Risk Assessment in Scientific Data Management using the Delphi method

Maria Beatriz Ferraz Cunha

Department of Engineering and Management, Instituto Superior Técnico

Abstract

The need to manage scientific data has become increasingly evident as data are increasingly being sought to be reused in the long term. Scientific Data Management must be applied throughout the scientific data lifecycle, including the creation, processing, analysis, preservation and re-use of data, as well as the access to data. The data management process should result in a dynamic document called the Data Management Plan. This document should countain the answers to some questions regarding with the identification of the nature of scientific data, how scientific data should be shared, how data will be accessed and how they will be archived. All these activities have uncertainties associated, so it is necessary to apply a risk management. The work to be carried out involves the analysis and study of the literature and some existing international standards, with a view to developing a method that is simple and quick to apply and that allows to apply a risk management in the domain of data management.

Key words: Risk Management, Risk Factors, Scientific Data Management, Data Management Plans, Delphi Method, Q-sort, Risk Matrices.

1. Introduction

Increasingly, projects (research or engineering) are oriented to scientific data. Scientific data are objects, usually in the digital format, used as a working tool for some researchers and accepted by the scientific community as a necessary tool for the validation of scientific research.

The objectives of research projects in the context of scientific data management may not always be the same. Thus, it is possible to define two "domains" of the problem: the first "domain" corresponds to the projects dedicated to the production and use of scientific data and the second "domain" refers to the projects dedicated to the management of scientific data. Whatever the purpose of the project, there should always be done a scientific data management that must be complemented with a Data Management Plan. A DMP is a dynamic document that, frequently, should be reviewed and updated.

In this document there should be answers to some questions, such as the identification of the nature of scientific data, how data will be shared, how data will be given and who should have it, and how the data will be archived. These are activities that have some associated uncertainty, as there may be interferences (risk factors) that affect their results.

One of the main objectives of scientific data management is to protect them so that they can be used in the long term. On the other hand, risk management is applied when it is intended to protect certain assets from certain risks. Thus, considering the scientific data as assets of a project, it can be concluded that, from this point of view, data management and risk management share the same objective regarding data protection. It is therefore clear that the management of scientific data must be complemented and accompanied by risk management. In order to study how the scientific data management can be complemented with risk management, two standards were analyzed: ISO 31000: 2009 and ISO 31010:2009. From these standards, it was possible to know the risk management process, as well as its activities. The analysis of these standards also allowed to acquire some knowledge regarding the existing techniques that can be applied in the diverse activities of the process of risk management. Based on these standards, a method that is considered ideal for the conduct of risk management in the context of scientific data management is suggested and it's presented in this paper.

2. Literature Review

2.1. Scientific Data Management

Before the evolution of the technology, the results of the scientific investigations (generated from scientific data) were shared by physical means, like books, articles or magazines. By this mean, the knowledge was transmitted from generation to generation. With the technological advances, the most used became the digital format so, nowadays, scientific data are stored in computers and shared in a digital way. Digital scientific data are objects, in a digital format, used by the investigators as a work tool. These data are accepted by the scientific community as a necessary tool for the validation of scientific researches. Sharing the data in a digital way provides a faster dissemination of the data.

With the concept of scientific data in mind the question that we should answer is: how can we guarantee that the scientific data are available for the future generations without suffering modifications or losses? Scientific data need to be created, stored, archived, preserved and should always be available for future use (Fernandes et al, 2012). Considering these requirements, a scientific data management should be applied. A scientific data management is required when it's necessary to support the production of high quality data, increase the exposure of the research and protect the data to avoid the lost of them or bad utilization. (Fernandes et al, 2012). This process includes some of the following activities: data organization in repositories, promote the data access, prevent losses and data modifications, guarantee the security of confidential data and cooperate with research communities in data creation and utilization.

Scientific data management should be done during the all scientific data "lifecycle" – data curation – which envolves the maintenance, preservation and value adding to scientific data all over its life (Figure 1.1).

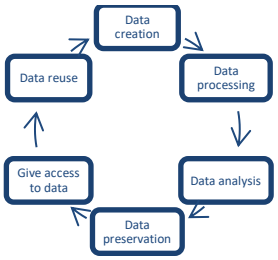

Figure 1.1: Scientific Data Lifecycle

According to DCC (Digital Curation Center) one of the main activities of data curation is adding value to data. It's possible to add value to data by adding information about them like the creation context, the reason for the creation, the way of how data was created and what alterations the data suffered during them lifecycle (Abbott, 2008). These informations must be documented in Data Management Plans and Metadata.

2.2. Data Management Plans and Metadata

As we said before, suplementar information about data increases its value and makes easier the data management process. This information can be written in Data Management Plans (DMP) or in the Metadata. A DMP is a dynamic document and the main goal of this document is to protect digital data from potential threats that can affect the projects. By this reason, nowadays, it's required for scientific communities, by the funding agencies, that in each project has to be defined and apllied a DMP and this step has to be included in the apresentation of a proposal for a project. A DMP describes the produced data, the standards used to describe data (metadata), the entities responsible for data, the sharing and accessing policies and what are the devices/ equipments needed to guarantee data sharing, archiving and preservation. According to DCC (DCC, 2013), which is an internationally known center in data curation area, the elements of a DMP are:

- Administrative data: ID, funding agency, project name, project description, name, investigator ID, first and last version dates and information about the related policies;
- Created data: type of data that are going to be created/collected and how data is going to be created/collected;
- Metadata: metadata description;
- Ethical and legal compliance: information about how the legal and ethical issues are going to be solved;
- Data archiving and security copies: information about how data is going to be archived and kept in security during and after the research;
- Data selection and preservation: which data should be managed and preserved – definition of the long-term data preservation plan;
- Data sharing: Defining how data is going to be shared and which constraints are;
- Responsabilities and resources: Defining who is going to be responsible for data management and what resource is going to be needed.

There are some principles, denominated by FAIR principles, that should be applied in a DMP execution. According to these principles, data need to be

Findable, Accessible, Interoperable and Re-usable (Mark Wilkinson, 2016).[1]

Another document that is very important is metadata. Metadata are informations that describe some data resources, for example, date of data creation, data description, format, title and informations about data producers. According to DCMI (Dublin Core Metadata Initiative), metadata are "data about data" used to described physical and digital data.

2.3. OAIS Model

In order to support the data management process, it was developed a reference model called Open Archival Information System (OAIS), in 2003. This model is also known as an international standard ISO 14721:2003. The main goal of this model is to establish a system where each organization entity knows its responsibility related with data creation, sharing and utilization (Lavoie,2014). According to this model, the entities that should be present in scientific data management process are: data producer, data consumer and respository manager (Figura 1.2).

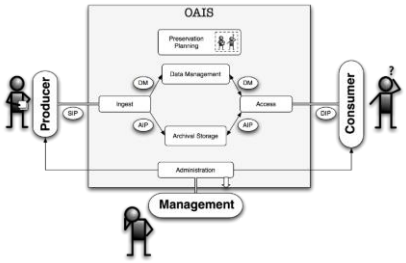

Figure 1.2: OAIS model

The activities that should exist in this system are: the deposit of data in the repository, data storage, data management, all system administration, data preservation and access planning (Arellano, 2008). According to this model, the data producer is responsible for the deposit of data in the repository, repository manager is responsible for data storage, data management, administration of the entire system, data preservation and access planning, and the data consumer is responsible for accessing and reuse data. All these activities mentioned in OAIS model are subject to risks because they are subject to uncertainties. This is a reason to explore the risk and risk management concepts.

2.4. Risk Concept

The research units are more concerned about the risk factors that can affect their projects. However, before beginning the risks vigilance it's important to exactly know the meaning of risk. Some years ago, the risk was known as a danger situation that could occur any time because it was an act of god (Mendes, 2002). These events were seeing as natural events, so the only way to reduce its impact were estimating when these events would happen and trying, somehow, avoid them. Since the population started to think about the future and since they felt the need of explain and justify these uncertain events, the risk concept has been developed (Mendes, 2002).

According to the ISO 31000:2009 standard – Risk Management (principles and guidelines) – the risk can be defined as the "effect of the uncertainty in the objectives". Risk can also be defined as a potential factor for the realization of negative and undesirable consequences in an event (Rowe,1988). Another definition for risk is suggested by The Royal Society (1992), and according to this author the risk can be defined as the probability of a certain adverse event happen during a given period of time. There is another risk definition that is more mathematical and according to this definition, the risk of an event can be calculated by the multiplication of the probability and impact values (Cox, 2009).

With all these definitions in mind, we can say that the risk concept has three concepts associated with it: probability, uncertainty and impact (Pickering et al, 2010). The probability is the possibility of some event happen. When it's not possible to exactly predict the result of an event, we can say that we are in a uncertain situation. The impact is related with the positive and negative effects that comes from an event. To better control the risks, it's important to execute a risk management.

2.5. Risk Management

A risk management is a continuous process, composed by a set of principles and supported in an appropriated structure to the research unit and its external environment. This process consists in the risks assessment and its useful to decide which steps should be taken to reduce or eliminate the risks and which are the control measures that should be applied. This process should be seen as a priority for the organization because a good risk management increases the probability for the company to be well succeded and decreases the probability of failure and the level of uncertainty associated with the results of the organization.

[1] http://datafairport.org/fair-principles-living-document-menu

There are some international standards, for example, ISO 31000:2009 –Risk Management (principles and guidelines) – and ISO 31010:2009 – Risk Management (risk assessment techniques) – that support the risk management process. The main goal of the first standard (ISO 31000:2009) is to provide a set of principles and guidelines that could be used by the organizations and can make the risk management process more efficient. The guidelines provided by this standard don't promote the uniformization of the risk management process in the organization. Each organization should adjust its risk management process considering the context of the organization. According to ISO 31000:2009, the risk management process can be executed based on the framework presented in Figure 1.3. This process envolves five activities: establishing the context, risk assessment, risk treatment, communication and consultation and monitoring and review. Risk assessment includes three activities that are risk identification, risk analysis and risk evaluation. The second standard mencioned before is ISO 31010:2009. This international standard is used as a support for ISO 31000:2009 and provides some practices on selection and application of systematic techniques for risk assessment.

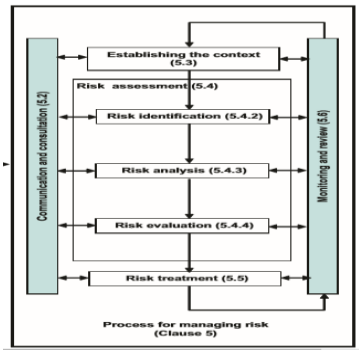

Figure 1.3: Risk management process (adapted from: ISO 31000:2009)

2.5. Risk Management in scientific data management

As it was said before, the scientific data management concept has been increasingly used and as a consequence of this fact, the utilization of data management plans is also more frequent. Since there is some uncertainty associated to the context of data management, it's necessary to apply a risk management. Risk management can complement the execution of a DMP because both areas share the same objective that is the data protection. However, if the scientific communities had to apply all the guidances of the international standards, mencioned before, this process would be very complex and it would take a lot of time. In this paper it's presented an easy way of complementing the risk management process with DMP, using the Delphi method, Q-sort method and other techniques also mentioned in the following chapters.

Another International standard that was analysed was the ISO 27005:2011 – Information Security Risk Management - that provides some guidelines for information security risk management. ISO 27005:2011 provides some examples of possible threats and vulnerabilities that can affect the information security.

Considering that the asset of this standard (information security) can be compared with the asset of this study (data security), this standard was very important and relevant because it was the basis for the risk factors identification activity. Some threats considered important were related with privacy, invasion, lack of motivation, cyber attacks, natural disasters, problems in archiving, technological problems, and others.

3. Solution hypothesis and proposed method

3.1. Methods for risk factors identification and evaluation

According to ISO 31010:2009, analysed before, there are a lot of tools/ methods that can be used in risk evaluation. However, there are some methods that are more frequently and commonly used, and those were analysed as solution hypothesis. For the risk identification activity, the most used methods are the Brainstorming and the Delphi method.

3.1.1. Brainstorming: risk identification

The brainstorming method is used when it's necessary to reach a conclusion for a specific problem in the shortest time possible, and for this is generated a spontaneous list of ideas given by the group of participants (Diehl et al, 1991). This technique has some strengths, like the implementation of the technique is fast, it's easy to obtain new ideas and solutions, the costs with equipments are very low or don't exist, this technique can be used in many different areas, it allows the creativity development and also the capacity of team working. On the other side some weaknesses of this technique are that it requires presential meetings, it requires an appropriated space for the implementation, the transportation costs can be high and it can occur personalities dominance. One version of this technique is the Brainwriting, developed by Bernd Rohrbach. With this technique it is given the

opportunity of sharing some ideas but using a paper. Although there still is a weakness that consists in the fact that both techniques require the physical presence of the participants.

3.1.2. Delphi method: risk identification

In a post-war movement in the 1960s, Dalkey and Helmer (researchers at the Rand Corporation) began a study whose goal was to develop a technique to predict the effect of technology development on military projects, drawing on the opinion of a number of experts (Dalkey et al, 1963). According to this study, when didn't exist an evidence basis about an issue, one way to achieve it was collecting and synthesizing the specialists' opinions (Dalkey, 1963). The result of this study was the development of the Delphi method that establishes three basic conditions: the anonymity of the participants in the study, the statistical representation of the results and the constant revelation of the expert panel feedback, from round to round (Rozados, 2004).

The consensus among the experts is constructed with the aid of a set of questionnaries, conducted in several rounds, that allow the collection of data on a given domain (Dalkey et al, 1963). The questionnaires should be prepared with a certain rigor in order to avoid ambiguities. This technique has some strengths, for example, it avoids bias of responses by dominance of personalities, it ensures the anonymity of participants, the number of factors considered by a group is greater, it allows an intense and interactive connection with experts and enrichs knowledge of both parties. On the other hand, the weaknesses of this technique are that the results are as valid as the experts' answers, there is the possibility of a misinterpretation of the questionnaire, there is a dependence on third party response, there is a high consumption of time and it requires high written communication skills of the researcher.

3.1.3. Risk Matrices: risk evaluation

A risk matrix can be defined as a mechanism used to characterized and classify the identified risks (Markowski et al, 2008). Before using a risk matrix, it is important that all the risks were previously identified using, for example, the methods explained before. A risk matrix is characterized by a set of cells, represented in lines and columns, where to each cell exists a determined risk level associated. Normally, risk matrices presented colors to distinguish different levels of risk. Green cells represent a low risk level, the yellow color represents a medium risk level and the red cells represent a high risk level.

3.2. Proposed Method

According to the international standard ISO 31000:2009, studied before, the first activity of the risk management process should be the definition of the context, that is explained in the first and second chapters of this paper. Then it's necessary to identify the risks, and for this activity, two techniques were studied. Analysing the strengths and weaknesses of the techniques mencioned before, it is possible to say that any of these techniques require some prior knowledge of the participants about the subject. Because this topic is not yet widely spoken in the literature and because the time for conducting this research was limited to face-to-face meetings, the method considered to be applied in this study was the Delphi method. As previously mentioned, this method corresponds to a communication process that aims at reaching a set of convergent opinions on a real world issue (Hsu et al, 2007), having the advantage that doesn't need personal meeting to obtain results. On the other hand, the Delphi method is the ideal method that extracts and maximizes the benefits of group-based expert methods while eliminating or minimizing its drawbacks. Analysing the advantages and disadvantages of this method, it was considered that this is a good method for risks identification. For risks evaluation, the most used technique is risk matrix, and this technique was used in this investigation, in a theorethical way.

4. Method application

As it was said before, the considered method for risks identification was the Delphi method. The first step to apply this method is to justify why this method is going to be used to solve the problem. The desire of obtain feedback from many specialists of different areas with different experiences, and from different parts of the world and the need to guarantee the anonymity of the experts, led to the selection of this method.

4.1. Delphi's questionnaire structure

The second step consists in the definition of the questionnaire structure. According to Konow and Pérez (1990) it's important to have clear questions, a questionnaire must not be very extense and it's very important the presentation of the concepts. There are two approaches to apply this method: using the "classical" approach or using the "modified" approach. In the first one, it's presented to the specialist an open question, where they can write their opinions about the theme. On the other hand, in the "modified" approach, in the first round of the Delphi method should be presented a list, constructed

before, with all the relevant factors for the study. In this survey, it was decided to use the "modified" approach so the construction of an initial list with some risk factors was needed. This list was generated using the knowledge acquired with the literature review. To identify the risk factors, it was important to clarify that, using the OAIS model as a reference, three types of entities can be defined: Data Producers (DP), Repository Managers (RM) and Data Consumers (DC). According to each entity and the associated type of data (original data, archived data or accessed data) there are different risk factors associated. Original data are newly created data by a DP. Archived Data are original data that are stored in a data repository. These archived data are managed by the RM. Accessed data are data taken from a repository after its acess.

By the analysis of the literature, the risk factors identified were:

- (RM) Loss of Archived Data due to technological obsolescence;
- (RM) Loss of Archived Data due to component faults;
- (RM) Loss of Archived Data due to technological faults;
- (RM) Loss of Metadata about Archived Data;
- (RM) Outdated Metadata about Archived Data;
- (RM) Unauthorized Data Access;
- (RM) Improper use of Archived Data;
- (RM) Wrong interpretation of the Data; Management Plan or Metadata;
- (RM) Data embargo is violated;
- (RM) Loss of Archived Data due to Organizational failure;
- (RM) Destruction of Archival Data due to an attack;
- (RM) Destruction of Archived Data due to a natural disaster;
- (RM) Loss of Data Archiving due to the lack of competencies of the Repository staff;
- (DP) Loss of Original Data due to component faults;
- (DP) Loss of Original Data due to an human error;
- (DP) Destruction of Original Data due to an attack;
- (DP) Original Data with confidentiality requirements not identified;
- (DP) Destruction of Original Data due to a natural disaster;
- (DP) Insufficient Data Management Plan;
- (DC) Insufficient or Incomplete Metadata about Accessed Data;
- (DC) Accessed Data is not trustworthy;
- (DC) Inaccurate Metadata about Accessed Data;
- (DC) Non understandable Metadata about Accessed Data;
- (DC) Plagiarism in the reuse of Accessed Data;
- (DC) Outdated Accessed Data;
- (DC) Accessed Data unusable;
- (DC) Violation of terms for Data Reuse.

There are some platforms that support the online implementation of the Delphi method, for example, the Delphi Decision Aid and e-Delphi. However, one of the goals of this investigation was to serve as a test of a platform that had been developed at the same time of this investigation execution. The used platform is Decspace. The main goal of this platform is to present itself a solution for the problem of the lack of an easy-to-use framework. It was developed to make many MCDA (multicriteria methods) available and to make it possible to add more methods without programming effort (Costa et al, 2017). For the application of the Delphi method it was used the Inquiry method that combines Delphi with the Q-sort method. Q-sort method is used in the study of qualitative issues, but the distinguishing feature of this method is to ask the experts to order the presented factors, according to the distribution provided. The use of the Decspace served essentially as a test, and as some weaknesses were detected, improvements were suggested.

4.2. Experts identification and selection

After the definition of the questionnaire structure and after defining why this method is going to be used, the third step is considered one of the most important steps in a Delphi study, because the characteristics of the experts significativately influenciate the quality and confidence in the results (Powell, 2003). It was not defined a maximum or minimum number of experts, so the number of participants in this study was dependent on the number of contacts done and the receptivity to them. In the beginning of the study, the questionnaire was just sent to a Portuguese community of investigarors, professors and responsibles for some universities documentation systems. Some invitations were sent to professors from Instituto Superior Técnico, Universidade de Évora, Politécnico de Beja, Universidade do Minho, and others, and also to some data repositories.

4.3. Assessment of consensus and stopping criteria

In a Delphi study it's important to define which are the main goals in order to determine the level of study stopping. It was initially established that the stopping criteria would be based on the level of consensus of the participants and the maximum number of rounds. There are some statistical measures, like the average value, the standard deviation and the Kendall's coefficient, that can be used to evaluate the consensus level and evaluate when the Delphi study should stop.

5. Discussion of the results

5.1. Round 1

The first round of this study started on 27/08/2017 and finished on 25/09/2017. As it was said before, first the questionnaire was just sent to a Portuguese community but then, during the first round, because the number of answers were low, it was decided to extend the questionnaire to foreign communities like DCC (Digital Curation Center) and RDA (Research Data Alliance). The number of answers in the final of the first round were eleven, which is a very low number of answers for a delphi study and according to the expectations. Considering the answers of these experts to the characterization questions it's possible to conclude that the majority of them worked in universities in different areas like engineering, physics, information science, decision analysis, biology, computer science, data science and molecular biology. It was also asked about the role of the specialist in data domain, and according to their answers the majority is dedicated to data producing.

About the results, the main goal of this round was to obtain an ordinated list according to the importance, in the experts' opinion, of the presented risk factors. The experts had the opportunity to classify each risk factor in "important" (value 1), "neutral" (value 0) and "not important" (value -1). Also in this round the experts could give some suggestions of risk factors that they considered important to be in the list. The only suggested that was given was: "(DP) Non-standard or incomplete Metadata: Some metadata has to be added by the original author, e.g. subject matter, and if this is not done, or done in a non-standard way, retrieval can be difficult". This suggestion was analysed and it was concluded that it's very similar to another that already was in the list, so it was not added to the list. The first round of this Delphi study was essentially useful to have a sense of what items should be kept on the list of risk factors because they are effectively considered important, and which ones could be eliminated. However, given the low response rate, it was not possible to effectively exclude any of the risk factors. The first round therefore served primarily as a first approach to experts, leading them to question the importance of risk identification in the field of scientific data management. This being a topic that has not yet been approached, with this first round it was possible to introduce these new concepts to the experts so that they could increase their knowledge about this problem.

5.2 Round 2

After finished the first round, the second one was started. With this second round the main goal was to go a little deeper into the question and obtain the classification of risk factors through different scores, thus, it is possible to identify the true order of classification of risk factors. For this round, the Decspace "default" scale was used. In this second round, it is possible to observe the score that each risk factor obtained in the first round, as it is presented at the beginning of each risk factor, for example: (7) (RM) Loss of Archived Data due to the lack of competencies of the repository staff. Given the low response rate of the first round, it was considered relevant not only to invite the experts who responded to the first round to this second round, but also to send the study to other experts. This second round started on 25/09/2017 and finished on 08/10/2017. In this round the number of obtained answers were just eight (and only seven of the eleven experts of the first round answered to this second round). The results of the second round are presented below (Figure 1.4):

Round 2	*Risk Factor*	*Identifier*
14	(RM) Loss of Archived Data due to the lack of competencies of the repository staff	A
13	(DC) Insufficient or incomplete metadata about Accessed Data	B
8	(RM) Unauthorized Data Access	C
7	(DP) Insufficient Data Management Plan	D
6	(DC) Inaccurate metadata about Accessed Data	E
6	(RM) Loss of metadata about Archived Data	F
5	(RM) Loss of Archived Data due to technological obsolescence	G
5	(DP) Original Data with confidentiality requirements not identified	H
3	(RM) Loss of Archived Data due to component faults	I
3	(RM) Loss of Archived Data due to technological faults	J
3	(DP) Loss of Original Data due to an human error	K
3	(DC) Non understandable metadata about Accessed Data	L
2	(DC) Accessed Data is not trustworthy	M
2	(RM) Outdated metadata about Archived Data	N
1	(RM) Data embargo is violated	O
0	(DP) Loss of Original Data due to component faults	P
-1	(DC) Accessed Data unusable	Q

-1	(RM) Loss of Archived Data due to an Organization failure	R
-1	(RM) Wrong interpretation of the Data Management Plan or Metadata	S
-1	(DC) Violation of terms for Data Reuse	T
-5	(RM) Improper use of Archived Data	U
-8	(DC) Plagiarism in the reuse of Accessed Data	V
-9	(DC) Outdated Accessed Data	W
-12	(RM) Destruction of Archived Data due to an attack	X
-12	(RM) Destruction of Archived Data due to a natural disaster	Y
-14	(DP) Destruction of Original Data due to an attack	Z
-17	(DP) Destruction of Original Data due to a natural disaster	AA

Figure 1.4: Round 2 results

Analyzing the results of the second round, some situations emerged that were considered to be the subject of analysis. The first important situation is the importance attributed by the panel of experts to the risk factor "(RM) Loss of Archived Data due to the lack of competencies of the repository staff", being that this factor obtained the maximum score in the two rounds of the study. Given this situation, it can be concluded that, according to experts, this risk factor is the most important in the field of scientific data management. According to ISO 27005: 2011, one of the main vulnerabilities to which information security is subject is associated with the people involved in the organization. In this area, it is understood that the lack of skills of the persons responsible for the management of archived data in repositories may be a major risk factor, since, due to the lack of training in the security and management of scientific data, the incorrect use of certain equipment or even the lack of monitoring to the team responsible for the management of archived data, may cause undesirable changes or changes to the archived data. Another factor that scored high in the second round was the following: "(DC) Insufficient or incomplete metadata about Accessed Data". The metadata corresponds to supplementary information on scientific data. It is therefore extremely important that metadata is well defined and complete so that the data accessed can be more easily used and understood.

An important curiosity of this study was that the factors that obtained lower scores were all those related to the destruction of original or archived data due to natural disasters and / or attacks. One factor that may explain the low score of these risk factors is that the probability of these events occur is relatively low, and experts are therefore likely to consider these events as not posing a serious danger to the safety of scientific data. However, original or archived data losses due to technology failures or problems in some components have already scored higher. This can be explained by the above-mentioned reason, of the actual probability of a technological failure being greater than the probability of a natural disaster occurring. Another factor that scored a very high classification was "(RM) Unauthorized Data Access". Following this factor, the following risk factors were positively classified: (DP) Insufficient Data Management Plan", "(DC) Inaccurate metadata about Accessed Data" e "(RM) Loss of metadata about Archived Data". The first risk factor referred to is undoubtedly a risk factor that is very much addressed in the literature and raises great concerns in the process of managing scientific data. As has been mentioned, some data is confidential and can't be accessed by anyone, so unauthorized access to such data represents a risk. On the other hand, all the other risk factors that follow are like measures that must be taken to prevent unauthorized access to the data. In other words, the existence of a DMP that is not complete or insufficient, the fact that the data accessed has inaccurate metadata and the possibility of loss of the metadata of the archived data are three factors that may lead to unauthorized data access, since there is no documentation (or incomplete) on the data.

To complement the Delphi method and considering the experts' opinions it was developed a risk matrix, to better analyse and evaluate the risks (Figure 1.5). This matrix was developed in a subjective way according to the obtained results, without any scientific method and is just an example of a possible risk matrix. According to each project or organization, the risk matrix can be different.

Probability					
	Very High			A	K
	High	W	V	G	U
	Medium	E	B, D, S	I, J, H	P
	Low	F, L, M	C, N	X, T	Z
	Very Low	Q	O, R	Y	AA
		Low	Medium	High	Very High
		Impact			

Figure 1.5:Suggested risk matrix

6. Conclusions

After the implementation of the method and analyzing the results obtained, it is possible to identify some faults that occurred and points that could have been improved. The fact that this is a relatively recent issue, made the identification of possible risk factors based on a review of the existing literature, take up a

large part of the time available for implementing the Delphi method.

Another difficulty that was found was that the language to be used in the Delphi study had to be clearly and easily understood by anyone. This is undoubtedly a difficulty detected during this investigation, but that is a rather present difficulty in our daily lives. When you have some knowledge about a certain subject, in this situation acquired with the reading and analysis of the state of the art, it is quite difficult to express our ideas so that they are understood by those who are not acquainted with the subject. This was an aspect that made it difficult to develop a list of risk factors.

After this investigation, it is possible to conclude that although the subject under study is not yet widely spoken in the literature, the results obtained, despite the low response rate, were useful to realize the perception that the experts (who were not experts in the area under study, but whose research and projects depend on scientific data) have on the problem of risk management in the domain of scientific data management. According to the results, the importance of the implementation of a data and metadata management plan as a complement to the scientific data is significant because, since all the necessary information on the data management process is contained in these documents, the risk to which the data are subject is reduced. On the other hand, it is also well known that, according to experts, risk factors with relatively low probability of occurrence make them less important. However, control and prevention measures must be taken because, despite the low probability of occurrence, the impacts can be severe.

Finally, it is important to emphasize the importance of the risk management process in the field of data management by helping to identify potential risks to which scientific data may be subject and, measures are easier to take. On the other hand, a good definition and understanding of the data management plan eliminates some risk factors.

7. References

Abbott, Daisy. Annotation. DCC Briefing Papers: Introduction to Curation. Digital Curation Centre: Edinburgh, 2008. Disponível em: http://www.dcc.ac.uk/resources/briefing-papers/introduction- curation/annotation

DCC. (2013). Checklist for a Data Management Plan. v.4.0. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/data-management-plans.

Fernandes, D., Bakhshandeh, M., & Borbinha, J. (2012). Survey of data management plans in the scope of scientific research. INESC-ID. TIMBUS Timeless Business.

Costa, Ana Sara; Barbosa, André; Bom, Francisco; Figueira, José Rui; Borbinh, José. 2017. DecSpace: A Multi-Criteria Decision Analysis Framework. Instituto Superior Técnico. Lisboa, Portugal.).

Cox, L. A. (2009). Risk Analysis of Complex and Uncertain Systems, Springer, New York.

Dalkey, N. C., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. Management Science, 9 (3), 458-467.

Diehl, M., & Stroebe, W. (1991). Productivity loss in idea-generating groups: tracking down the blocking effect. Journal of Personality and Social Psychology, 61, 392-403

Hsu, C.-C e Sandford, B.A. 2007. "The Delphi Technique: Making Sense of Consensus". Pratical Assessment, Research & Evaluation (12:10)

Konow, Irene; Pérez, Gonzalo. Método Delphi. In.: Konow, Irene; Pérez, Gonzalo. Métodos y Técnicas de Investigación Prospectiva para la Toma de Decisiones. Chile: Fundación de Estudios Prospectivos, 1990

Lavoie, Brian (2014). The Open Archival Information System (OAIS) Reference Model: Introductory Guide. 2nd Edition. OCLC Research. Great Britain.2014-7916

Markowski, A, S. &Mannan, M, S. (2008). Fuzzy risk matrix. Journal of hazardous materials. pp 152-157.

Mendes, Felismina, (2002). Risco: um conceito do passado que colonizou o presente. Revista Portuguesa de Saúde Pública, 20 (2), 53-62.

Pickering, Alexander; Cowley, Stephen (2010). Risk matrices: implied accuracy and false assumptions, Journal of Health & Safety Research & Practice, 2(1), 9-16.

Powell, C. 2003. "The Delphi technique: myths and realities", Journal of Advanced Nursing (41:4), Feb, pp. 376-382

Rowe, W. (1988). An Anatomy of Risk. Malabar, Robert E. Kreiger.

Rozados, Helen Beatriz Frota. Indicadores como ferramenta para gestão de serviços de informação tecnológica. 2004. Tese (Doutorado em Comunicação e Informação) - Programa de Pós-graduação em Comunicação e Informação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2004

The Royal Society. (1992). Risk: analysis, perception and management. England: The Royal Society.