. We have to investigate the growth on the different class levels and when it stands it may become possible to predict the characteristics of the VLE. Such important insight may help the decision-making managers, educationalists, and teaching staff to act more decisively. Our aim is to declare parameters such as total numbers of courses and communities, non-courses, user types, life time periods, session periods, session activities, seat times, etc., which can be set out in time to discover growth or consolidation patterns.

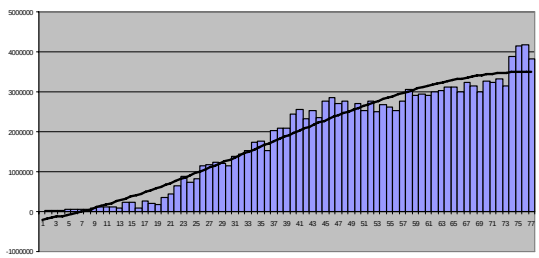


Figure 3: Indication of Growth of Users over 77 Months Time Period

Diffusion

Diffusion is (to cause something) to spread in many directions. Many investigations on the diffusion of innovations indicate that there are standard patterns to reveal when innovations take place, whether it concerns rural innovations or technological, organizational or educational innovations. Everett Rogers has been conducting research in the field of diffusion for more than four decades (Rogers, 2003).

The time element of the diffusion process allows classifying adopter categories and drawing diffusion curves. The adoption of an innovation follows a normal, bell-shaped curve when plotted over time. The adopter categories are innovators, early adopters, early majority, late majority, and laggards. When the cumulative number of adopters is plotted, the result is an S-shaped curve too, as is depicted in figure 4.

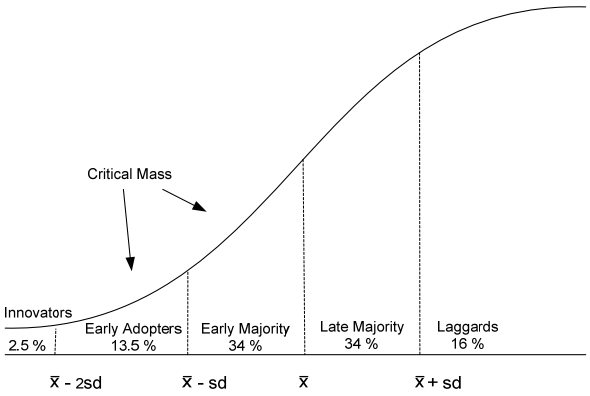


Figure 4: Adopter Categorization according to E.M. Rogers

Unfortunately it is difficult to thread the growing diffusion patterns, such as geographical spreading, because the logging data is collected anonymously. However, within a dataset an indication may be derived for educational spreading such as the increase of a certain profession (computer engineering, mathematics, languages, etc.) or a certain science class (Engineering, Social Sciences, etc.) of an institute.

On the other hand the ratio between the science classes' shares does not differ that much over the last period of 7,5 years, despite an average increase of more than 40 % new courses and communities every month. In figure 5 the relative partition of all courses and institutes in science classes is presented over a period of 7,5 years.

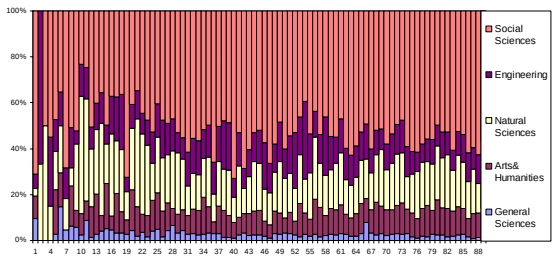


Figure 5: Relative Partition in Science Classes over 90 months Period

Again, the vast amount of data asks for more investigation to discern educational spreads within the institutes.

Uses

According to the Cambridge online dictionary uses mean to put something such as a tool or skill to a particular purpose. We want to explore which applications are used in courses and communities on the separate class levels, on what time and for how long. We also want to explore if changes occur over time, such as increases or decreases of certain applications or activities.

Although we do not know the content of the messages of the discussion board, because of the privacy reasons, it is obvious that communication is a significant part of the uses. Figure 6 indicates that the communicative actions are a steady part of more than 50 % average of the courses and communities activities for a period of 38 months.

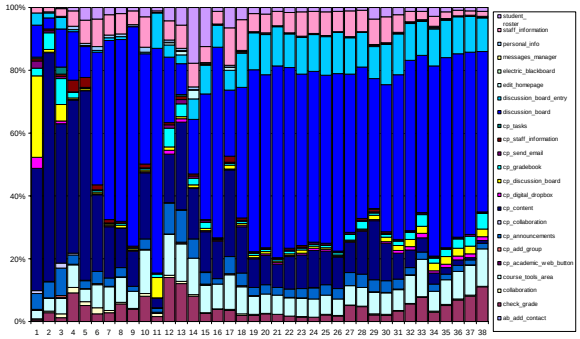


Figure 6: Relative Partition of Activities over 38 Months Period

This means that the VLE's are used according to their purpose. That is to say that it is the communicative part which was the great advancement of the VLE. In the beginning there were discussions that it looked like VLE's were used as logistic course systems for easy distribution of files and applications in stead of interactive learning.

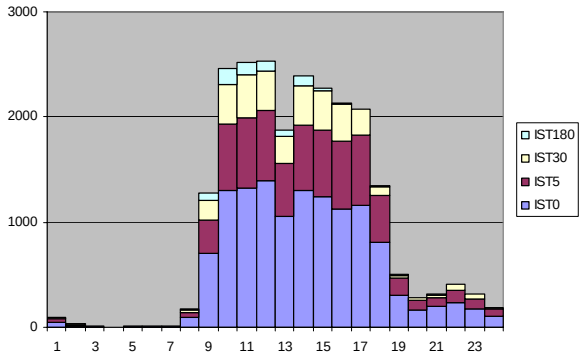
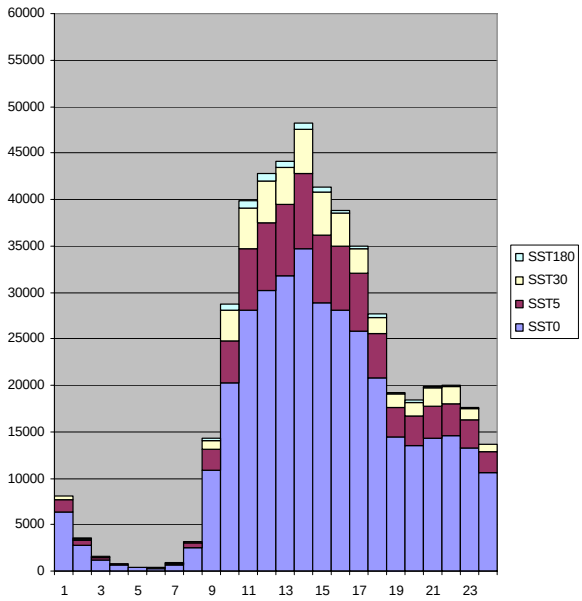


Figure 7: Activities of Instructors set out in a 24 hours day scheme

A first impression about the activities and its duration gives us the idea that the VLE is used within or during the lectures, because the activities of teachers diminish during the lunch hours, see figure 7. On the other hand the student activities increase during those same hours.



In figures 7 and 8 the activities are divided in short sessions with a time length from 0 to 5 minutes, from 5 to 30 minutes, from 30 to 180 minutes, and sessions with lengths longer than 180 minutes. It is not known yet which activities are done in the shorter and longer sessions.

In the Blackboard VLE courses are present with no students at all. A typical and probable reason for this is the first attempt

of a teacher or lecturer to explore the functionalities of the VLE. After a period of testing a true course was set up and the try-out course remained for further testing. Such courses and communities without any students are indicated as non-course or non-communities.

We assumed that any course could have only one or two instructors, but after a more detailed study it appeared that some courses were set up for professional communities, where every student or member is also instructor or community leader. They inventively "misused" the functionalities of the VLE to create a professional peer community for sharing and communicating.

DISCUSSION

Every year new students arrive at the university which causes growth of the student population. And every year students leave the universities when they graduate or fail or just leave for another reason. What influence do these figures have? Each year or each period in an academic year new courses are coming online. Are all of these really courses or just try-outs or updated courses or are they automatically generated for administrative reasons? Periodically reorganizations take place to rearrange faculties. It is possible that for such reasons course have multiple names?

How do we deal with these questions? What possible errors are introduced when corrected or not corrected for these possible flaws? More in depth research is needed to answer such questions.

CONCLUSION

It is still early to draw conclusions, but we argue that with the study of logging data the characteristics of a VLE can be determined. Usually qualitative research conducted with surveys and interviews is applied to measure satisfaction and worthiness of investments, but with the logging data based on traceable facts (with the help of Nolan's and Rogers's studies in mind) standard parameters for benchmarking efforts may be defined, which is an important advancement in assessing the value of the VLE.

Because of the vast amount of data we will continue our research with the help of data mining techniques for probably discovering new patterns in the logging data, for instance between the duration of sessions and the applied activities of that session.

REFERENCES

Buelens, H., Roosels, W., Wils, A., & Rentergem, L. v. (2002, 2002, Sep 2nd). *One Year e-Learning at the K.U. Leuven, an Examination of Log-Files.* Paper presented at the European Conference: The New Educational Benefits of ICT in Higher Education, Rotterdam.

Deinum, J. F. (2003a). *Evaluatie Brainbox I; Implementatie en Knelpunten [in Dutch]* (Evaluation report). Groningen: Rijksuniversiteit Groningen.

Deinum, J. F. (2003b). *Evaluatie Brainbox II; Ervaringen van Docenten en Leerlingen [in Dutch]* (Evaluation report). Groningen: Rijksuniversiteit Groningen.

Deinum, J. F. (2003c). *Evaluatie Brainbox III; Statistieken en Eindconclusies [in Dutch]* (Evaluation report). Groningen: Rijksuniversiteit Groningen.

Groot-Kormelink, J. B. J., & Bos, M. v. d. (2002). *ICTO evaluatie 2000 - 2002 [in Dutch]*. Delft: Delft University of Technology.

Huizer, C. G. (2002). *Evaluatie Blackboard, de basis voor het onderwijs support systeem [in Dutch]* (Evaluation report). Delft: Delft University of Technology.

Moonen, J. (2003). Simplified Return-On-Investment; a new approach. *Interactive Learning Environment, 11*(2), 147 - 165.

Mutsaers, E.-J., Zee, H. T. M. v. d., & Giertz, H. (1998). The evolution of information technology. *Information Management & Computer Security, 6*(3), 115 - 126.

Nolan, R. L. (2000). Information Technology Management since 1960. In A. D. Chandler & J. W. Cortoda (Eds.), *A Nation Transformed by Information; How Information has shaped the United States from Colonial Times to the present* (pp. 217 - 256). Oxford: Oxford University Press.

Nolan, R. L., & Croson, D. C. (1995). *Creative Destruction: A Six-Stage Process for Transforming the Organization*. Boston, Massachusetts: Harvard Business School Press.

Nolan, R. L., Croson, D. C., & Seger, K. N. (1992). The Stages Theory: A Framework for IT Adoption and Organizational

Learning. In *America's Information Technology Agenda*. Cambridge: John F. Kennedy School of Government.

Nolan, R. L., & Gibson, C. F. (1974). Managing the Four Stages of EDP Growth. *Harvard Business School, 52 (1)*, 76 - 88.

Rogers, E. M. (2003). *Diffusion of Innovation* (5th ed.). New York: Free Press.

Zanden, A. H. W. v. d., & Jonker, S. A. (2002). *Evaluatie ICTO Programma TU Delft [in Dutch]* (Evaluation report). Delft: Delft University of Technology.

Zanden, A. H. W. v. d., & Veen, W. (2004a, Apr 3 - 5). *Hypothetical Model for Change and Progress of ICT in Education.* Paper presented at the 13th International Conference on Management of Technology, Washington, USA.

Zanden, A. H. W. v. d., & Veen, W. (2004b, Jun 21 - 26). *Innovation in Higher Education; Demand-driven or Market-pushed.* Paper presented at the EDMEDIA 2004, Lugano, Switzerland.

Zee, H. T. M. v. d., & Koot, W. J. D. (1989). I/T-assessment; een kwalitatieve en kwantitatieve evaluatie van de informatieverzorging vanuit een strategisch perspectief [in Dutch]. *Informatie, 31 (11)*, 837 - 851.