A BUSINESS INTELLIGENCE SOLUTION FOR BUSINESS CONTINUITY AND SAFETY MANAGEMENT IN PUBLIC UNIVERSITIES

[a]ATHANASIOS PODARAS, [b]TOMÁŠ ŽIŽKA, [c]DANA NEJEDLOVÁ, [d]DAVID KUBÁT

Technical University of Liberec, Faculty of Economics, Department of Informatics, Studentská 1402/2, 461 17, Liberec 1, Czech Republic
email: [a]athanasios.podaras@tul.cz, [b]tomas.zizka@tul.cz, [c]dana.nejedlova@tul.cz , [d]david.kubat@tul.cz

Abstract: The article introduces a modern business intelligence solution for facilitating business continuity and safety management proactive decisions in public organizations and units, which is currently tested in a public university for its effectiveness. The tool's data dimensions, hierarchies and facts are based on the business continuity points method which is a modern approach for estimating proactively the recovery time and predicting the criticality level for individual business functions. From the constructed dataset, selected safety – related and highly critical business functions are used to validate the proposed contribution. The same functions are further used for estimating their availability rates and compare the results with the rates proposed by the university business continuity experts. The conducted research results indicated high accuracy when predicting criticality levels as well as computing availability rates for safety critical functions in the public university. The proposed BI tool facilitates both online analytical processing operations as well as machine learning activities.

Keywords: availability, business continuity, business continuity points, business intelligence, machine learning, public university, safety critical business functions, safety management

1 Introduction

The interruption of modern business operations and especially those related to safety management is an issue which is thoroughly discussed and analyzed by experts from the business sector as well as academic researchers. Business continuity Management (BCM) is a topic which is strongly related to safety management. More precisely, the part of an integrated business continuity management entitled Business Impact Analysis is defined as "*a process that identifies and evaluates the potential effects (financial, life/safety, regulatory, legal/contractual, reputation and so forth) of natural and man-made events on business operations*" (Gartner, 2017).

Regardless of the important role of business continuity policies in the secure and uninterrupted operation of core business operations within modern organizations, the application of standard BCM regulations is so far considered to be a hard task. A study conducted by Urbanec & Urbancova (2014) reveals that modern organizations are skeptical in terms applying standard BCM strategies.

In universities, the situation is rather equivalent to other public organizations. The recent COVID-19 outbreak forced academic institutions to conduct research and teaching activities online as a result of the necessary health and safety countermeasures. Such crisis response activities require high network availability in order to ensure the uninterrupted operational mode of each public university to a minimum acceptable level. However, business functions which are rather ignored during the normal operational period, during the epidemic period have been treated as highly important. Distant lectures have been implemented to ensure that the spread of the epidemic is controlled. Immediate information distribution via email or the web site of every public university regarding exceptional regulations and countermeasures against the spread of the disease has been considered as crucial. In general, the importance of ensuring the continuity and the availability of core IT infrastructure and the safety of personnel in public organizations, including universities, throughout the epidemic times has been undoubtedly realized to a considerable extent.

The current research attempts to highlight the importance of business intelligence systems towards the formulation of effective business continuity and safety management policies. For the needs of the current research a number of university BCM regulations have been considered. However, the most complete BCM guide is provided by Columbus College (2018). Based on the study of different university BCM regulations, it can be concluded that every academic institution follows BCM strategies adjusted to its individual needs. As a consequence, it is not hard to define common recovery priorities for every academic institutions. Yet, the thorough study of multiple university and college BCM templates, can facilitate a basic pattern software based BCM and safety management solution when data regarding common business functions is utilized.

Due to the above mentioned BCM peculiarities among university institutions, standard theoretical methodologies and approaches or even sophisticated software tools that could support demanding BCM activities and knowledge discovery within the domain such as the establishment of recovery priorities for unique business functions, are not available in the literature. The same holds for other public organizations as well as enterprises that operate in the private sector.

In order to fill this gap, a standard mathematical method for classifying individual business functions, entitled business continuity testing points (or simply business continuity points) (Podaras et al, 2016) has been recently developed. The method focuses on estimating the recovery complexity of an individual business function. This estimation can facilitate the classification of a business function as critical or non-critical (Podaras, 2018) with the help of specific mathematical computations and data mining rules. So far the approach has been based on empirical lab computations and a dataset that has been constructed by the research team. In the present study, a real data set from a public college is used for further validation of the method. Data about 42 critical business functions is gathered, and used for testing the validity of the BCPTs approach.

For the purposes of the current paper, it is considered necessary to focus more on highly critical operations and systems which are crucial for ensuring the safety of the university staff as well as the students throughout the conducting of routine academic activities. The study of other university BCM policies (Rowan University, 2014; Columbus Technical College, 2018; Pace University, 2020) reveal that strict resumption timeframes and infrastructure availability are crucial prerequisites for ensuring the sustainable operation of safety-related processes as well as safety critical units and systems.

For the above stated reason, 7 safety – related operations have been chosen out of the 42 functions for which BCM data was collected, in order to further validate the BCPTs method and propose a business intelligence solution for BCM based on dimensions, facts, hierarchies and rules stemming from a mathematical and strictly validated approach.

Based on the above, the goal of the present article is the proposal of a business intelligence solution which is aimed to support decision making with respect to the rapid response to unexpected disruptions regarding safety – related business operations and ensure effectiveness in terms of business continuity and safety management in public organizations and units. The goal is supported by a number of important research objectives as follows:

- Incorporation of the business continuity points recovery parameters in order to design a conceptual business intelligence BCM tool and develop its physical data warehouse solution to support the proposed mathematical approach, classify accurately each business function in terms of recovery priority and compute proactively its maximum allowed downtime (or maximum recovery time).
- Utilization of real business continuity data to test the validity of the business continuity points as well as the functionality of the proposed business intelligence tool. In the present study, data regarding safety –related operations

in a public university are stored in the data warehouse via a developed web-based application.
- Demonstration of the descriptive and the predictive decision making possibilities of the proposed solution via online analytical processing and machine learning practical examples respectively.
- Utilization of the estimated maximum recovery time to compute the availability rates of the same safety – critical operations.

Based on the above stated objectives the rest of the paper is organized as follows:

Section 2 is devoted to the problem statement and the provision of the necessary background information. Section 3 includes a brief delineation of the business continuity recovery time effort estimation mathematical approach entitled *Business Continuity Points* for which a data warehouse schema is proposed as a database solution to host data for safety critical business functions. Business intelligence and data warehouse fundamentals are also included. Section 4 is used for analysis of the conducted results, including a thorough discussion with respect to the accuracy of the proposed solution in business continuity and safety management operations. The article is finalized with the conclusions and the future research directions.

2. Problem Statement and background information

"A Safety Critical System is such a system which has the potential and may cause accidents either directly or indirectly. Failure of such systems can result in loss of life, property damage, environmental harm and financial loss. Safety is dependent on proper operations of such systems" (Srinivas Acharyulu & Seetharamaiah, 2015). It is thus, important to classify such systems as highly critical in terms of recovery priority establishment stemming from the computation of their recovery time, and bearing in mind that such systems should operate without or with minor interruptions. As a consequence, the incorporation of mathematical tools and software solutions for criticality ranking of industrial business functions which are dependent on safety critical systems becomes a clear necessity. An interesting mathematical approach by Torabi et al (2014) refer to proactive recovery time estimation of critical business functions based on multiple criteria decision making. However, the method implements criticality ranking for a group of BFs and does not focus on the peculiarities and the unique technological, user and process related features as well as the environmental parameters of an individual BF.

Moreover, software tools which have been designed and developed for the BCM domain (Šimonová, S., & Šprync, O., 2011) though proactive, they serve as tools which manage operational failures by focusing exclusively on the technical aspects of the business functions, and do not take into account the environmental aspects (i.e. experience of the end user, users' motivation) of an individual business function. Additionally, mathematical models which are proposed in combination with ICT - based solutions (Sahebjamnia et al, 2014) behave as reactive (not proactive) BCM and disaster recovery planning (DRP) solutions for the resumption of critical operations after their failure.

The current contribution is proposed based on the gap which is realized from the study of the available literature, according to which none of the sophisticated BCM tools and methods computes the recovery time of individual business functions based on input data that stem from the unique technological and the environmental features of this function. Additionally, "data collection is an important activity throughout the BCM development process" (Engemann & Henderson, 2012) and every "resilient organization, through an enhanced sensing capability, integrates business intelligence in order to improve situational awareness" (Starr, 2003). The current research is devoted to the construction of a business intelligence software tool which can efficiently support data collection towards the precise classification and recovery time estimation of a given business function. Moreover, the computed timeframes can be used as a standard input for system availability measurement for safety – related functions in public organizations. The data warehouse features are conceptualized based on the business continuity points method. In the current work, real data from a public university are used for validating the initial BCPTs method, the suggested BI tool as well as the availability result for safety functions in the public institution.

3. Tools and Methods

3.1 The Business Continuity Points (BCPTs) approach

The approach (Podaras et al, 2016) focuses on the proactive estimation of the recovery time effort for an individual business function and its corresponding criticality ranking. The algorithmic process for calculating the Recovery Time is below depicted (Fig.1).

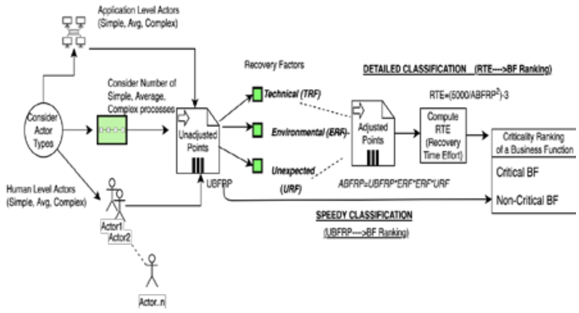

Fig. 1. The summarized model of the BCPTs Approach. (Source: own work).

For the better interpretation of the derived results we briefly mention that for the estimation of the recovery time effort we have to consider a set of recovery complexity parameters.

3.1 Unadjusted Business Function Recovery Points (UBFRP)

In order to compute the specific value human and application level actors have to be considered along with their corresponding impact (weight) on the recovery process. Moreover, the number of the involved processes and the level of complexity of each process has to be calculated. Summing up all these unadjusted parameter values, the unadjusted business function recovery points variable is computed (Fig.1).

The general function which is utilized to compute the UBFRP value is provided by Eq. (1):

$$UBFRP = \sum_{i=1}^{n}(HA_i * W_{HA_i}) + \sum_{i=1}^{n}(AP_i * W_{AP_i}) + \sum_{i=1}^{n}(BF_i * W_{BF_i}) \quad (1)$$

Where, HA=Human Actor i, AP=Application or Technical Actor i, BF= Business Function (or process or activity) i and W is the weight or importance of the given parameter. The corresponding values for each level of importance (W) are defined as follows:

Simple: 0.5, Average: 1 and Complex: 1.5.

Example: based on the available BCM data from the Columbus Public College, the following parameters are considered, regarding the safety critical operation named as *emergency communication*:

Human Actors: 6 Human Actors (including 1 Process Manager (Complex Level: 1.5), and 1 backup employee (Average Level: 1)

Technical Actors (mainly software tools): 5 defined Technical Actors (SW and IT infrastructure) , including 2 complex (1.5), 2 average (1) and 1 simple (0.5).

8 delineated highly important processes (complex: 1.5) based on the function description and the recovery strategy overview.

Thus the UBFRP value is computed as follows:

UBFRP = (1*1.5+1*1)+ (2*1.5+2*1+1*0.5)+(8*1.5)=20 points.

3.2 Adjusted Business Function Recovery Points (ABFRP)

For the computation of the specific value the following recovery complexity paremeters are considered:

- Technical Recovery Factors (TRF)
- Environmental Recovery Factors (ERF)
- Unexpected Recovery Factors (URF), and
- Recovery Time Effort (RTE)

$$ABFRP = TRF * ERF * URF * UBFRP \qquad (2)$$

$$RTE = \frac{5000}{ABFRP^2} - 3 \qquad (3)$$

The Business Continuity Points method relies on the *recovery complexity* concept following the software/system complexity principle of Karner (1993) to estimate the effort required to develop an information system.

According to lab computations, the criticality ranking can be, however, determined without the computation of the resumption timeframes for specific UBFRP values (Podaras et al., 2016). The current work presents the developed decision support data warehouse schema, which currently implements the criticality ranking of individual processes without the computation of the resumption timeframes. The specific classification is entitled "speedy" criticality ranking. The more detailed criticality ranking data warehouse solution is currently under development.

Based on preliminary lab computations and after validating the general BCPTs business rules (Podaras, 2018) the following decision making algorithm has been generated and applied in the current BI solution.

Rule 1: "speedy classification of a business function based on UBFRP"

Empirical lab computations led to the construction of a data set including 46 business functions which has been used for machine learning classification of a business function based on UBFRP input (Podaras, 2018). The classification rule induced via this study is the following:

```
IF UBFRP<9.7 Points THEN
       IF UBFRP<14.45 Points THEN
              IF UBFRP<20.89 Points THEN
                   Criticality Level = L2
                   (Critical Operation
                   RTE_MAX=24Hours)
                   ELSE Criticality Level
                   = L1 Critical
                   Operation
                   RTE_MAX=2Hours)
              END IF
       Criticality Level = L3
       (Non-Critical Operation
       RTE_MAX=72hours)
       ELSE Criticality Level = L2
       (Critical Operation RTE_MAX=24Hours)
       ENDIF
Criticality Level = L4 (Non-Critical
Operation RTE_MAX=168Hours)
ELSE Criticality Level = L3 (Non-Critical
Operation RTE_MAX=72Hours)
END IF
```

Based on this rule, a speedy classification can be implemented with approximately high accuracy. However, only the maximum recovery time that is mapped to the corresponding IVL (Gibson, 2010) can be assigned. Precise recovery timeframes cannot be determined via the speedy classifier.

Rule 2: The Recovery Scenario (RS) Selection

The rule-based Recovery Scenario (RS) selection of individual operations is illustrated as a decision tree which has been derived via the R software package (Yadav & Roychoudhury, 2018) after importing and processing the lab-based empirically derived data (Fig. 2)

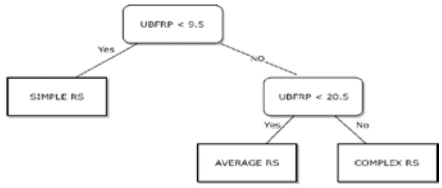

Fig. 2. The decision tree for selecting the appropriate recovery scenario. Source: (Podaras, 2018).

The semantics used in the above illustrated decision tree has the following meaning:

- Simple RS: TRF=URF=ERF=0.85. The value is constant which means no international units are utilized (Podaras et al., 2016)
- Average RS: TRF=URF=ERF=1 and
- Complex RS: TRF=URF=ERF=1.15.

As it was previously stated, the scenario selection and the computation of these parameters is important for the precise computation of the recovery time which is not included in the current version of the BI solution.

3.2 Business Intelligence Data Warehouse Preliminaries

Multiple academic researchers and business experts have provided precise delineation and definition with respect to the business intelligence data warehouse systems. A representative definition considers a data warehouse as "a collection of methods, techniques, and tools used to support knowledge workers — e.g., senior managers, directors, etc. — to conduct data analysis that helps with performing decision making processes and improving information resources" (Golfarelli & Rizzi, 2009). When data warehouse systems are integrated, a standard procedure regarding the design process is the consideration of multiple dimensions, the facts which indicate the measurable variables of these dimensions and the key attributes for dimensions and facts (Romero & Abelló, 2010). The data warehouse schema consists of several dimensions and a single fact is known as multidimensional schema or star schema.

In the multidimensional schema, "facts correspond to events which are usually associated with numeric values known as measures and are referenced using the dimension elements" (Caniupán et al., 2012). Moreover, "dimensions are modelled as hierarchies of elements, where each element belongs to a category. The categories are also organized into a hierarchy called hierarchy schema." (Caniupán et al., 2012).

Finally, based on the traditional design approaches regarding the relational as well as the object-oriented database models, three relevant design categories are distinguished, that is the conceptual, logical and physical design (Vaisman & Zimanyi, 2014). In the results section both the conceptual and the physical design of the proposed data warehouse are illustrated due to their importance.

4. Results and Discussion

4.1 The Proposed Business Intelligence Data Warehouse Tool based on the Business Continuity Points – Granularity and Hierarchies

The concept of "granularity" in business intelligence data warehouse theory appears as a "solving problem technique in which a complex problem is subdivided into smaller components or granules to facilitate information processing" (Yao, 2019). Furthermore, information granularity in all the categories of applications, acts as an element that elevates aggregate models to higher abstraction levels and enables quantification of different results in individual levels in order to ensure enhanced knowledge-based and data-oriented decision support (Pedrycz et al., 2014). In a data warehouse schema, the granularities are illustrated via the hierarchy schema where the predetermined level of detail according to which the dimensional data should be analysed is represented (Fig.3). The Business Continuity Points approach includes the following dimensions:

D1: Business Operation Level – the organizational operations for which the recovery complexity data should be stored in order to implement the proactive criticality ranking are classified according to the following hierarchy levels: business function, business process, activity and task. The task level is the lowest level of detail (granularity) for which recovery data will be analyzed.

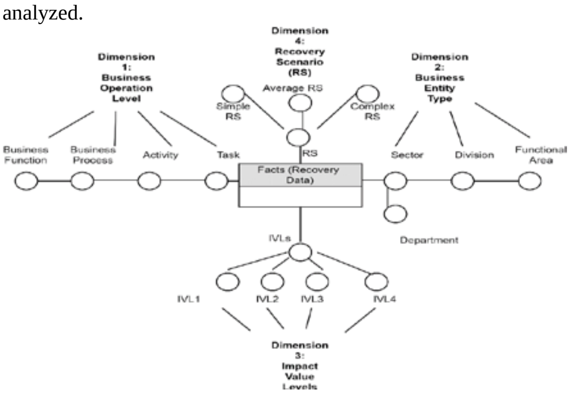

Fig. 3. Dimensions and hierarchical schema of the BCPTs business intelligence data warehouse model (speedy criticality ranking). Source: (Source: own work).

D2: Business_Entity Type- The specific dimension is utilized for defining the type of business entity, which might be a functional area, a division, a sub-division, a sector, or a department. The smallest unit size indicates the data granularity regarding the business units involved in the recovery process of an individual function.

D3: Impact Value Levels- this dimension refers to the recovery time effort (RTE) value based on the classification RTO and MAO values proposed by (Gibson, 2010). Furthermore, information regarding the corresponding impact value levels is included (IVL types). According to the specific classification, the lowest level of detail is IVL 4.

D4: Recovery_Scenario- The dimension includes the list of possible recovery scenarios. The proposed scenarios regarding the recovery procedure are categorized as simple, average and complex (Podaras et al., 2016). During the speedy classification of a business function, the scenario type can be determined by the system, after the input of the data regarding the involved Actors and Business Processes has been terminated (Podaras, 2018).

Facts and measures: the facts entity includes the calculated values of the recovery parameters. The fact include measurement regarding the total weight recovery complexity parameters namely the UHW, UAPW, UBFW and UBFRP values after considering the input parameters (attributes) regarding the number of simple, average and complex actors (both human and application level types) which participate in the execution of the business operation, as well as the number of the involved simple, average and complex business activities.

4.1.1 Conceptual Schema

The core elements of the classical conceptual data warehouse schema include the dimensions as well as the single facts in the form of entities. The highest level conceptual schema does not include the attributes of each dimension that indicates the hierarchies as well as the data granularity. Nevertheless, a further detailed conceptual schema can provide information regarding all the included attributes as well as the primary key (PK) and foreign key (FK) semantics for both the dimensions and the facts. The high –level (Fig. 4), as well as the detailed (Fig. 5) conceptual models for the proposed BCPTs data warehouse, are currently illustrated.

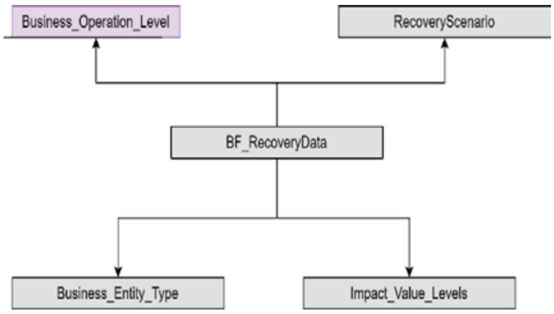

Fig. 4. The high-level conceptual BCPTs data warehouse model (speedy criticality ranking) (Source: own work)

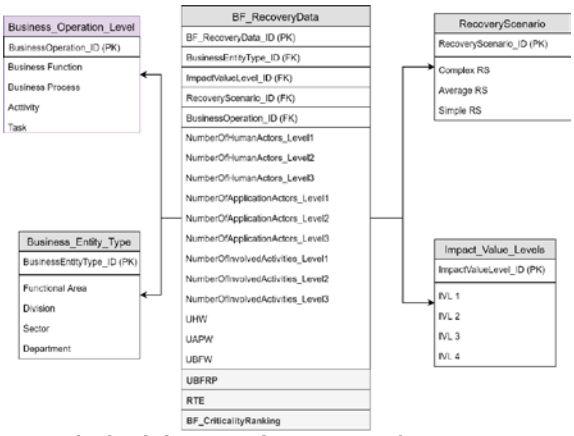

Fig. 5. The detailed conceptual BCPTs DW schema.
(Source: own work).

4.1.2 Physical schema and Web-Based System Architecture

Based on the above illustrated conceptual data warehouse schema, a corresponding physical data model has been developed. The physical model includes information about the data types of each attribute as well as the length of each data type (Fig. 6). The model was developed in MySQL (Welling & Thomson, 2017). The developed physical data warehouse is used as the back end repository where the business continuity data shall be stored via a simple and user-friendly web application. The application is developed in PHP programming language (Welling & Thomson, 2017). The proposed BCPTs application interface is utilized for the recovery complexity computations. Moreover, the connection of the API and the data warehouse is supported by HTML 5.0 and CSS technology (Brown et al., 2014).

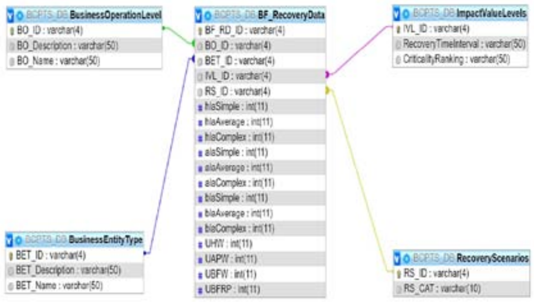

Fig.6. The physical data model of the BCPTs data warehouse

The input (Fig. 7) and output screens (Fig. 8) which include the data that are stored in the target BCPTs data warehouse are below depicted.

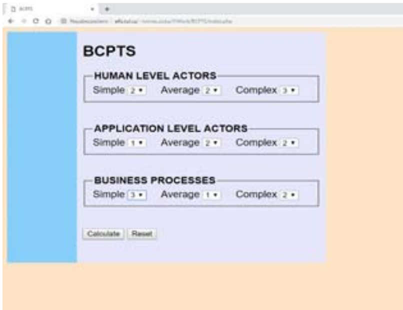

Fig.7. Input BCPTs screen

Fig.8. Ouptut BCPTs screen

The overall BI system architecture is also included for the better interpretation of the web-based solution (Fig. 9).

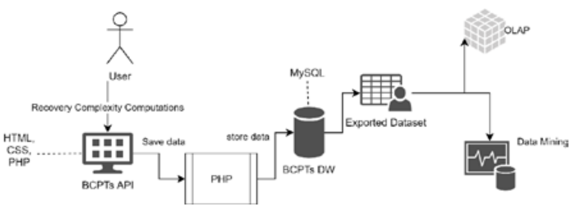

Fig. 9. The proposed BCPTs BI System Architecture

The proposed architecture's main advantage is that the data loading process is not complicated and not time demanding due to the avoidance of core ETL (Extraction, Transformation and Loading) activities such as data formatting, validation and transformation. The users can directly store the business continuity data regarding their departments, divisions, functional areas. Thus, the problem of data extraction from various sources based on flat files such as excel, text documents, or email communication, which indicate mandatory ETL activities, is solved with the existence of the BCPTs API. It can be observed that the application can be used for exporting data in a spreadsheet format (.xls or .csv) and utilize it for online analytical processing operations and data mining predictive decision making activities.

4.1.3 Online Analytical Processing (OLAP) Operations

The online analytical processing services can be provided via exporting the data as a spreadsheet document or directly via the database solution in the form of a query. The proposed schema is based on the relational database design and implementation approach. As a consequence, SQL (Structured Query Language) queries can be utilized as explanatory representative business continuity OLAP (Sohrabi & Azgomi, 2019) descriptive operations.

The OLAP covers the analysis services task where the analysis of the recovery data is based on the UBFRP value for a single operation. The granularity is based on the operational level (function, process, activity, and task) as well as the unit level.

Example: The following SQL query is a representative and simple OLAP aggregate operation. The executed query computes the average UBFRP for individual business functions from a determined operational level:

```
SELECT BF_RecoveryDATA.BusinesOperation_ID,
Avg(BF_RecoveryDATA.UBFRP) AS AvgOfUBFRP
FROM BusinessOperationLevel INNER JOIN
BF_RecoveryDATA ON
BusinessOperationLevel.BusinessOperation_ID
= BF_RecoveryDATA.BusinessOperation_ID
GROUP BY
BF_RecoveryDATA.BusinessOperation_ID;
```

4.2 Safety Critical Computations Based on the Exported BCPTs Spreadsheet Data: Evidence from Real University BCM Data

4.2.1 Validation of the BCPTs speedy classifier and the computations supported by the proposed BI tool

The currently proposed contribution, as every proposed business intelligence solution, serves as a tool for effective and efficient predictive decision making. Predictive analytics that facilitate crucial decisions regarding future trends in public organizations are based on machine learning activities. The current version of the proposed solution supports classification of critical business operations based on the UBFRP input. As a consequence, according to Rule 1 (Section 3) the UBFRP input recovery variable can indicate the Maximum Recovery Time Effort (RTE_{MAX}) required to recover a business operation.

The currently developed web-based business intelligence application (Fig.5, Fig.6) has been used to validate the business continuity management policies based on the BCPTs computations. Taking into consideration real BCM data from a public university (Columbus Technical College, 2018) a spreadsheet dataset in .csv format has been created. Business continuity parameters have been used as the input data to infer robust maximum recovery time computations via the proposed business intelligence schema. The data have been stored in the physical database from which they have been exported in the form of a .csv file. Part of the data set is below depicted (Tab.1). The table includes arbitrary data related to 7 selected safety-critical operations in a public university. The full data set can be accessed via the link:

Tab.1 Business continuity data for safety-related functions in a selected public university (Source: own work based on data from (Columbus Technical College, 2018)

Business Function	Number of Human Actors	Number Of Involved Processes	Number Of Technical Actors	UBFRP (points)
Emergency Communication	12	2.5	5.5	20
Public Information	10.5	4	5.5	20
Risk Management	3	2.5	3	8.5
Police and Security	9	2.5	1	12.5
Mail Services	12	3.5	1.5	17
Core IT Systems	9	2.5	2.5	14
Emergency Services	13.5	2.5	0	16

Based on the inferred UBFRP computations and according to the BCPTs classifier (Podaras, 2018) specific impact value levels can be assigned for every individual function which can then be compared to the proposed by the university BCM experts impact value levels. These levels do not appear in the utilized university BCM guide but they have been inferred via mapping the proposed by the university BCM recovery team resumption timeframes for these functions. According to this value a corresponding IVL has been defined and compared to the predicted IVL via the BCPTs approach and the currently proposed business intelligence tool. In this way, the BCPTs accuracy can be verified (Tab.2)

Tab.2 The comparison between the predicted IVL with the proposed by university BCM team members IVL

Business Function	UBFRP (points)	Impact Value Level (IVL) (BCPTs prediction)	Proposed Recovery Time (by the BCM Team) (hours)	Proposed IVL (by the BCM team) (hours)
Emergency Communication	20	L2	24	L2
Public Information	20	L2	24	L2
Risk Management	**8.5**	**L4**	**24**	**L2**
Police and Security	12.5	L3	48	L3
Mail Services	17	L2	24	L2
Core Technology Infrastructure	14	L3	72	L3
Emergency Services	16	L2	24	L2

From the above recorded predictions it can be concluded that the BCPTs classifier and the incoproration of the currrent business intelligence web application may infer highly accurate business continuity management predictions for safety-related business functions in public universities. Based on the selected case study the predictive accuracy of the BCPTs classifier is 85.71% (6 out of 7 criticality ranking predictions have been proved correct for safety-related operations).

4.3 Data Mining Tasks

The exported spreadsheet data can be further used for machine learning activities such as classification and regression techniques. So far, the BCPTs speedy classifier has been based exclusively on UBFRP input variable for the prediction of the Impact Value Level. Based on the data exported by the currently proposed BI solution (Tab.1), more input (explanatory) variabes can be utilized for predicting criticality ranking for individual business function. Moreover more association rules among the incorporated variables can be explored. The machine learning path for BCM knowledge discovery is illustrated (Fig.10)

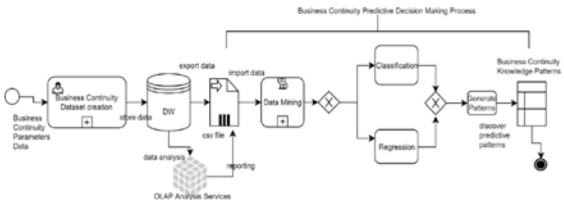

Fig. 10. The machine learning path for business continuity predictive knowledge discovery via the proposed BCPTs BI tool. Source: (Source: own work).

Regression analysis tasks can be also implemented in order to predict recovery time values. However, this is not feasible when relying exclusively on the speedy BCPTs classifier. Data mining regression analysis tasks can be implemented after considering the appropriate recovery scenarios, a set of technical, environmenatal and unexpected recovery factors (TRF, ERF, URF) and after estimating the Adjusted Points variable (ABFRP) for conducting precise recovery time computations.

Another issue which requires further clarification is the possibility to boost the predictive accuracy of the speedy classifier. Several robust ensemble classification techniques, such as random forests, k-folds cross validation, k-NN (nearest neighbor), logistic regression and support vector machine (SVM) may be incorporated. However, importing the conducted .csv reports into a sophisticated machine learning software package is also demanded. One of the most commonly utilized machine learning free software tool is the R package (Rahlf, 2017). Currently the possibility to connect the proposed business intelligence tool with the R package for faster machine learning activities is under consideration.

In order to test the predictive accuracy of the speedy classifier (as explained in rule 1), the full dataset (42 critical business functions) has been imported into the R-Package. We used the CART decision tree algorithm (Breiman et al, 1984) along with the 10-folds cross validation (Machine Learning Mastery, 2018) and the random forests (Breiman, 2001) classification algorithms in order to test the BCPTs accuracy in predicting Impact Value Levels (else criticality ranking or recovery priorities) for the entire data set. The ensemble machine learnning techniques have been used to avoid overfitting. The evaluation metrics used for measuring the accuracy of the three tested machine learning techniques has been the confusion matrix. The advantage of the data mining is the possibility to conduct additional classifiers based on different inputa variables and to investigate diverse association rules among the variables included in the data set. For example, we could focus on exploring the accuracy in IVL predictions based only on human actors or considering only the number of involved processes as input. We may also use association rules to explore relationships among the included variables (Fig. 11).

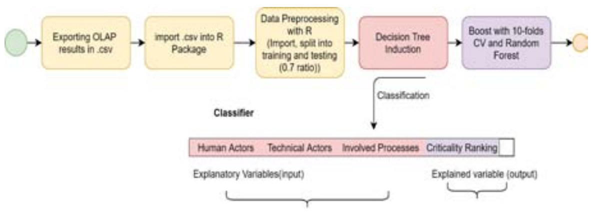

Fig.11. The data mining classification procedure for critical business functions based on a public university data set.

For practical demonstration, the entire dataset has been used to predict IVL based on UBFRP input. The data set is composed of 42 records and 1 input (UBFRP) and 1 output variable (IVL). The data preprocessing procedure includes importing data and splitting data in a logical ratio beteween training and testing data. In our example, the splitting ratio has been 0.7 (70% training data, 30% testing data. For the induction of the decision tree, the library *rpart()* has been loaded.

For the advanced classifiers more packages have been required. The 10folds CV requires the *e1071* and *caret* libarires, while the *randomForest* library has been loaded to investigate the random forest classification accuracy. The induced decision (Fig.12) tree as well as the random forest out of bag (OOB) error plots (Fig.13) are ilustrated. The OOB estimate relies on observations which are not considered in the bootstrap sample. The error plot (Fig. 13) shows that 500 trees were induced for enhancing the classification robustness. The prediction error is reduced and stabilized after a specific number of induced trees (<200 trees).

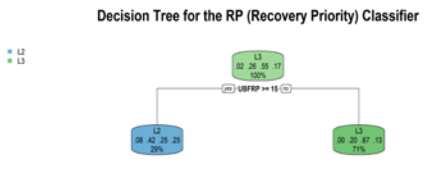

Fig. 12. The decision tree for predicting IVL criticality ranking of crucial business functions in a public university based on the full data set (Source: Author)

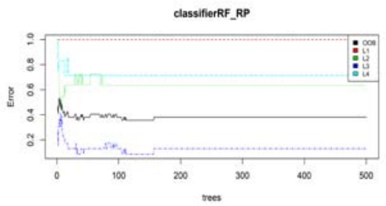

Fig. 13. The IVL classification out of bag (OOB) error plot based on the random forest based on the full data set (Source: Author)

The accuracy of the decision tree, 10-Folds CV and the random forest algorithms is ilustrated on the following summarized table (Tab. 4).

Tab.4 Comparison of the predictive accuracy of the three different classifiers with respect to the recovery priorities for the entire set of critical business functions in the public university.

IVL Classifier	Accuracy (based on the confusion matrix results)
Decision tree (CART)	75%
10folds CV	75%
Random forest	91.66%

After considering the above illustrated results, it can be concluded that the proposed business intelligence tool can be considered as a rapid and effective software solution for business continity and safety management. So far, the research has been focused on the establishment of balanced recovery priorities for safety critical operations in public organization. However, a complete safety management policy should also rely on the robust availability measurement of these operations.

4.4 Computing availability rates for safety-related operations via RTEmax input for an integrated safety management framework

Another dimension that must be considered for more effective safety management policies in public organizations and units is the estimation of the availability of safety critical operations. An interesting recent study (Spang, 2017) highlights the importance of a safety management system in order to protect workers in all industries form electrical hazards, and defines it as a "formal and proven system for the safe execution of work activities". Moreover seven core safety management principles are indicated including "balanced priorities" in terms of protecting the "workers, the public and the environment". The business continuity points is proposed a mathematical method for setting balanced recovery priorities in the occasion of a failure of a safety critical system and the involved industrial functions.

Additionally, a description of the Integrated Safety Management system as provided by the U.S. Department of Energy (2008), indicates, among others, the "operational excellence" as an important safety management principle. The study relates operational excellence with high reliability achieved through "focus on operations, quality decision-making, open communications, deference to expertise, and systematic approaches to eliminate or mitigate error-likely situations". According to the second principle, the business continuity points is also aimed to serve as an operational excellence tool for controlling the reliability and, more precisely the availability of a safety critical function and the corresponding systems. Based on the above, the business continuity points can serve as a crucial part of an integrated safety management under the below summarized framework (Fig.14)

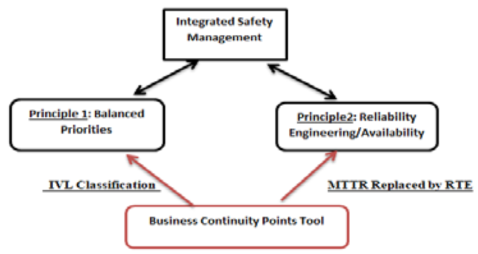

Fig. 14. Business Continuity Points as Part of an ISM – Proposed Framework (Source: Author)

Due to the fact that the current version of the developed business intelligence solution exports business continuity spreadsheet reports, a visual basic for applications (VBA Excel) software interface has been utilized for estimating the availability rates for the selected safety – related business functions based on the RTEmax vaue that is used. The VBA application has been developed to estimate availability rates based on the formula (4) (Rance, 2013) :

$$Availability = \frac{AST - DT}{AST} \times 100\% \qquad (4)$$

where, AST = Agreed Service Time, DT=Downtime. DT value is also mentioned as Mean Time To Repair (MTTR) in other availability formulas (Garcia et al, 2016)

Assuming that AST=22 hours/day/week (154 hours/week).The expected downtime will then be, DT=2 hours/day or 14 hours/week.

As a result, the weekly availability for this function is:

A_{WEEK}= [(154-14)/154] * 100%= 90.9%.

In the case that a maximum unplanned downtime interval of 8 hours/week is permitted the availability rate is then estimated as

A= [(154-14-8)/154]*100% = 85.71% , which indicates the maximum tolerable eekly availability rate, so that an organization will not suffer significantly negative consequences.

Based on the utilized case study, it has been attempted to estimate yearly availability rates based on the BCPTs proposed maximum recovery time, with respect to the same safety related business functions. The following facts have been assumed:

- each business function is highly critical, AST=24 hours/day (8760 hours/year),
- the DT value has been replaced by the RTE_{MAX} proposed by the BCPTs tool to compute the BCPTs weekly availability rates,
- the DT value has been replaced by the maximum tolerable downtime proposed by the university BCM team, in order to compute the proposed by the university weekly availability rates and compare them with our computed results,
- one outage incident per year is considered, the duration of which is RTE_{MAX} .

For the above mentioned computations a simple VBA excel tool has been developed and utilized (Fig. 15).

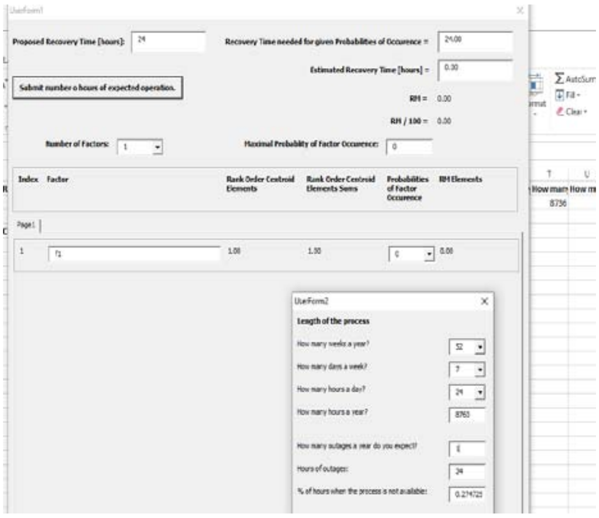

Fig. 15. The VBA Excel form for estimating availability rates for individual business functions based on recovery time input

Tab.5 Predicted availability rates for safety related university business functions.

Business Function	Proposed Maximum Recovery Time by the BCPTs BI tool (hours)	Proposed Maximum Downtime period from the university BCM team (hours	Predicted Yearly Availability Rate based on RTE_{MAX} (BCPTs approach)	Proposed Yearly Availability Rate by the University BCM experts
Emergency Communication	24	24	99.72%	99.72%
Public Information	24	48	99.72%	99.45%
Risk Management	**168**	**24**	**98.07%**	**99.72%**
Police and Security	**72**	**48**	**99.17%**	**99.45%**
Mail Services	24	24	99.72%	99.72%
Core Technology Infrastructure	72	72	99.17%	99.17%
Emergency Services	24	24	99.72%	99.72%

The conducted results indicate satisfactory performance levels for both the BCPTs approach and the proposed BI tool in terms of availability estimation for safety critical operations. It can be noticed that only for two business functions, taht is, the *risk management* and the *police and security* the computed availability slightly deviates from the proposed by university BCM experts rates. However, 5 out of 7 estimations were highly accurate, while in the case of the public information function the proposed by our approach availability rate was higher than the one proposed by the BCM experts.

5. Conclusions and future research directions

Business intelligence solutions are important for ensuring resilinece in public organizations. The big data manipulation is a modern challenge of paramount importance for business continuity and safety management. Public universities include several safety-related business functions for which rapid restoration after unexpected interruptions and high availability rates are crucial for their smooth operation. The current work has been focused on the development of a modern business continuity and safety management tool based on the business intelligence data warehouse concepts. The dimensions, facts and the defined information granularity has been designed based on the business continuity points (BCPTs) method, that is utilized for the proactive recovery priority level definition as well as the proactive computation of the resumption timeframe for individual business functions. The method is based on the computation of recovery complexity and effort estimation parameters which have been inspired by the Use Case Points approach. In the present article, the BCPTs approach has been validated via a real data set from a public university. The proposed tool is supported by a web interface which facilitates the BCPTs computations and enables the estimation of the criticality ranking and the recovery time effort estimation. The exported spreadsheet data can be used for OLAP operations and data mining activities. From the utilized data set which is composed of 42 university business functions we selected 7 safety related functions to measure the accuracy of the BCPTs classifier and the computations conducted by the proposed BI solution. The entire dataset has been also investigated with machine learning classification techniques, namely the decision trees, the 10Folds cross validation and the random forests with respect to the IVL predictive accuracy. The estimated accuracy in predicting the critility level (impact value level - IVL) has for the safety-related operations has been 85.71% which is highly promissing. For the full dataset the decision tree classification and the 10folds cross validation techniques were 75% accurate based on the confusion matrix results stemming from the testing data (30%) over the full dataset. The random forest technique was 91.66% accurate. Finally, the safety-related functions, have been used for investigating the accuracy of the current tool in estimating their availability rates. The availability rates computed via the suggestd BI tool, have been highly accurate when compared to the proposed availability rates by the university BCM experts. A developed VBA Excel simple interface has been used to support availability computations which stem from the exported spreadsheet data. Only 2 out 7 business functions slightly deviated from the proposed by experts rates. In general the results conducted throughout the present resarch are highly encouraging for the proactive business continuity and safety management in public organizations and especially for universitities which has been the target domain of the present article.

Nevertheless, crucial future research activities include the further validation of the BCPTs BI solution by extracting data from more universities, the incorporation of more data features in order to infer more advanced machine learning classification techniques and, also, the enrichemnt of the current web-based BI interface by including more functionalities. One of them is the connection of the application with sophisticated machine learning software packages such as the R package. However larger data volumes from other universities should be gathered. This task is demanding due to the fact that BCM data is sensitive and confidential in most cases. However, data from 5 more universities which is currently processed and analyzed by the research team have been so far collected and will be used for future investigation. Moreover, the incorporation of the current standalone VBA tool in the BI solution for computing the availability rates is planned. Finally, the current interface requires further testing for estimating accurate resumption timeframes based on several recovery scenarios for every individual function. In this way, machine learning regression tasks can be performed. Similar BCM BI solutions can be also proposed for other public organizations.

Literature:

1. Gartner. Magic Quadrant for Disaster Recovery as a Service. 2017. https://www.gartner.com/doc/3746618/magic-quadrant-disaster-recovery-service. (Accessed 13 January 2019).
2. Urbanec, J. and Urbancova, H. *Adoption of Business Continuity Management Standards in Czech Organizations*. Scientia Agriculturae Bohemica. 45(1). 2014. 66-74 pp.DOI: 10.7160/sab.2014.450109
3. Columbus Technical College. *Business Continuity Plan 2018-2019*. 2018. Retrieved from: https://www.columbustech.edu /skins/userfiles/files/2018-2019%20Business%20Continuity %20Plan.pdf.
4. Podaras, A., Antlova, K., & Motejlek, J. *Information Management Tools for Implementing an Efficient Enterprise Business Continuity Strategy*. E M Ekon. Manag. 19. 2016. 165-182 pp. http://dx.doi.org10.15240/tul/001/2016-1-012 .
5. Podaras, A. (2018). *Measuring the Accuracy Levels Regarding the Dual Business Function Criticality Classifier*. IEEE Access. 6. 2018. 41598-41606 pp. doi: 10.1109/ACCESS.2018.2860976.
6. Rowan University. *Business Continuity Management Policy*. 2014). https://confluence.rowan.edu/display/POLICY/Business+Continuity+Management. (Accessed 9 January 2019).
7. Pace University. *Security and Emergency Management*. Business Continuity Planning. 2020. Retrieved from: https://www.pace.edu/security-emergency-management/business-continuity-planning .
8. Srinivas Acharyulu, P.V. & Seetharamaiah, P. *A Framework for Safety Automation of Safety-Critical Systems Operations*. Safety Science. 77. 2015. 133-142 pp. doi: dx.doi.org/10.1016/j.ssci.2015.03.017.
9. Torabi, S. A., Rezaei, H., & Sahebjamnia, N. *A New Framework for Business Impact Analysis in Business Continuity Management (with a Case Study)*. Safety Science. 68. 2014. 309 – 323 pp. doi: http://dx.doi.org/10.1016/j.ssci.2014.04.017.
10. Šimonová, S., & Šprync, O. *Proactive IT/IS monitoring for business continuity planning*. E M Ekon. Manag. 14. 57-65 pp. 2011.
11. Sahebjamnia, N., Torabi, S. A., & Mansouri, A.*Integrated Business Continuity and Disaster Recovery Planning: Towards Organizational Resilience*. European Journal of Operational Research. Vol. 242. No. 1. 2014. pp. 261-273. ISSN: 0377-2217.
12. Engemann, K. J., & Henderson, D.M. *Business Continuity and Risk Management: Essentials for an Organizational Resilience*. Rothstein Associates Inc., Connecticut, USA, 2012.
13. Starr, R. *Enterprise Resilience: Managing Risk in the Networked Economy*. Available at: http://www.boozallen.com/content/dam/boozallen/media/file/Enterprise_Resilience_Report.pdf, 2003 (Accessed 8 February 2019).
14. Karner, G. *Use Case Points—Resource Estimation for objectory Projects*. Objective Systems SF AB, 1993. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.604.7842&rep=rep1&type=pdf, (accessed 8 December 2018).
15. Gibson, D. (2010). *Mitigating Risk with a Business Impact Analysis*. Managing Risks in Information Systems, second ed., Jones & Bartlett Learning, Burlington, NJ, USA. 2010.
16. Yadav, M.L., & Roychoudhury, B. Handling missing values: a study of popular imputation packages in R. Knowl. Based Syst. 160. 2018. 104-118. https://doi.org/10.1016/j.knosys.2018.06.012.
17. Golfarelli, M., & Rizzi, S. *Data warehouse design. Modern principles and methodologies*. McGraw-Hill, 2009.
18. Romero, O., & Abelló, A. *A Framework for Multidimensional Design of Data Warehouses from Ontologies*. Data Knowl. Eng. 69. 2010. 1138–1157 pp. https://doi.org/10.1016/j.datak.2010.07.007.
19. Caniupán, M., Bravo, L., & Hurtado, C. A. Repairing Inconsistent Dimensions in Data Warehouses. Data Knowl. Eng, 79–80. 2012. 17–39 pp. https://doi.org/10.1016/j.datak.2012.04.002.
20. Vaisman, A., & Zimanyi, E. *Data Warehouse Systems: Design and Implementation*. Data Centric Systems and Applications. Berlin: Heidelberg: Springer-Verlag, 2014.
21. Yao, M.X. *Granularity Measures and Complexity Measures of Partition-Based Granular Structures*. Knowl. Based Syst. 163. 2019. 885-897 pp. doi:https://doi.org/10.1016/j.knosys.2018.10.015
22. Pedrycz, W., Al-Hmouz, R., Morfeq, A. and Balamash, A.S. *Building Granular Fuzzy Decision Support Systems*. Knowl. Based Syst., 58 (2014) 3-10 pp. https://doi.org/10.1016/j.knosys.2013.07.022.
23. Welling, L., & Thomson, L. *PHP and MYSQL Web Development*. Fifth ed. Addison-Wesley. USA, 2017.
24. Brown, T.B., Butters, K., & Panda, S. *Jump Start HTML5: Get Up to Speed With HTML5 in a Weekend*. First ed., Sitepoint Pty. Ltd, Autralia, 2014.
25. Sohrabi, M. K., & Azgomi, H. (2019). *Evolutionary Game Theory Approach to Materialized View Selection in Data Warehouses*. Knowl. Based Syst. 163. 2019. 558-571 pp. https://doi.org/10.1016/j.knosys.2018.09.012.
26. Rahlf, T. *Data for Everybody*. Data Visualization with R: 100 Examples. Springer International Publishing, Switzerland, 2017. pp. 1–4. ISBN 978-3-030-28443-5.
27. Breiman L., Friedman, J.H., Olshen, R.A. & Stone, C.J., *Classification and regression trees*. Chapman and Hall. New York, USA. 1984.
28. Machine Learning Mastery. *A gentle introduction to k-fold cross-validation*. 2018. Retrieved from: https://machinelearningmastery.com/k-fold-cross-validation/.
29. Breiman, L. *Random forests*. Mach Learn. 45(1). 2001. 5–32 pp.
30. Spang, R. *Integrated Safety Management: Creating an All-Inclusive Electrical Safety Program*. IEEE Industry Applications Magazine. 64-70 pp. 2017. ISSN Online: 1077-2618. DOI: 10.1109/MIAS.2016.2600729.
31. U.S. Department of Energy – Office of Environmental Management Headquarters (2008). *Integrated Safety Management System Description*. 2008. https://energy.gov/sites/prod/files/em/EMHQISMSDescription5-7-08.pdf (Accessed 9 May 2020).
32. Rance, S. Defining availability in the real world. 2013. Available at: https://itsmf.custompublish.com/getfile.php/1283611.1559.wcsxycquer/Defining+availability+in+the+real+world.doc
33. García, I.E.M., Sánchez, A.S. & Barbati, S. *Reliability and Preventive Maintenance*. MARE-WINT. New Materials and Reliability in Offshore Wind Turbine Technology. 235-272 pp. Springer, 2016. DOI 10.1007/978-3-319-39095-6.

Primary Paper Section: I

Secondary Paper Section: IN