



(Volume 4, Issue 4)

Available online at: www.ijarnd.com

Share point and file share search engines: Can personal data be found for EU GDPR compliance?

Rade Nikolic

Leeds Beckett University, Leeds, England

ABSTRACT

This article aims to investigate the ability of Share Point and File Share Search Engines to ensure compliance with GDPR regulations. The value of this research is to contribute to academic theory and business literature on the capability of search engines to find personal data in internal and external data repositories and systems for GDPR compliance. Long term project outcomes are expected to lead to empirical results which added value to existing theory and be a foundation (or framework) for future research. Quantitative methods are used through a questionnaire and experiments to obtain data from a convenience sample from an IT and non-IT population of executives, administrators, and professionals who have exposure to Share Point and File Share data repositories. Evidence from the experiments demonstrated that under three conditions, the hypothesis can be rejected. This research, considering its limitations, has proven that personal data, metadata, and files cannot be found in all cases in Windows File Share and Share point repositories. In conclusion, improvements in search engines or enhancements of operational manipulations with data and metadata in files must be performed based on the conditions stated in the research conclusions to ensure accurate compliance with GDPR.

Keywords— Metadata, GDPR, Search Engine

1. INTRODUCTION

GDPR, according to Lynskey (2015) [[1]], is an outcome to better control individual data, clearly getting individual ownership, confirming Koop's (2011) [[2]] 'right to be forgotten' even in Big Data environments. Recognised measures are within the policies and procedures according to Bitar and Jakobsson (2017) [[3]] with some potential opportunities recognised in language technology impacting ethics in academic and industrial researches [[4]].

Endeavour to criticise compliance theory through general data protection regulations and as such it was argued that the transparency and documentation are the essences of the regulation, giving more pressure to organisations through severe penalties [[5]]. Lynskey (2015) [[1]] recognises that at the time when the first-time data protection legislation was adopted in EU in 1995, only 1% of the population was using the Internet, whilst European Parliament already developed data protection concept in the mid 1970s as a reaction to the evolution of data processing. Koops and Leenes (2013) [[6]] are advocating that every new system should include 'privacy by design' in the system which in their opinion is what GDPR is embracing.

Even though 'personal data' and 'data subject' are in the early draft version of the GDPR still are presented as an identifier or a person [[8]]. GDPR is impacting in every way both, online systems [[6]] and internet-of-things (IoT) systems [[7]].

Research done by a survey in 2017 among 500 IT decision makers [[9]] confirmed that just over 40% of the participants are not taking the final GDPR deadline as a priority. At the same time, according to the same research 75% of participants clearly agreed that fines for not being compliant with GDPR can severely damage organisations, knowing that the average IT budget in 2017 was 2.5% from the company revenue worldwide [[10]] compared with the fines that could be up to two per cent world-wide turnover for a company [[11]].

Aspect identified by Bitar and Jakobsson (2017) [[3]] on auditing for internal and external systems to provide secure organisation, the value of this research is to contribute to academic theory and business literature on the capability of search engines to find personal data in internal and external data repositories and systems for GDPR compliance. Main theory, explored through Mudassar (2013) [[12]], where Share Point is introduced over search engine functionality with data governance capacities, that data/metadata can be found, were used as starting point for this research. Project relevance is anticipated as the General Data Protection Regulation (GDPR) came into force on 25th May 2018. The regulation aims to provide more control to EU citizens over their data.

The research project was determined through a set of hypotheses analysing the subject [[13]] in both repositories:

- **Hypothesis 1:** Changing the textual data in the file into the personal data will cause that location of the data will be able to be found through the search engine for the specific syntax.
- **Hypothesis 2:** Changing the metadata of the file to personal data will cause that location of the data will be able to be found through the search engine for the specific syntax.
- **Hypothesis 3:** Changing the textual data in the file into the personal data will cause that location of the data will be able to be found through the search engine for the specific syntax regardless of metadata.

Findings were reinforced by the questionnaire to get facts on the user's behaviour, attributes, and opinions in using search engines, Windows File / Share Point data repositories and metadata. Conclusions were drawn and recommendations made on whether enhancements need to be made for bodies to be more efficient for GDPR compliance.

Given that data has become a resource [[14]], the author places value on relating this research within the context of EU legal, economic and social intentions of its legislature. Rationalizing various traditional and modern philosophical dimensions, GDPR is presented as an attempt to protect and regulate data security mainly through the scope of protecting individual and civil rights influencing other countries to perform the same level of universal protection.

Increase of information and data in modern academic and business world is growing rapidly impacting and releasing new business opportunities and values either from the PC computing or mobile phones or Internet of things however security and protecting privacy becomes more important and even more difficult to manage given the circumstances of data sharing through the internet or through other channels [[15]].

Data privacy within the New Institutionalism theory for data privacy and protection by Moller (2017) [[16]], GDPR business and academic impact on IoT and HSM devices and big data in open data share scientific communities. Through data governance theories, the author highlights a need for anticipation for preparations to achieve full compliance of the new regulation framework (GDPR) that has a unique territorial scope. Furthermore, metadata is identified as a vital element of data governance through main data governance theories. Both Share Point and File share document management repositories are introduced, as an important segment of organisational collaboration strategy linked with data governance and their specific functionalities related to the ability to find data through the search engines.

2. MATERIALS AND METHODS

Colburn (2000) identified computer science to be formalized in a way that given effort can be repeated on a computer, within a different scientific method than in engineering science given the abstract nature of code and programming through the workflow: "Natural Phenomenon - Abstract model - Machine implementation" [[17]]. This analogy is used for computer science as in a natural science observing phenomena, constructing a hypothesis and testing them through experiments adhering relationship between various elements, setting its paradigm as meeting specific requirements [[17]]. The central topic of the proposed study is the EU General Data Protection Regulation and its impact on all bodies worldwide.

The research area is gathered in topics such as GDPR, Data Governance, Data repositories (external and internal) and Search Engines. This research was based on theory-driven hypothesis that data privacy and GDPR compliance personal data are to be found in repositories within the out-of-the-box search engines. The deductive research approach was taken to explain the relationship among variables through experiments and as such precept that the research is independent of the object of observation [[19]] adhering that the further researches should explore other systems and search engines to give an answer whether personal data can be found to be compliant with the EU GDPR. In that sense, the inductive approach can be inducted to some extent through the open-ended questions in the survey to support potential further researches. Research was carried out using the deductive research approach to examine causal relationships between variables, editable data files - Office Documents file formats and mark-up textual files (XML, HTML) search engine results and controlled variables, browser, repository version and configuration of the search engine in order to collect data whether Microsoft search engines in both data repositories (File Share and Share Point) are GDPR compliant. Choice of the research strategy is driven by hypothesis aligned with the philosophy adopted for the research [[19]] and based on existing theoretical knowledge and the literature review, research strategy was carried out through experiments and survey.

Through experiments it was explored whether search engines in Share Point and File Share repositories can find personal data for the GDPR compliance within three identified variables:

- Independent variable, IV textual data file | .doc .txt .xml .html .ppt .xls
- The dependent variable, DV search engine results
- Controlled variable, CV configuration of the search engine, repository version, browser

Group of files, relevant to the research are identified to be exposed for a planned manipulation (Experiment group) and not to be exposed for a planned manipulation (Controlled group) in order to measure the difference [[19]] that indicated the ability for a search engine to find personal data:

- Controlled group - files that the content is not changed (.doc .txt .xml .html .ppt .xls)
- Experiment group- copy of the controlled files with changed content - editable data files that can carry GDPR personal data (.doc .txt .xml .html .ppt .xls).

Internet-mediated web survey was identified as a more appropriate one to gather data than postal or telephone survey and structured interviews. Based on the non-sampling frame, (convenience sample choice of IT and Non-IT personnel operating with data

repositories), design requirements are based on decision which data need to be measured, matching types of variable, identified in Dillman’s data types in Saunders et al (2009) to collect quantitative data through “investigative questions” [[19]] that have at least one of each: attribute, behavior and opinion questions. Based on the premises that both groups of participants had similar understanding and that there is no expert knowledge currently on this regulation, a limited amount of information regarding the law itself was needed to be shared without influencing participants. Ethical approval for this study was issued from Leeds Beckett University, UK.

3. RESULTS

3.1 Experiment results

Data collected, and the results of the experiments can be presented in several dimensions:

- File share search experiments results
- Share Point search experiment results
- Data content experiment results
- Metadata content experiment results

3.1.1 File Share Search Engine: Results in figure 1 indicate the ratio between found files and non-found files. One file that could not be found on the inserted search query was .xlsx file (MS Excel file).

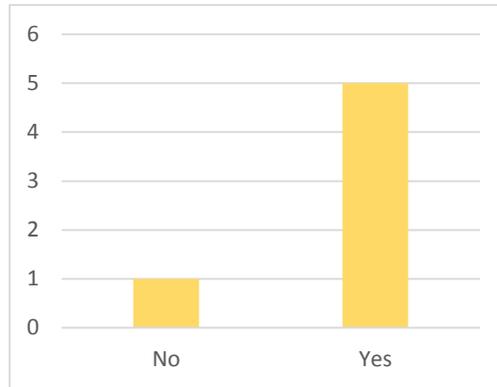


Fig. 1: Total of Files found in the File Share repositories

In the File Share experiments, there were three files changed with metadata content presented in table 1. The search results indicate different results on query syntax that was used for measuring whether the data could be found.

Table 1: File Share search engine files report on found metadata content

File ID	Filename	File type	Data Content	Data Folder	Content Found
1.	Exp_Random.docx	MS Word (.docx)	Meta Data	ExperimentGroupMeta	No
2.	Exp_Random.xlsx	MS Excel (.xlsx)	Meta Data	ExperimentGroupMeta	No
3.	Exp_Random.pptx	MS Power Point (.pptx)	Meta Data	ExperimentGroupMeta	No

As the content on metadata could not be found in all the files, queries need to be changed and modified to collect more data on the specific difference that could be found. Therefore, eight queries (Table 2) examined metadata files.

Table 2: Query syntax examined files to find metadata content

Query ID	Query syntax	Query type of GDPR	Data content	Data Folder	Content Found
1.	‘David Gilmour’	VARCHAR string	Meta Data	ExperimentGroupMeta	Yes
2.	‘david’	VARCHAR string	Meta Data	ExperimentGroupMeta	Yes
3.	‘gmail’	VARCHAR string	Meta Data	ExperimentGroupMeta	Yes
4.	‘front’	VARCHAR string	Meta Data	ExperimentGroupMeta	Yes
5.	‘146’	NUMERIC string	Meta Data	ExperimentGroupMeta	No
6.	’77.146’	NUMERIC string	Meta Data	ExperimentGroupMeta	Yes
7.	‘7.146’	NUMERIC string	Meta Data	ExperimentGroupMeta	No
8.	‘14’	NUMERIC string	Meta Data	ExperimentGroupMeta	No
9.	’07.06.1974’	TS string	Meta Data	ExperimentGroupMeta	Yes
10.	’07-06-1974	TS string	Meta Data	ExperimentGroupMeta	Yes
11.	’07.06’	NUMERIC string	Meta Data	ExperimentGroupMeta	No

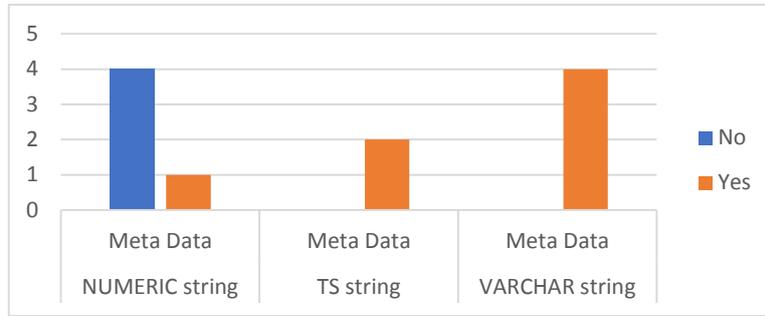


Fig. 2: File share metadata query results found on TS, Varchar and Numeric strings

Data in figure 2 indicates the success rate on tested VARCHAR and Time Stamp (TS) search strings against the NUMERIC strings in the experiments on metadata search queries on the File Share search engine. From 11 files measured VARCHAR search query strings were found in 4 cases, TS strings in 2 cases and in NUMERIC strings 1 case total 7 strings were found and 4 not found as a NUMERIC search query string.

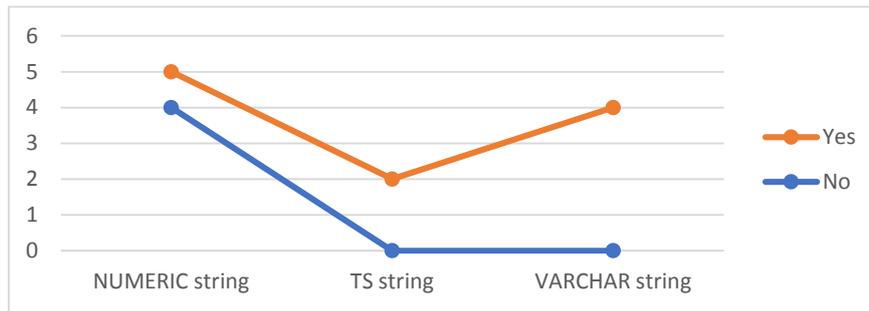


Fig. 3: Query syntax trend analysis on found content through the search engine for metadata

The chart in figure 3 indicates a successful search trend for a selected search query for the examined case in experiments on metadata grouped by the content found in figure 13. Data in the chart from figure 12 indicates the trend on searched queries through various data types.

3.1.2. File Share Search Engine: In the Share points search engine experiments there were six files changed with data content and as per table 3, results of the data content search gave the same results. In cases where experiments covered metadata search within the Share Point repositories following results were captured in table 3.

Table 3: Query syntax examined on files to find metadata content in Share Point

Query ID	Query syntax	Query type of GDPR Data content cases	Data content	Meta Data type	Content Found
1.	'David Gilmour'	VARCHAR string	Meta Data	Author/Title	Yes
2.	'david.gilmour@gmail.com'	VARCHAR string	Meta Data	Other	No
3.	'@gmail'	VARCHAR string	Meta Data	Other	No
4.	'192.168.77.146'	NUMERIC string	Meta Data	Author/Title	Yes
5.	'77'	NUMERIC string	Meta Data	Author/Title	Yes
6.	'07-06-1974'	TS string	Meta Data	Author/Title	No
7.	'6/7/1974'	TS string	Meta Data	Author/Title	Yes
8.	'1974'	NUMERIC string	Meta Data	Other	No
9.	'49, Route de Frontenex, 1207 Geneva'	VARCHAR string	Meta Data	Other	No

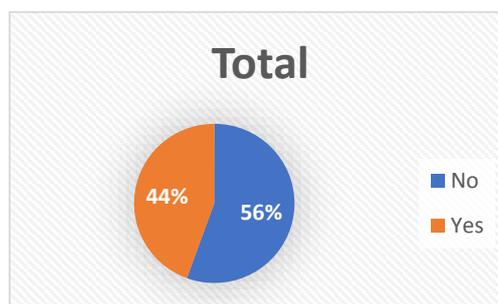


Fig. 41: Ratio between found and non-found files for the metadata content in Share Point repository

Results from figure 4 are indicating that the majority of examined search queries were not successful and the files were not found in 56% of cases where 44% per cent of all sample search queries were found.

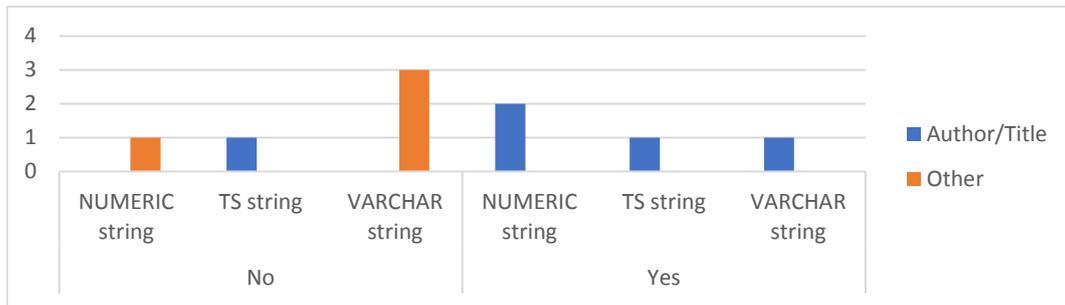


Fig. 5: Share point share metadata query results found by Author/Title metadata or other metadata

Data from the chart presented in figure 5 is indicating dimension on found metadata files based on the populated Author/Title type of metadata for Numeric, Timestamp and Varchar string type with the anomaly of non-found metadata for one-time stamp examined case. From total queries number of queries examined 4 were found from each data type, NUMERIC, TS and VARCHAR conditioned by Author/Title metadata, whilst 5 queries were not found for other data type and one was not found on Author/Title condition. In the experiment, the structure of 3 folders and 15 files were used per repository. The structure is organized purposely to measure the ability of search engine from different dimensions. For the purposes of analyzing overall efficiency based on the stated conditions of the search engines in table 4 is shown files type and folder where the file was found through the search query.

Table 4: File type and data search queries found in both repositories

File ID	File Type	Data found File Share (rf)	Data found Share Point (rs)
1.	MS Word	True	True
2.	MS Excel	False	False
3.	MS PowerPoint	True	True
4.	XML file	True	True
5.	TXT file	True	True
6.	HTML file	True	True

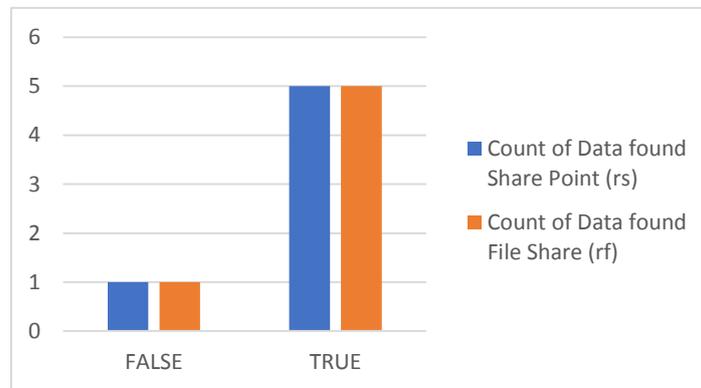


Fig. 6: Files found based on the search results queries in both repositories

Data from figure 6 indicate the ratio between files found in both repositories. Total files examined in the experiment were 6 in each repository. Under circumstances that non-Microsoft editable files (.txt, .xml and .html) cannot add metadata in their file list attribute, which creates false condition, data presented in figure 7 indicates 83% of all the files were found and 17% of files were not found for Data Content search in both repositories.

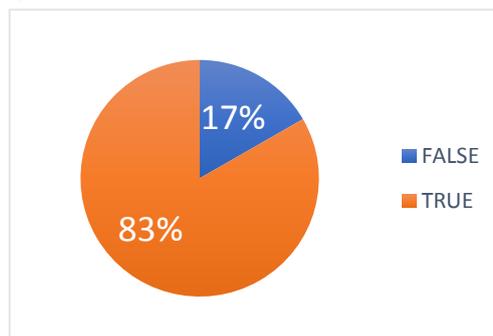


Fig. 7: Files found in all repositories for data queries in all experiments

3.2 Survey results

Questionnaire in the form of the Internet-based survey was conducted from the 1st of March 2018 until the 22nd of March. From 141 invited participants on the 22nd of March 68 participants completed a questionnaire which is 48.2% of invited participants.

Results of the questionnaire can be divided into the following sections:

- Results on participant’s behaviour and usage of search engines and data repositories
- Results on participant’s opinion regarding the GDPR readiness in their organisations
- Results on participant’s personal questions on participants attributes.

72.1% of participants recognised Document collaboration and libraries function of share point platform that is used from all share point platforms functional parts, whilst 35.3% of participants using lists functionality in Share Point platform. 22.1% of participants recognised web site development and content management usage in share point, 17.6% of participants are using surveys functionality and 19.1% of participants are not using any share point functionality (figure 8).

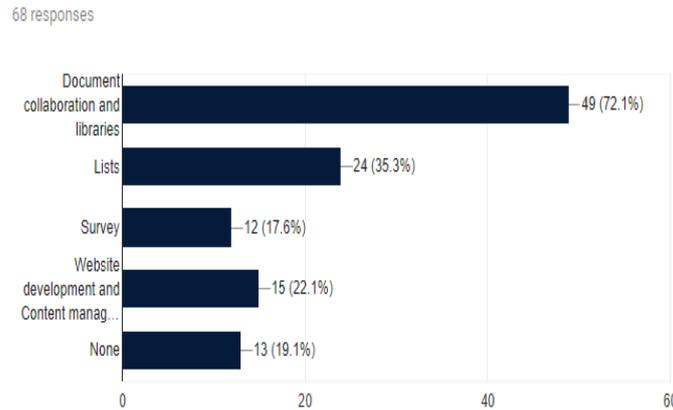


Fig. 8: Questionnaire results on participant’s response on the usage of share point functionalities

Results on how participants find data in repositories (figure 9), 67.6% of all the participants find the data through manual search, 44.1% of participants are using a repository search engine function whilst 16.2% are not using at all Share Point or File Share repositories. 8.8% of all participants are using security and compliance tools to find the data and finally, 4.4% of participants are using third party tools to find the data.

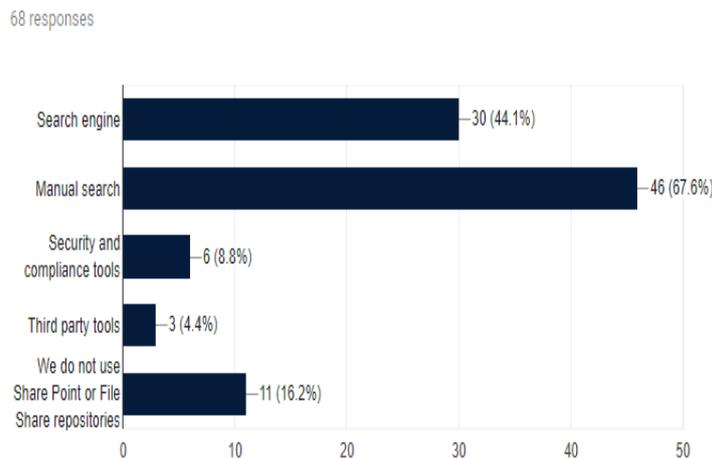


Fig. 9: Questionnaire results on participant’s behaviour on finding data in Share Point and File share repositories

One of the key parameters from the survey is the participant’s usage of Meta tags. Results from figure 10 are indicating that in every category of metatags, majority of the population is not using metatags as a functional element that could enhance search later in search engines. 28 Participants indicated that they are never using Meta tags ‘Comments’ and ‘Title’, 29 participants never used Meta tag ‘Author’, 34 participants never used Meta tag ‘Company’ and in 38 cases participants never used Meta tag ‘Manager’. Meta tags that are used Frequently by participants are ‘Title’ for 14 participants, ‘Author’ in 13 cases, ‘Comments’ and ‘Company’ in 9 cases and ‘Manager’ in 5 cases. Meta tags that are used Occasionally by 12 participants for ‘Comments’, 8 participants used Meta tag ‘Author’, 7 participants used Meta tag ‘Manager’ and only 6 participants declared themselves that were using meta tags ‘Company’ and ‘Title’.

Declared Meta tags that participants used in the category rarely in 18 cases for Meta tag ‘Comment’, in 11 cases for Meta tag ‘Author’. 15 participants expressed themselves that they rarely used Meta tag ‘Company’ and 14 participants used rarely Meta tag ‘Manager’. Only 10 participants used rarely Meta tag ‘Title’. As very frequently participants selected Meta tag ‘Author’ in 7 cases only and in 4 cases for Meta tags ‘Company’ and ‘Manager’. Only one participant selected that was using Meta tag ‘Comment’ very frequently and 10 participants used very frequently Meta tag ‘Title’.

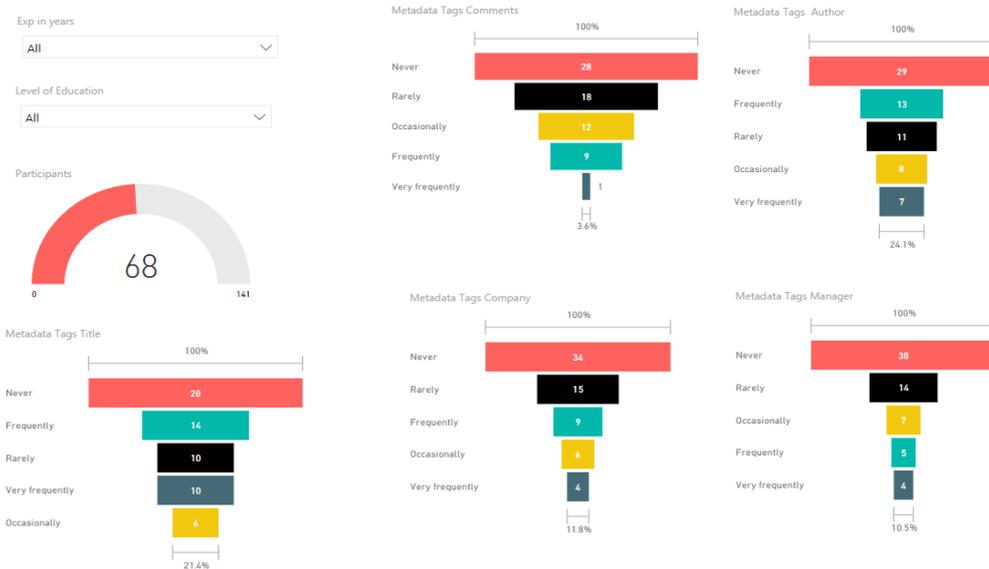


Fig. 10: Questionnaire results on participant's usage of metadata

Results on the usage of the search engine in different IT systems from Figure 11 shows that majority of participants 54.41% responded that are using Google Internet search engine 'very frequently', 32.35% 'frequently' and only 4.41% are using Google search engine 'occasionally', 'rarely' and 'never'. In regards of search engine in manufacturing execution systems (MES), participants chosen in 63.24% that they 'never' used its search engine, 'occasionally' used it in 11.7% of cases, 'rarely' 13.24% of cases, 'frequently' 7.35% and 'very frequently' in 4.41%. In the case of the Share Point search engine, 39.71% of participants 'never' used the search engine, whilst 19.12% of participants were using it 'frequently'. 'Occasionally' participants searched in Share point in 13.24% of cases, 'rarely' in 19.12% and 'very frequently' in 8.82% of cases.

File share search engines are used by participants of the survey in 39.71% 'frequently', 25% 'occasionally', 17.65% 'never', and 8.82% 'rarely' and 'very frequently'. In the case of Active Directory search engine usage, 41.18% of participants 'never' used its search engine, 'occasionally' it was used in 30.88% of all participants whilst 14.71% chosen the option 'rarely'. Only 7.35% of participant's marked 'frequently' usage of Active Directory search engine, whilst 5.88% of participants used it 'very frequently'. Finally, Enterprise Resource Planning System (ERP) and its search engine were 'never' used for 48.53% of participants, 'occasionally' it was used in 17.65% of cases whilst 13.24% of participants marked this usage as 'frequently'. This search engine is used 'rarely' in 11.76% of cases and only 8.82% are using this search engine 'very frequently'.



Fig. 11: Questionnaire results on participant's usage of search engines in data systems

To analyse the trend on participation in a survey figures 12 and 13 show results on the participation per day per opinion on the reasons for GDPR compliance where 15 participants completed the survey before 1st of March 2018. 28 participants finished the survey by the end of the 2nd of March whilst 10 participants finished it until the 3rd of March. From the 3rd of March 2018 until the 10 of March 2018 there were less than 4 participants per day.

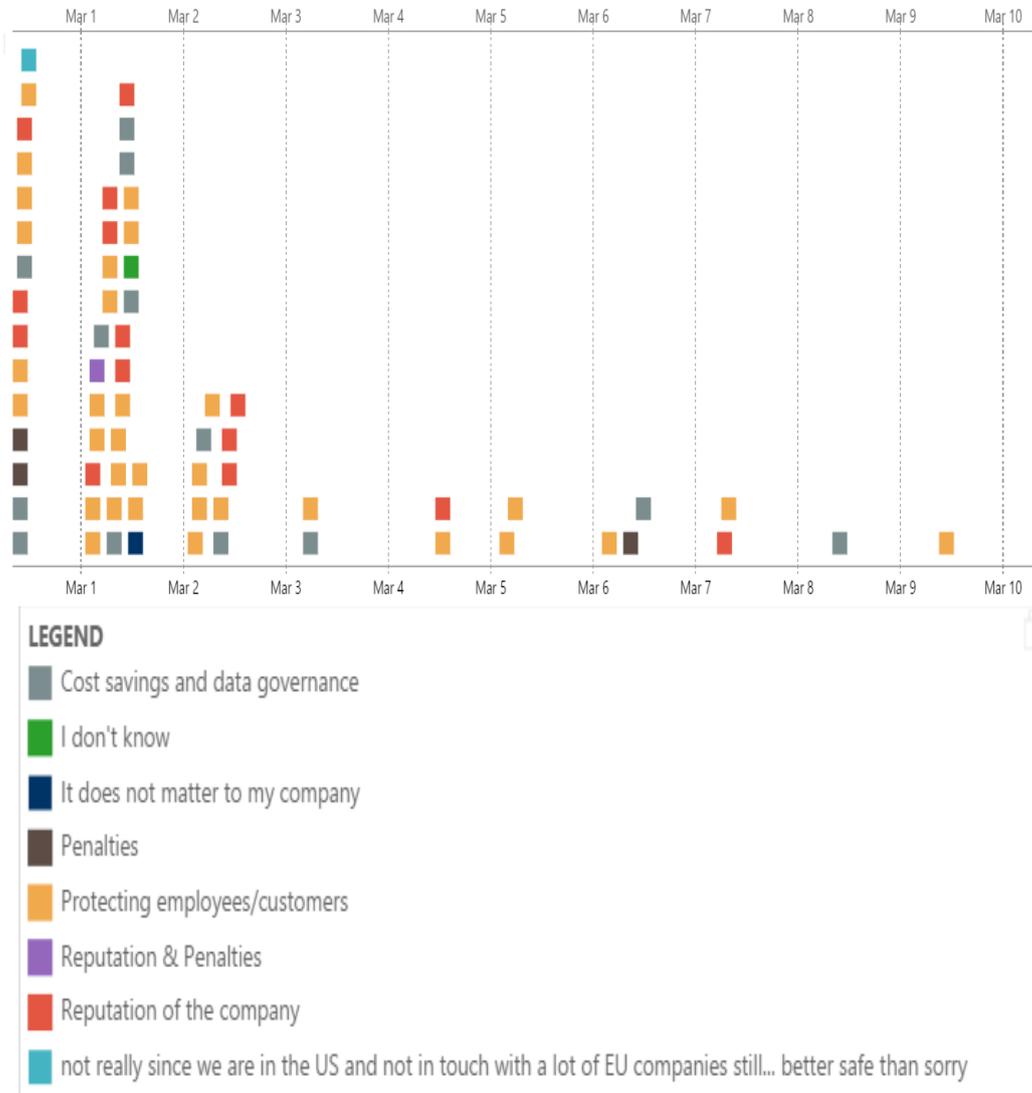


Fig. 12: Questionnaire results on survey’s participation per day through the reasons for GDPR compliance

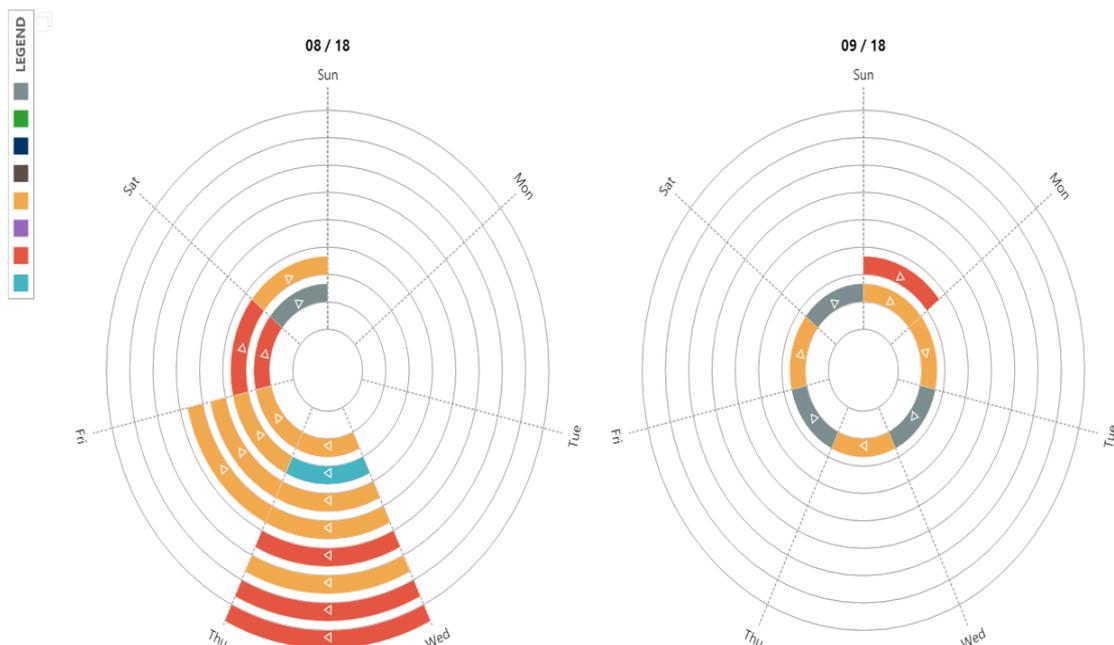


Fig. 13: Questionnaire results on the survey’s participation per day through the reasons for GDPR compliance per week

Results on participant’s level of education per reason of the GDPR compliance shown in figure 14 are indicating that 36 participants of the survey are with Master of Science degree, 6 participants hold PhD degree, 15 participants bachelor’s degree, 9 with a high school degree, 1 with business school degree and 1 with professional qualification.

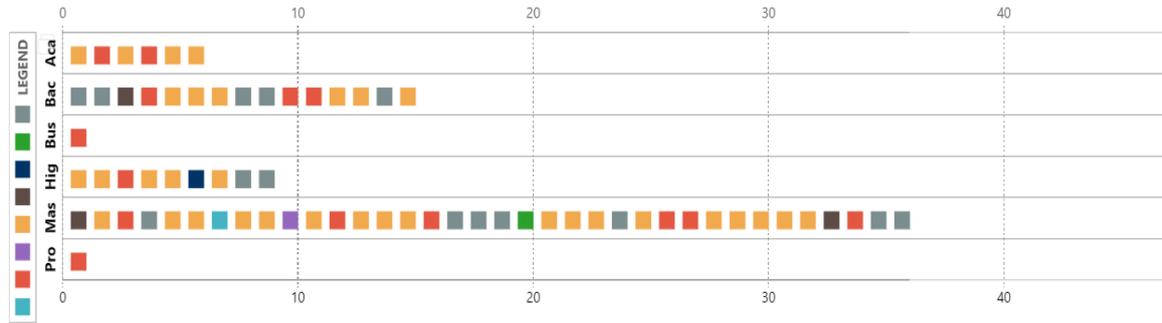


Fig. 14: Questionnaire results on the survey's participation per day through the reasons for GDPR compliance per level of education

In collaborating with its co-workers, results shown in figure 15, participants responded that through Windows File Sharing 29 participants are very frequently 13 participants marked that are frequently collaborating, whilst occasionally option was selected in 13 cases as well. 10 participants selected option never whilst only 3 participants are using Windows File Sharing services rarely. The surprisingly high response of participants 49, who never used Linux File share, 7 participants are using it occasionally, 6 participants responded that are using it rarely, whilst, 5 out of 68 frequently. Only 1 participant is using the Linux File Share very frequently. One drive sharing services were never used by 22 participants of the survey.



Fig. 15: Questionnaire results on participant's usage of different data repositories

Frequently one drive sharing per figure 15, is used by 17 participants, whilst 15 participants chosen option very frequent usage of this service. Only 9 participants marked occasional usage of one drive sharing whilst 5 participants were rarely using it.

One of the important outcomes of the survey was the indicator of how participants are using Share Point services in sharing files and collaboration with others. Surprisingly high number of participants, 21 of them marked that are never used Share Point data repositories in collaboration with other colleagues and co-workers whilst 18 participants selected frequent usage option. 13 participants of the survey chosen occasional usage, 13 participants as well marked this usage as very frequent whilst only 3 users are rarely using it.

Participants were asked as well to rate their overall organisation's readiness for each topic that is regulated through the GDPR. For the Data breach alert, readiness 24 participants answered that they don't know, ready organisations were selected by 16 participants, good rated 12 of participants, whilst 12 selected poor readiness and only 4 of participants selected very good readiness level on Data Breach alert.

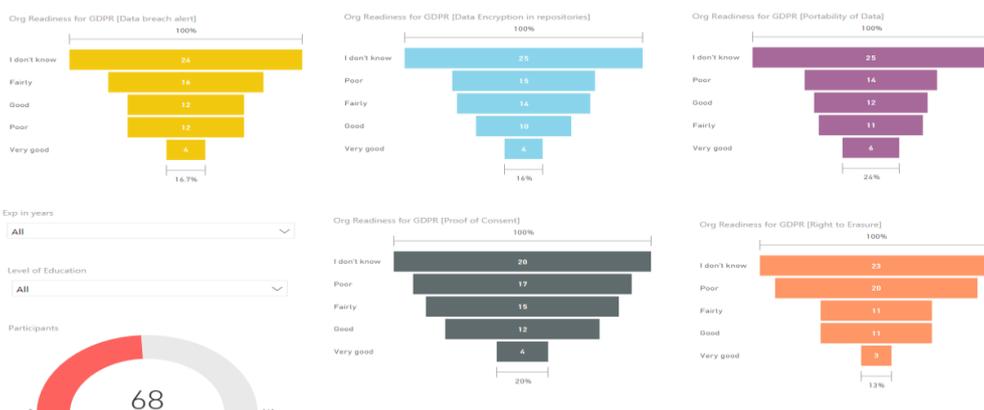


Fig. 16: Questionnaire results on participant's opinion regarding the Organisational readiness for GDPR

Regarding the GDPR topic in figure 16, data encryption in repositories, participants rated their organisation being ready with the answer ‘I don’t know’ by 25 participants, poor readiness by 15 participants whilst only 14 participants chosen fairly readiness for the data encryption in repositories. Only 10 participants marked their organisational readiness as ‘good’ and 4 participants marked as ‘very good’.

A very similar outcome is obtained for the topic Portability of data where 25 participants did not know their organisation’s level of readiness and 14 of participants marked that the readiness level is poor in their organisations. 12 participants chosen ‘good’ level of readiness, ‘fairly’ level of readiness was selected in 11 cases whilst only 6 participants chosen ‘very good’ level of readiness on the topic Portability of Data.

Proof of consent and right to erasure parts of the GDPR have the lowest rates on the organisation’s readiness answered by participants. Only 12 participants rated ‘good’ for ‘proof of concept’ and 4 participants as ‘very good’ on the same topic, whilst 11 participants rated ‘good’ for ‘right to erasure’ and 4 as ‘very good’. Majority of participants did not know on both topics 20 organisation’s readiness for the topic ‘Proof of Consent’ whilst 23 participants did not know readiness level for ‘Right to erasure’ topic of the GDPR. 17 participants marked ‘poor’ and 15 marked ‘fairly’ readiness on ‘Proof of consent’ topic whilst 20 participants marked ‘poor’ and 11 marked ‘fairly’ readiness on ‘Right to erasure’.

Results regarding the participants’ opinion (figure 17) on the reasons why would the organization need to be compliant with GDPR, majority of participants 33 of them answered that protection of employees or customers is the main reason whilst ‘Cost savings and data governance’ was chosen by 14 participants as well as protecting ‘Reputation of the company’ selected by 14 participants. Only 3 participants selected penalties as the main reason to be compliant with GDPR. Remaining 4 responses were individual opinions without a clear indication to group them together.

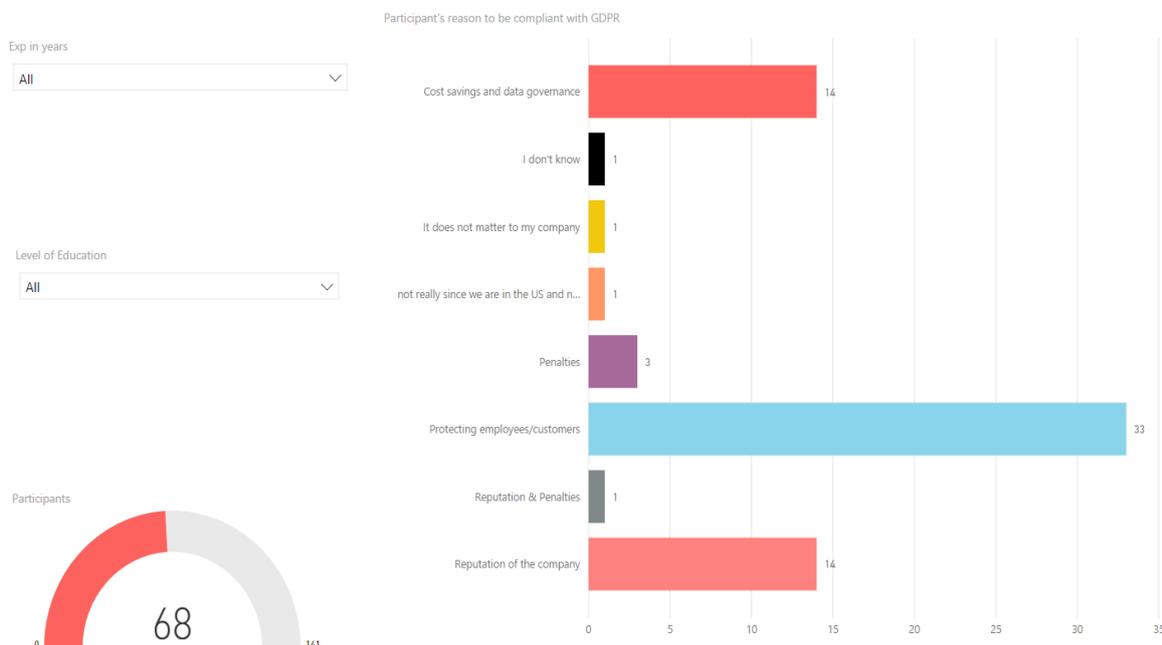


Fig. 17: Questionnaire results on participant’s opinion on reasons to be compliant with GDPR

4. DISCUSSION

Evidence from the experiments demonstrated that under three conditions hypothesis can be rejected:

Condition 1

Findings on metadata numeric string search in File Share repository is that every numeric string metadata content search is conditioned to be found if the numeric string of the search data includes the ninth character at least, formulated in condition: $(\alpha \rightarrow \beta) \wedge (\neg \alpha \rightarrow \mu)$.

This construct contains the following:

- μ - Axiom for negative meta data search results in File Share
- α - Axiom for a ninth character in numeric search string included
- β - Axiom for positive meta data search result in File Share

Condition 2

Findings on metadata in Share Point search engine variations are formulated in condition: $(\gamma \rightarrow \delta) \wedge (\neg \gamma \rightarrow \theta)$.

This construct contains the following:

- θ - Axiom for unsuccessful metadata search
- γ - Axiom for metadata populated in sections of Author/Title
- δ - Axiom for Successful metadata search

Condition 3

Experiments results from the chapter 5.2 are indicating that all the search queries were found based on the condition that the metadata are populated either in Author or/and Title metadata section otherwise it can't be found. One case of Author/Metadata was not found based on the condition FDCC = FDCI.

File Data Content Converted time stamp (FDCC) = 6/7/1974

File Data Content Inserted time stamp (FDCI) = 07-06-1974

FDCC ≠ FDCI

Results and data obtained through the experiments through out-of-the-box search engine configuration in a File share and Share Point data repositories and their ability to find data were supported by the results of the survey. The objective of the research is to find the links between survey where there are indicators on the reliability of the structured data within the data repositories and participants behavior regarding the data and metadata management in order to investigate the ability of Share Point and File Share Search Engines to ensure compliance to GDPR regulations.

Evidence from the questionnaire is grouped by the following variables supporting the experiment's findings in:

- Usage of search engine
- Usage of metadata
- Usage of search engines (Windows File Share and Share Point)
- Usage of File repositories (Windows File Share and Share Point)

Search engines are used by the half of asked participants and half or participants are still manually searching for data. Metadata are not widely used and only half or participants are using any form of metadata.

Both Windows File share and Share point search engine are used by majority of participants and both repositories are the most used by more than 2/3 or participants.

Authors should discuss the results and how they can be interpreted in perspective of previous studies and of the working hypotheses. The findings and their implications should be discussed in the broadest context possible. Future research directions may also be highlighted.

5. CONCLUSION

Through evidence presented derived from experiments and questionnaire, it emanates that personal data files in files and its metadata cannot be found in all cases in Windows File Share and Share point repositories, hence improvements of search engines or enhancements of operational manipulation with data and metadata in files must be performed based on conditions stated in research conclusions. Evidence from the results of the questionnaire are partially validating experiment's effort related to the need for a metadata search and search engine usage versus manual search option, nonetheless, they are validating fully experiments effort related to use of both Windows File Share and Share Point search engines and usage of their repositories.

- **Supplementary Materials:** The following are available online at www.mdpi.com/xxx/s1, Figure S1: title, Table S1: title, Video S1: title.
- **Author Contributions:** This research was submitted and rewarded for the Master of Science degree at Leeds Beckett University.
- **Funding:** This research received no external funding.
- **Acknowledgements:** The author thanks, M. Sc. Radmila Velichkovic, M. Sc. Deborah Sebold Thomas and Dr Philippa Trevor for their kind help and discussions.
- **Conflicts of Interest:** The authors declare no conflict of interest.

6. REFERENCES

- [1] Lynskey, O. (2015) "The Foundations of EU Data Protection Law", Oxford: Oxford University Press.
- [2] Kooops, B., J. (2011) "Forgetting Footprints, Shunning Shadows. A critical analysis of the "Right to Be Forgotten" in Big data Practice", Scripted, Volume8, Issue3, December 2011, Available at <https://script-ed.org/wp-content/uploads/2011/12/kooops.pdf>. [Accessed, 16/11/2017]
- [3] Bitar, J., Jakobsson, B. (2017) "GDPR: Securing Personal Data in Compliance with new EU Regulations", M.Sc. thesis University of Lulea: Luleå tekniska universitet.
- [4] Lewis, D., Moorkens, J., Fatema, K. (2017) "Integrating the Management of Personal Data Protection and Open Science with Research Ethics", Proceedings of the First Workshop on Ethics in Natural Language Processing, pages 60–65, Valencia, Spain, April 4th, 2017 Association for Computational Linguistics [Accessed, 16/11/2017] <http://www.aclweb.org/anthology/W17-1607>
- [5] Billgren, P., Ekman, L. W. (2017) "Compliance Challenges with the General Data Protection Regulation", M.Sc. thesis University of Lund: Dept. of Informatics, Lund University School of Economics and Management
- [6] Kooops, B., J. Leenes, R. (2013) "Privacy regulation cannot be hardcoded. A critical comment on the 'privacy by design' provision in data-protection law", International Review of Law Computers & Technology · July 2013
- [7] Ford, D. T., Qamar, S. (2017) "Seeking opportunities in the Internet of Things (IoT): A Study of IT values co-creation in the IoT ecosystem while considering the potential impacts of the EU General Data Protection Regulations (GDPR)", M.Sc. thesis University of Umea: Department of Informatics
- [8] Hornung, G. (2012) "A General Data Protection Regulation for Europe? Light and shade in the commission's draft of 25th January 2012", [Accessed, 17/11/2017] <https://script-ed.org/wp-content/uploads/2012/05/hornung.pdf> Appendix
- [9] Varonis Systems, Inc (2017) "One Year Out: Views on GDPR", [Accessed, 16/11/2017] <https://info.varonis.com/hubfs/docs/2017-GDPR-survey-results.pdf>

- [10] Gartner (2017) "Healthcare Providers vertical industry comparison", [Accessed, 16/11/2017] http://www.gartner.com/downloads/public/explore/metricsAndTools/ITBudget_Sample_2012.pdf
- [11] Roode, M. D. (2016) "Privacy Enhancing Technologies", Research Gate, May 2016 [Accessed, 16/11/2017] https://www.researchgate.net/profile/Mike_De_Roode/publication/301350955_Privacy_Enhancing_Technologies_A_software_engineering_approach_to_design_PETs/links/5714dc5208aebda86c0d2485.pdf
- [12] Mudassar, A., (2013) "SharePoint Governance", M.Sc. thesis University of Agder
- [13] Horn, R. (2009) "Researching and Writing Dissertations", London: Chartered Institute of Personnel and Development.
- [14] The Economist (2017) "Data is giving rise to a new economy" 6th May 2017, Available from <https://www.economist.com/news/briefing/21721634-how-it-shaping-up-data-giving-rise-new-economy> [Accessed: 08 January 2017]
- [15] The Economist (2010) "Data, data everywhere" 25th Feb 2010, Available from <http://www.economist.com/node/15557443> [Accessed: 20 January 2017]
- [16] Moller, C. (2017) "The Evolution of Data Protection and Privacy in the Public Security Context - An Institutional Analysis of Three EU Data Retention and Access Regimes", UK: Queen Mary, University of London
- [17] Colburn, T. (2000) "Philosophy and Computer Science", London: Routledge Taylor & Francis Group
- [18] Saunders, M., Lewis, P. and Thornhill, A. (2009) "Research methods for business students", 5th Edition Essex: Pearson Education
- [19] Saunders, M., Lewis, P. and Thornhill, A. (2009) "Research methods for business students", 5th Edition Essex: Pearson Education