The evaluation of external data resources for Business Intelligence applications: the example of the Czech Republic

Vladimir Bures, Tereza Otcenaskova, Veronika Jasikova
Department of Information Technologies, University of Hradec Kralove, Czech Republic
vladimir.bures@uhk.cz
tereza.otcenaskova@uhk.cz
veronika.jasikova@uhk.cz

Abstract: *Contemporary business requires relevant data for appropriate, effective and reliable decisions. The concept of Business Intelligence which has been successfully established in a lot of organisations necessitates gathering relevant internal as well as external data which are often hardly acquirable. The aim of this paper is to analyse both the quality and availability of the sample of the external data resources on the basis of selected criteria and to enable the evaluation of their potential. The research reveals that the resources are currently not as suitable and utilisable for the purposes of Business Intelligence as they are expected to be. Although there are exceptions such as data provided by particular governmental or national institutions, the majority of external data resources does not correspond to the requirements posed to their availability, credibility, update regularity, stability and technical accessibility.*

Key words: evaluation analysis; Business Intelligence; Czech Republic; data resource availability; data resource credibility; data resource stability; decision making process.

1. Introduction

Decision-making is a complex process which requires support by any available tool. Modern Information and Communication Technologies (ICT) provide managers with a set of methods, techniques, tools, principles and procedures, which have already been successfully applied in various kinds of business oriented issues ranging from business process management [17] to knowledge management implementation [2]. Moreover, the process of mutual integration of advanced ICT and business strategy has been observed for several years. Nevertheless, one specific approach is of high importance nowadays. Due to huge amount of data produced every day by particular organisations, the concept of Business Intelligence (BI) has been established. Sums, percentages, grouping or trends are one of the most common data analysis keywords [1]. However, these need to be derived from available data with the help of particular methods or tools. Therefore, the concept has successfully spread out and has become familiar in the business domain. While in 2003 searching the string 'Business Intelligence' on the web gave you 911.000 hits [1], contemporary, eight years later, the Google search engine provides you with approximately 29.400.000 hits.

The core objectives of this paper are to emphasise that there is a certain segment of the Business Intelligence concept that is not treated appropriately, and to conduct in-depth analysis of the current state of external data resources. The main purpose of the research is to stimulate the discussion and highlight that while external data are crucial for proper decision-making in modern business administration, their resources and integration with internal data resources remain mostly underestimated by ICT providers.

To outline the general structure of the paper, the first part deals with the general concept of BI. Then the research problem is formulated. The next section describes the utilised methodology together with the rationalization of applied methods. Afterwards, the research results are discussed. Finally, the implications and limitations of the research are mentioned.

2. Business Intelligence review

There is no broadly accepted definition of the Business Intelligence concept. Definitions vary from very general ones, such as the definition provided by Scheps, who states that BI is 'any activity, tool, or process used to obtain the best information to support the process of making decisions' ([16], p.11). Nevertheless, there are also definitions dealing with specific technologies and principles. For instance,

Vercelis argues that BI is a 'set of mathematical models and analysis methodologies that systematically exploit the available data to retrieve information and knowledge useful in supporting complex decision-making processes' ([19], p.xiv). Apparently, the necessity to work appropriately with data resources is recognizable even from existing Business Intelligence definitions. Several authors also emphasise, although implicitly, that there are several data resources available for Business Intelligence applications. For example, Sabherwal and Becerra-Fernandez [15] specify BI with the help of further technologies, since they differentiate among Business Intelligence, Knowledge Management, Data Warehousing, Data Mining and Decision Support Systems (see Table 1).

Tab. 1: Distinction between Business Intelligence and other Related Technologies

	Business Intelligence	**Knowledge Management**	**Data Warehousing**	**Data Mining**	**Decision Support Systems**
Inputs	Data, information	Data, information, knowledge	Data (from multiple systems)	Data	Data, information, knowledge
Nature of Inputs	Internal or external, structured or unstructured	Internal or external, structured or unstructured	Internal, structured	Internal, structured	Internal or external, structured
Outputs	Information and explicit knowledge	Tacit knowledge and explicit knowledge	Data (in a single logical repository)	Information	Decision recommendation
Components	Information technologies	Information technologies, social mechanisms, structural arrangements	Information technologies	Information technologies	Information technologies
Users	Across the organisation	Across the organisation	IT personnel	IT personnel, others trained in IT	Specific targeted users

Source: [15].

Probably the most comprehensive definition used as a starting point in this research is provided by Novotný, Pour and Slánský, who define BI as a complex of ICT approaches and applications, which almost exclusively support only the analytical and planning activities in an enterprise and which are based on the principle of multidimensionality that represents the ability to investigate reality from several perspectives [12].

Apparently, the BI cannot be perceived as one specific technology or tool. Revealing its content, several technologies which can be hierarchically decomposed in the pyramid described in Figure 1 should be considered.

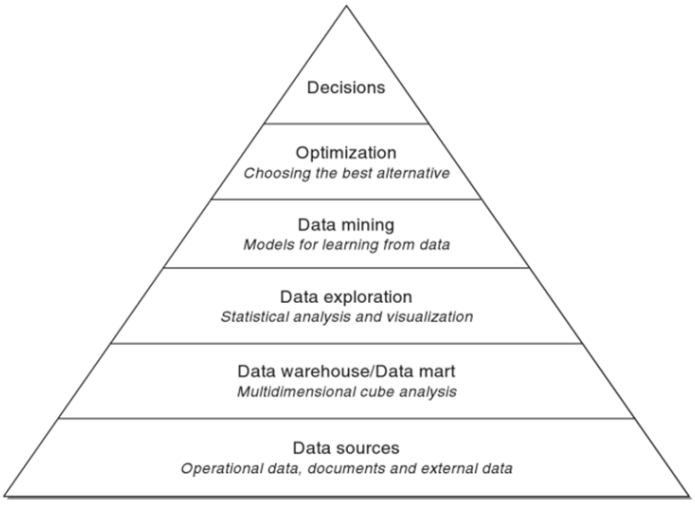

Fig. 1: Building Blocks of Business Intelligence Systems *(source [19])*.

While the pyramid reveals only the purpose of the technologies, further specifications remain hidden. Nevertheless, it stresses the fact that data resources represent the basis of all consequent activities and endeavours. Therefore, the scheme described in [19] can be used (see Figure 2). This Business Intelligence framework is in concordance with the hierarchical decomposition depicted in the pyramid. Several levels are identified. Moreover, this framework goes beyond enterprise boundaries and emphasises the openness of the concept. Although this framework is inferred from the ICT-based definition (see above), it reflects the human side of the discussed issue. Therefore, the environment is represented not only by resource systems (for example Enterprise Resource Planning - ERP, or Customer Relationship Management - CRM), but also by the end users from different departments of an organisation. Generally, the first level of the framework is represented by Data Transformation Components.

Huge amount of coeval data is captured in the transaction systems, in the Figure 2 denoted as Resource systems. However, these data do not have a form suitable for further analysis. For decision making, today's business user asks for the continually up-to-dated information. Therefore, the Data transformation components level is of critical importance. This level is responsible for processing data in their original form and converting them into unified structures providing the possibility of fast information retrieval. One of the common attitudes of working with data at this level is ETL, comprising three step operation Extract-Transform-Load which transforms the input data from their crude form into format needed by the end users. This refreshment is usually executed periodically. The challenge of shortening the data updating period is discussed in [9]. Authors also describe the usual occurrence of anomalies in the data consistency and offer methods of avoiding them. The set of topics connected to the new system integration into the running business is covered by the enterprise application integration (EAI) including proven manners used within this process. One of the key issues in this area is security of the whole process depicted in [20], where authors focused their research mainly on the EAI security problematic of service-oriented architecture (SOA) systems.

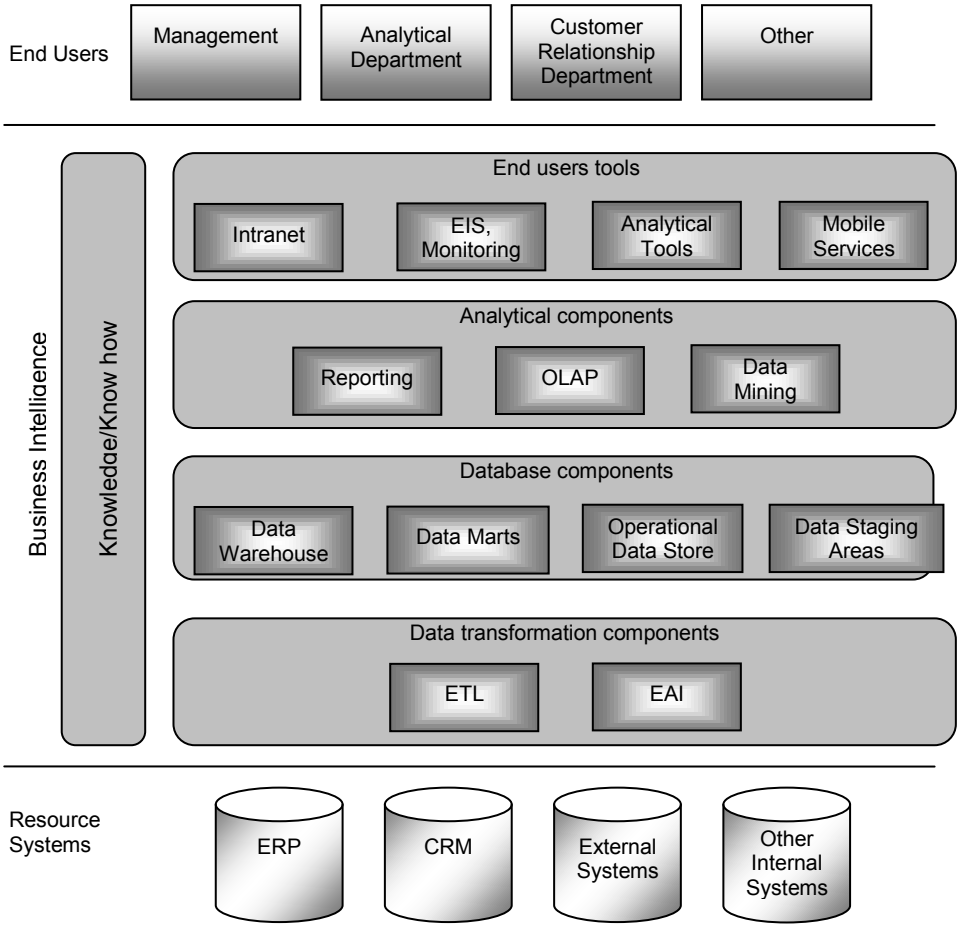

Fig. 2: Business Intelligence Components *(Source: [12])*.

The next level comprises particular data storage elements. Data in their new format are stored in data warehouses, marts and other special data staging areas. One of the most common practices recently

used is the operational data store (ODS). ODS is not a substitute to any of classical data warehouse system and the main advantage of this approach is processing of complex data operations [11]. In general, we can call these systems database components. In comparison with the original resource systems this level consists of the structures focused on fast information gaining instead of the efficient data storing. It leads to the need of data redundancies, duplicities and aggregated values warehousing. Classification and marking of the most suitable data storing system is the topic of many software measurement methods and analyses which are discussed in detail in [5]. The included case study is focused on specific benchmarking of the most used Business Intelligence platforms.

Analytical components level is characterized by the multidimensional data representations built on warehouses mentioned above. This module usually consists of logical multidimensional cubes physically stored in special On-Line Analytical Processing (OLAP) databases or in the tables of relational databases arranged into so-called star or snowflake schema. This data architecture is mainly oriented on the reporting purposes providing wide scale of aggregating and drilling possibilities using unified query languages. This analytical level is also designed for data mining operations searching for hidden dependencies and revealing latent knowledge. Both reporting and data mining represent the added value to the organisation and therefore fulfil the Business Intelligence software main purpose. Moreover, these serve as an information and knowledge source for further planning and decision making. For example the decision tree method is becoming more and more popular these days in the realm of information classification and prediction making. As stated by Rokach and Maimon [14], there are certain advantages of this approach such as algorithms simple implementation, non-parametric nature, robustness and the ability to process both quantitative and qualitative variables. Decision trees can be easily converted to the data classification rules and therefore they provide the wide range of possibilities of their representation such as the expression in the common language. Another method leading to revealing new connections by item classification is the cluster analysis described by Everitt et al. [4]. The gist of the method is in dividing data items into classes based on their properties. Consequently, the members of each class show the similarity in maximum of observed attributes. Further techniques such as genetic algorithms can be used at this level [18]. Apparently, in case of BI several knowledge-based approaches and technologies are of crucial importance. Therefore, the traditional ICT-based perception should be extended to Information, Communication and Knowledge Technologies (ICKT).

Any visualization, reporting and monitoring tools can be used as a superstructure based at the last mentioned level. This level enables the end user data visualization oriented on the result simplicity and lucidity, usually achieved by using charts, graphs and ratios. Visualized information also differs on each user's intention and only the relevant data are shown. Intranet or organisational web portals are tools usually employed at this level. Ability to visualize required data can be used as additional criterion of their usability evaluation [8]. Use of the decision support systems is another domain where data in this form are required as an input for further processing. One of recently discussed topics is using the cognitive hierarchy process in on-line decision making originally presented by Gavalec and Mls [6]. It is one way how to optimize the on-line decision making process optimization using the fuzzy cognitive maps. In SOA of the information system, the data visualization layer is represented by mobile devices such as smart phones or tablets. The importance of this approach is increasing nowadays because more and more organisations are assimilating the mobile way of leading and managing business processes. Kozel and Mohelská [10] are dealing with the efficiency of using of the mobile technologies in this area called e-business or e-commerce.

3. Problem formulation

The field of BI is apparently well developed and particular technologies are applied, scrutinized and further elaborated, not only in economically developed countries, but also in emerging economies such as the economy of Vietnam [13]. However, the lack of attention has been paid to the external resources that serve as a vital connection between the Business Intelligence framework and its environment. Although commonly used during the development of particular Business Intelligence applications, their overall attributes and structures are generally unknown. Moreover, the evaluation of their appropriateness for the Business Intelligence applications is neither described in scientific resources nor in resources accessible via data providers. The external data are of high importance for successfulness of the Business Intelligence application in practice. Therefore, the main objective of the conducted research was the comprehensive description of available data resources and their overall assessment. The primary research question was focused on the identification of the level of utilisability, which the existing external data resources used for Business Intelligence applications achieve. Moreover, the intention to find out under which conditions these can be used during the

Business Intelligence development was pursued. The focus of the research was mostly aimed to the Czech business environment. However, the used methodology can be successfully applied in any developed country.

4. Research method

Based on the internet search on data providers' web sites, literature reviews, recommendations given by experts from relevant areas and telephonic unstructured interviews with representatives of data providers, the sample of 93 available data resources was collected. Considering the amount of data and information currently provided and daily amended not all data resources can be examined within such research. Therefore, the aim was to gather data from the most relevant providers and to cover the main domains for the purposes of potential end users from various backgrounds. The elementary content analysis revealed that resources can be clustered into three main categories reflecting the nature of provided data. These categories are:

- macroeconomic indicators,
- legal (business) entities,
- markets, products and commodities.

Among macroeconomic indicators, the information from the national statistical office and the national bank was gathered and evaluated. Moreover, the data from commercial banks were also considered. Legal (business) entities were represented by institutions providing information about particular organisations or business sectors such as profits, share prices, or number of employees. Markets, products and commodities included mostly the stock exchanges, ministries, professional organisations and associations, and also commercial providers and retailers. The research was targeted on the particular national environment, international institutions and other data providers were included only to a limited extent. On the other hand they cannot be completely excluded, because they both significantly influence the development of prices and markets, and offer useful data and information. Although, considering these limitations, the general findings are comparable to a certain extent.

Consequently, the level of utilisability of data resources was investigated using the selected criteria. Generally, the data resources can be evaluated according to their characteristics and various indicators. Among others the following might be considered: relevancy, redundancy, unambiguity, data provider reputation, objectivity, price, technique of data collection/retrieval, reliability, integrity, consistency, accessibility, accuracy, transparency, permission for further data utilisation, comprehensibility, or up-to-datedness. Nevertheless, the final criteria for this research were strictly pre-defined by the research applicant (see the acknowledgment) considering the practical aspects and needs. Particularly, these represent a combination of some of the aforementioned indicators. The utilised criteria are:

- Availability
- Credibility
- Update Regularity
- Stability
- Technical Accessibility.

Regarding the necessity of the unification of particular data resources assessment, it is crucial to unambiguously determine both the aforementioned criteria and the scale used for their evaluation. Table 2 illustrates the simple three-point scale and the determination of particular grades. While the assessment by grade 1 represents the best alternative, the grade 3 is the worst one.

Tab. 1: Assessment Criteria of Data Resources and their Determination

	Availability	Credibility	Stability	Update Regularity	Technical Accessibility
1	Free of charge, RSS	Primary data, known methodology	Public institutions, data retention and release is pursuant to the law	Topical data (schedule of their release etc. included)	CSV, XML, XLS
2	Registration required	Secondary data, known methodology	Large enterprises with long-term operation in the market	Release with delay	HTML, DOC, TXT, RTF
3	Charge required	Not known methodology	Other companies	Obsolete data	PDF, printed paper

Source: Authors' Research

The criterion ***availability*** characterizes where and under which conditions the resource is available. Financial costs linked with the data availability are considered as well, because these influence the utilisation of data resource especially from the perspective of the end users. The best evaluation is reached by resources available at no cost and possibly enabling the information provision through the RSS feed. Data acquirable with difficulties or after registration are graded by 2. The least favourable alternative comprises the resources which charge users for the data provision. The issues linked with the availability evaluation are quite difficult and ambiguous. The particular institution could have been assessed by grade 2 only because it provides more sophisticated data available solely after registration or after charging a fee. Nevertheless, other data (comparable with the offer of other institutions) could have been without charge. This fact should be taken into consideration while interpreting the outcomes.

Another criterion is ***credibility*** which plays its important role regarding especially the data utilisation. For the purposes of this research, the credibility is described as a subjective assessment of quality of provided data. The best resources include primary data with exactly defined methodology of their gathering and proceeding. It obviously facilitates better and easier data interpretation as well as enhances the relevancy of their utilisation. Lower grade on the scale represents secondary data. Nevertheless, this category is also required to embrace the source of provided or used data together with verified methodology. The resources where no source and data gathering methodology are included remain in the worst assessed category.

The ***stability*** describes the perspective of resource availability and operation in the future. The best assessment is assigned to public institutions. These are obliged to retain and release data regularly according to law. Moreover, this category includes stable, persistent and world widely influential institutions. The second less stable group comprises large enterprises operating in the market in the long term. Banks and various associations serve as an example. The third category covers predominantly private market-oriented subjects which have less probability to operate in the market from the long-term perspective.

Another criterion considers the ***update regularity***. The purposes of data update vary within particular analysed resource categories. Generally, the best assessed resources ensure both the up-to-date data and the information about the schedule of their release. The subject who requires and utilises the data can apparently on the basis of such reliable and topical data better plan its organisational processes or predict the future development. The second category embraces the resources publishing and releasing data with delay. The group assessed by grade 3 includes the obsolete data released with significant delay which could have negative consequence for the end user. Therefore, such data are not suitable for practical utilisation.

The last considered criterion is the ***technical accessibility*** describing the format of provided data and the technical options of their acquisition. From the perspective of the end users who utilise the data, the most appropriate resources enable the data automation. It is provided especially by csv, xml or xls formats. Editable, but unstructured formats represent a less user friendly category which includes for example the following formats: html, doc, txt or rtf. The last group comprises data that cannot be edited in formats such as pdf or data printed on paper.

5. Results and discussion

The conducted research dealt with three main tests focused on the achievement of the given research objectives. The first test was focused on evaluation of suitability of particular data resources. The second test was concentrated on the frequency of the assessment of particular criteria in each resource category. The last test was used to verify if the overall set of available data resources can be successfully utilised in prospective Business Intelligence applications.

The parameters were set considering the purposes of Business Intelligence utilisation for the ***evaluation of particular resources***. The description of two conditions for the selection of the most appropriate resources follows:

- the evaluation of particular resource criteria cannot include any grade 3
- the average value of all criteria cannot exceed the value 1.5.

Such filtering might be also differently adjusted according to the actual end user and his or her requirements. The analysis was executed utilising a common spreadsheet application Microsoft Excel. The analysis ensures the user-friendly and common environment for the parameters settings and is therefore widely applicable for different purposes of various end users.

The results from this selection reveal that within the legal entities, only 25 % passed in the test with aforementioned settings. Concerning the markets, products and commodities, only 28 % of the discussed resources met the demands. Finally, the macroeconomic indicators provide the most relevant resources. In this case, the number of recommended resources represents 71 %. These results support the potential assumption that the macroeconomic indicators which comprise quite a lot of public sector organisations are more reliable and more accessible in terms of costs. Moreover, these are usually linked with the legal issues such as various regulations about the frequency of data publishing and their availability. Generally, these outcomes prove the best preparedness and potential utilisation of the macroeconomic indicators from the institutional and legal environment.

Consequently***, the frequency of the assessment*** of particular criteria in each resource category was examined. The Table 3 illustrates the summarised results focused on the identification of maximum and minimum occurrences within the particular criterion. Apparently, the macroeconomic indicators proved again the best results considering the availability. Such finding is not so surprising considering the fact that these resources comprise institutions from public or banking sector. On the other hand, the technical accessibility remains problematic within this category. Two remaining resource categories demonstrate the highest frequency in up-to-datedness. These categories apparently necessitate regular updates because of the demands of both the institutions themselves and their customers. In particular cases (for example in cases of institutions providing prices of commodities and products) the data change practically continually and require regular updates in order to retain their relevancy and up-to-datedness.

Regarding the legal entities, the main problems are with availability and stability. The latter criterion is also significant in reference to markets, products and commodities. The interpretation of identified problems might be linked with the type of these institutions. A lot of these resources belong among small- or medium-sized private enterprises and therefore, the data provision is usually charged. These institutions are also not as stable as for example public sector ones because of the high competition, market or legal demands, and the like.

Tab. 2: Frequency of Assessment of Particular Criteria (maximum values are highlighted in critical categories)

Resource Category	Availability	Credibility	Stability	Update Regularity	Technical Accessibility
Macroeconomic Indicators					
1	**29**	26	9	27	7
2	2	5	22	3	20
3	0	0	0	1	**4**
Legal Entities					
1	4	1	3	**9**	2
2	0	4	1	3	10
3	**8**	7	**8**	0	0
Markets, Products and Commodities					
1	23	22	25	**32**	3
2	18	26	10	17	45
3	9	2	**15**	1	2

Source: Authors' Research

Considering the fact that the gathered data are ordinal in nature, the ***test of relative frequency*** was conducted. The relative frequency of a monitored attribute (in this case the occurrence of the best assessment of each criterion) is tested because such statistic offers helpful outcomes and worthwhile comparison.

Firstly, the null hypothesis was stated successively for each criterion as follows: 'The absolute majority (more than 50 %) of the resources does not prove sufficient attributes, in terms of availability, credibility, stability, update regularity, or technical accessibility, for Business Intelligence applications.' The last tested feature called 'Total' represents the null hypothesis stating that: 'The absolute majority of the resources (more than 50 %) do not demonstrate good results from the perspective of Business Intelligence considering all criteria.' The right-sided alternative hypothesis is considered, because it

complies the purposes of the intended test. Generally, the null and alternative hypotheses can be explicitly expressed:

$H0: p \leq \pi 0$

$H1: p > \pi 0$ (1)

The hypotheses as well as following formulas include a variable p which is a real relative frequency calculated as

$p = m/n$ (2)

The relative frequency represents the ratio of the number of times of event occurrence (m) and the total number of values in the sample (n). π0 is then the anticipated value of relative frequency. In this problem, the π0 was fixed at value of 0.5 (representing the absolute majority).

Moreover, the considered statistics has approximately the normal distribution, because in case of all criteria it fulfils the necessary condition for such approximation [7]. The formula for the condition follows:

$$n > \frac{9}{\pi(1-\pi)} \quad (3)$$

The significance level is determined at 5 % (α = 0.05). The rounded tabulated critical value, possibly the quantile of standardized normal distribution, is therefore u1-α = 1.645 (section Tables in [7]).

The value of test-statistic (U) is calculated according to the following formula [7]:

$$U = \frac{p-\pi_0}{\sqrt{\pi_0(1-\pi_0)}}\sqrt{n} \quad (4)$$

Firstly, the values of relative frequencies for the best assessment (grade 1) of each criterion were calculated. Afterwards, the corresponding test-statistics were determined using the abovementioned formula. All calculated values are in Table 4.

Tab. 4: Statistical Testing and its Crucial Determinants: Absolute Frequency (m), Relative Frequency (p), Test-statistic (U) and the Test Result (H0 acceptance or rejection)

Statistical Variable	Availability	Credibility	Stability	Update Regularity	Technical Accessibility	Total
M	56	49	37	68	12	222
P	0,60215053	0,52688172	0,39784946	0,731182796	0,129032258	0,47741935
U	1,97020822	0,518475847	-1,97020822	4,458892287	-7,154966694	-0,97385168
H_0	**rejected**	**accepted**	**accepted**	**rejected**	**accepted**	**accepted**

Source: Authors' Research.

The null hypothesis is accepted when the value of test-statistic is lower or equal than the critical value ($U \leq u1-\alpha$) [7].

Apparently, the research revealed significant differences among particular criteria. In case of availability, with respect to the value of test-statistic, H0 is rejected at the 5 % significance level. This proves that the data are relatively available for the purposes of BI. Considering the credibility, H0 is accepted at 5 % significance level. This signifies the fact that the resources include not sufficiently reliable institutions. The stability indicates the same result as the credibility which implies that the resources are not appropriately stable and both the future data availability and provider operation are not certain. On the other hand, the sample of resources demonstrates high assessment referring the data update. In this case H0 is rejected at the 5 % significance level. Such results confirm the current necessity and demands on the up-to-date data. The technical accessibility demonstrates the opposite situation when the H0 is rejected at the 5 % significance level. This shows the inappropriate preparedness for data automation and inflexibility of provided formats. Finally, the total frequency of the best criteria assessment is examined. The research concludes that H0 is accepted at the 5 % significance level in this case which proves the overall inadequate quality of discussed data resources.

The findings prove the necessity to improve the quality of data resources or at least to consider their limitations while using them. Basically, there are two reasons for the current state. Firstly, data providers are not forced to follow any albeit general directions or instructions. Secondly, those who use the data resources for their purposes, e.g. software engineers, do not make unified demands on these providers. Therefore, the certain level of standardization of both the requirements of software organisations on, and the assessment criteria of data resources would be worthy. This can be realised through the employment of various attitudes - one of the options can include the state regulation or the

involvement of international standard institutions to ensure the standard determination and their control.

6. Implications and limitations of the study

The findings confirm that the external data resources are not generally as suitable and utilisable for the purposes of BI as they should or could be. Therefore, the research substantiates the necessity to allocate attention to improvements of data from all the discussed perspectives. Nevertheless, the outcomes cannot be generalized, because each resource evaluation differs. The fact that the results assume a certain level of generalization should be considered while interpreting the research findings. Nevertheless, the systematic and systemic solution of the existing situation could contribute to following main domains:

- informatics - coping with the vast amount of data resources consisting of different data formats and distributed in various institutions can represent a challenge for research in the area of systems integration. Commonly known types of integration such data, application, or methodological ones, will have to be used in innovative ways to handle the described problem from the informatics point of view. However, the process of overall integration is not merely the technological problem. Further types of integration, such as strategic or political ones, necessitate being included and thus the current development of the information society characterized by the intense mutual integration of ICT with other activities will be reflected. Moreover, the organisations dealing with software engineering play an important role for these represent the most often end user of data resources. From their perspective, it is crucial to know which data resources are available as well as reliable together with the information about their providers.
- business administration - the beneficial use of Business Intelligence system requires its certain complexity and performance. Therefore, it is used mostly by large companies which can afford these tools and which are able to take the advantage of them. Business Intelligence solutions for small- and medium-sized companies, as defined by Eurostat statistics, are still not very common. Public data resources improvement could be the first step towards solving the situation. This precaution is also tightly connected to the area of public administration - in this way attempts to improve the information society at the national level such as the e-government development of establishment of the Smart Administration strategy can be supported. It would be appropriate to integrate the system of data resources into the concept of public administration digitalization and include services providing this information. This step would improve the access of the small- and medium-sized companies to this kind of data and hence introduce Business Intelligence principles into this type of businesses. This corresponds with Cao, Luo, and Yhang [3] who emphasise the necessity to balance technical and business interests of organisations.

It is necessary to mention that there are particular limitations of the discussed research which should be considered while interpreting the results. Firstly, the problem with the criteria evaluation needs to be stressed - the determination of categories of data quality (marks 1, 2 and 3) remains vague and therefore a certain generalization and simplification is necessary. Nevertheless, within the conducted research, the criteria were clearly defined and applied identically to all the data resources assessment to ensure relevant, objective and comparable outcomes. Secondly, the criteria were intentionally selected - as mentioned above, there are a lot of factors which influence the data resource quality. Generally, it is difficult to find an appropriate set of criteria suitable within any environment. Each end user has different requirements and priorities which data resource characteristics are important and relevant to his/her own context and purposes. Therefore, a research focused on the determination of general criteria and their importance remains a challenge for future. Both the standardization and the incorporation of software engineers' requirements would support the trustworthiness as well as enhance the usability of the methodology. Such methodology for data resources assessment would be then utilisable within various areas (for example for the purposes of private and public sector organisations).

7. Conclusions

The conducted research reveals that resources, which provide users with macroeconomic indicators, prove the best results considering the availability. This finding can be explained by the fact that these resources comprise institutions from governmental, public or banking sector. Nevertheless, the technical accessibility remains problematic within this category. Two remaining resource categories,

i.e. legal (business) entities and markets, products and commodities, demonstrate the best results in the area of up-to-datedness. The null hypothesis "The absolute majority (more than 50 %) of the resources does not prove sufficient attributes, in terms of availability, credibility, stability, update regularity, or technical accessibility, for Business Intelligence applications" is confirmed at the 5 % significance level. However, null hypothesis for two particular criteria is rejected. Availability and Update regularity prove to be at an appropriate quality level. Nevertheless, although several exceptions can be identified, for instance data provided by the Czech Statistical Office or the Czech National Bank, the majority of external input data resources are not mostly reliable, credible or stable enough to be effectively used for the Business Intelligence purposes.

While it is obvious from the second section of this paper that internal data resources are treated appropriately and several tools are already applied, the quality and accessibility of external data is expected to be ensured by their providers. Therefore, consequent decisions can be based on obsolete or unreliable data and hence cause wrong or even harmful consequences. The potential for further improvements is therefore obvious and these have to be discussed and implemented. Alongside the analysis of the current state of the external data resources, the aim of this paper is to initiate the discussion about the problem. While the analysis is successfully conducted and presented, the discussion and implementation of changes is the burden shifted to the professional community.

Acknowledgements

This paper was written with the support of the project Innovation Voucher provided by the Innovation Fund of the Hradec Králové Region and conducted for company GIST, s.r.o.

References

[1] BIERE, M., 2003. *Business Intelligence for the Enterprise.* 2nd ed., IBM Press, New Jersey, United States

[2] BUREŠ, V., 2006. Knowledge management and its implementation. In *Proceedings of the 2nd International Conference on Web Information Systems and Technologies* (WEBIST 2006). INSTICC Press, Setubal, Portugal, pp. 115-118

[3] CAO, L., LUO, D. and YHANG, C., 2007. *Knowledge* Actionability: Satisfying Technical and Business Interestingness. *International Journal of Business Intelligence and Data Mining,* Vol. 2, No. 4, pp. 496-514

[4] EVERITT, B.S. et al., 2011. *Cluster Analysis.* 5th ed., John Wiley & Sons, Ltd., Chichester, United Kingdom

[5] DELL'AQUILA, C., DI TRIA, F., LEFONS, E., and TANGORRA, F., 2008. Evaluating Business Intelligence Platforms: a case study. In *Proceedings of the 7th WSEAS International Conference on Artificial intelligence, knowledge engineering and data bases (AIKED'08),* WSEAS Press, University of Cambridge, Cambridge, United Kingdom, pp. 558-564

[6] GAVALEC, M. and MLS, K., 2008. Trend evaluation in on-line decision making. In *Proceedings of the 12th International Conference on Operational Research, Croatian Operational Research Society*, Zagreb, Croatia, pp. 267-274

[7] HINDLS, R., HRONOVÁ, S. and SEGER, J., 2003. *Statistika pro ekonomy*. 4th ed., Professional Publishing, Prague, Czech Republic

[8] HUB, M. and ZATLOUKAL, M., 2009. Usability Evaluation of Selected Web Portals. In *Proceedings of the 9th WSEAS International Conference on Applied Informatics and Communications* (AIC'09), WSEAS Press, Moscow, Russia, pp. 259-264

[9] JORG, T. and DESSLOCH, S. 2010. *Near Real-Time Data Warehousing Using State-of-the-Art ETL Tools.* In Enabling Real-Time for Business Intelligence, Springer-Verlag, Heidelberg, Germany, pp. 100-117

[10] KOZEL, T. and MOHELSKÁ, H., 2010. Models of firms with mobile oriented architecture. *E+M Economics and Management,* Vol. 13, No. 4, pp. 135-142

[11] LI, Z., HUANG, Y. and WAN, S., 2007. Model Analysis of Data Integration of Enterprises and E-Commerce Based on ODS. *Research and Practical Issues of Enterprise Information Systems II,* Vol. 1, Springer, Boston, United States, pp. 275-282

[12] NOVOTNÝ, O., POUR, J. and SLÁNSKÝ, D., 2005. *Business Intelligence: Jak využít bohatství ve vašich datech*, 1st ed., Grada Publishing, Prague, Czech Republic

[13] PHAN, T.T. and NGUYEN, T.C., 2011. Vietnamese Knowledge Base Development and Exploitation. *International Journal of Business Intelligence and Data Mining,* Vol. 6, No. 1, pp. 71-89

[14] ROKACH, L. and MAIMON, O., 2008. *Data Mining with Decision Trees: Theory and Applications.* World Scientific Publishing Co. Pte. Ltd., Singapore

[15] SABHERWAL, R. and BECERRA-FERNANDEZ, I., 2011. *Business Intelligence: Practices, Technologies, and Managemen.,* 1st. ed., John Wiley and Sons, Chichester, United Kingdom

[16] SCHEPS, S., 2008. *Business Intelligence for Dummies.* 1st ed., Wiley Publishing, Indianapolis

[17] SLAVÍČEK, V., 2011. Enhancing Business Process Management with Knowledge. *E+M Economics and Management*, Vol. 14, No. 1, pp. 123-134

[18] ŠIMONOVÁ, S. and PANUŠ, J., 2007. Genetic algorithms for optimization of thematic regional clusters. In *Proceedings of the 2007 IEEE International Conference on Computer as a Tool (EUROCON 2007),* IEEE Computer Society, Warsaw, Poland, pp. 798-803

[19] VERCELIS, C., 2009. *Business Intelligence: Data Mining and Optimization for Decision Making*. 1st ed., John Wiley and Sons, Chichester, United Kingdom

[20] WU, D., DONG, Y., LIN, J. and ZHU, M., 2005. *Construction of Security Architecture of Web Services Based EAI*. In Advances in Web-Age Information Management, 1st ed., Springer-Verlag, Heidelberg, Germany, pp. 601-609

JEL Classification: C12, C82, D81, L63

Appendix

The list of analysed data resources

MACROECONOMIC INFORMATION

Ministry of Finance of the Czech Republic	Bank Hypoteční banka, a.s.	Bank Československá obchodní banka, a.s.
Ministry of Industry and Trade of the Czech Republic	ING Bank	Bank Česká spořitelna, a.s.
Czech Statistical Office	LBBW Bank CZ, a.s.	Bank Komerční banka, a.s.
Czech National Bank	Bank mBank - retail banking BRE Bank SA	UniCredit Bank Czech Republic, a.s.
Institute of Health Information and Statistics of the Czech Republic	Oberbank AG	Raiffeisenbank a.s.
Prague Stock Exchange, a.s.	Volksbank CZ, a.s.	AXA Bank Europe
Patria Finance, a.s.	Czech Export Bank, a.s.	Banco Popolare Czech Republic, a.s.
Power Exchange Central Europe, a.s.	Wüstenrot Mortgage Bank a.s.	Citibank a.s.
RM-SYSTÉM Czech Stock Exchange, a.s.	Finance Media a.s.	Bank Evropsko-ruská banka, a.s.
FIO Bank	Economia a.s.	GE Money Bank, a.s.
Ministry of Labour and Social Affairs of the Czech Republic		

LEGAL ENTITIES

Access to Registers of Economic Subjects / Entities (Administrativní registr ekonomických subjektů)	Czech internet B2B market of Buying and Selling - ABC Českého Hospodářství
The European e-Justice Portal	Provider of demands and business opportunities - ePoptávka.cz
Central Register of Debtors Czech Republic	Provider of demands and business opportunities - AAApoptávka.cz
Creditinfo Solutions / Creditinfo Czech Republic	Czech Capital Information Agency (Česká kapitálová informační agentura)

Czech Credit Bureau	HBI company database
COFACE Czech	Czech News Agency
D & B	

MARKETS, PRODUCTS, COMMODITIES

Raw materials and energies	***Other useful Czech information resources***
GEOFOND - Czech Geological Survey	Public Database (Czech Statistical Office)
OKD, a.s. - Czech producer of hard coal	Czech Statistical Office
SROTY.CZ - scrap information service	Ministry of Finance of the Czech Republic
Steel Federation, a.s.	CzechTrade
TZB-info.cz - Czech portal providing information about construction, energy conservation, and related fields	Economic Chamber of the Czech Republic
Vodárenství.cz - Czech portal providing information	Czech Confederation of Commerce and Tourism
The Energy Prices - Czech portal providing the information about prices of energies	Confederation of Industry of the Czech Republic
Ministry of Industry and Trade of the Czech Republic	Territorial information provided by Business Info (mostly for small and medium enterprises)
Textile, apparel and leather industry	Colloseum, a.s. - Financial Markets, respectively eMAX - commodities, exchange rates, shares
Association of Textile-Clothing-Leather Industry of the Czech Republic	iKomodity.cz - Czech agriculture advertising commodity portal
Research Institute of Cotton, a.s. (Výzkumný ústav bavlnářský) and portal Textil.cz	Financial portal - Kurzy.cz
The European Apparel and Textile Confederation	European and worldwide enquiries, offers to buy, demands - B2B poptávka
The International Textile Manufacturers Federation	Economic newspaper - Hospodářské noviny
Agriculture industry	Daily newspaper E15.cz
Farm Accountancy Data Network Czech Republic	***Czech commodity exchanges***
Union of Growers and Processors of Oil Plants (Svaz pěstitelů a zpracovatelů olejnin)	Czech Moravian Commodity Exchange Kladno
Institute of Agricultural Economics and Information (Ústav zemědělské ekonomiky a informací)	Power Exchange Central Europe, a.s.
Agronavigátor - portal providing agriculture inforamtion	Commodity Exchange HRAPRAKO
Agricultural and Food Library (Zemědělská a potravinářská knihovna)	Říčany Commodity Exchange
The State Agricultural Intervention Fund (Státní zemědělský intervenční fond)	***Foreign commodity exchanges***
Agricultural Association of the Czech Republic	London Metal Exchange
COPA - Committee of Professional Agricultural Organisations in the European Union a COGECA - General Committee for Agricultural Cooperation in the European Union	CME Group - merge of exchanges Chicago Mercantile Exchange, Chicago Board of Trade and New York Mercantile Exchange
Food and Agriculture Organisation of United Nations	Tokyo Commodity Exchange
Ministry of Agriculture of the Czech Republic	NYSE Euronext
Other useful international information resources	NYXdata - NYSE Technologies
Organisation for Economic Co-operation and Development (OECD) Statistics Portal	Eurex
Market Access Database	Intercontinental Exchange
Bloomberg	

NOTES

- a.s. = Public Limited Company
- in case of no English equivalent data resource name, the translation is provided together with Czech original name of the institution/resource in brackets

This article should be cited as:

BURES, V., & OTCENASKOVA, T., & JASIKOVA, V., 2012. The evaluation of external data resources for business intelligence applications: the example of the Czech, *Journal of Systems Integration* 3 (1), pp. 32 - 44. [Online] Available at: http://www.si-journal.org. ISSN: 1804-2724