

# Risk aware Data Management in Metagenomics

Filipe Ferreira, Miguel E. Coimbra, Ricardo Vieira, Diogo Proença,  
Ana T. Freitas, Luís M. S. Russo, José Borbinha

*INESC-ID* / IST, Universidade de Lisboa, Portugal

{filipe.ferreira,miguel.e.coimbra,rjcv,diogo.proenca}@ist.utl.pt,  
atf@inesc-id.pt,lsr@kdbio.inesc-id.pt, jlb@ist.utl.pt

**Abstract.** The concept of e-Science applies to science that, increasingly, is executed through distributed global collaborations where extensive data collections are processed by large scale computing resources through scientific workflows. These data can be heterogeneous, with multiple provenances, created and transformed according to different data schemas, and the processing can be distributed by multiple locations. This emerging reality brings new challenges, like the preservation of these data workflows. To address specific concerns with Data Management (DM), the concept of Data Management Plan (DMP) was established. However, we claim actual guidelines for DMP ignore achievements by the digital preservation community indicating DM should be mainly a Risk Management (RM) activity. Therefore, we propose to reanalyze the concept of DMP considering the best practices of RM, and claim the concept should be complemented by a Risk Management Plan. The motivation, requirements and validation of this proposal is the project MetaGenFRAME, focused in Metagenomics.

**Keywords:** Risk Management, Digital Preservation, E-Science, Data Management Plan, Bioengineering, Metagenomics.

## 1 Introduction

The term e-Science usually applies to science that, increasingly, is being carried out through distributed global collaborations, comprising the processing of extensive data sets, and the usage of large scale computing resources [15].

The e-Science paradigm leads to new challenges, such as sharing information through distributed locations and the usage of processes that generate large data sets requiring large storage capacity. As a consequence, the concept of Data Management Plan (DMP) has been defined, intended to support the governance, in the context of a specific project, of the data to be created, stored, and shared. However, we can claim actual guidelines for DMP do not include important lessons and guidelines from the Digital Preservation (DP) community, especially those indicating DP should be mainly a Risk Management (RM) activity. Furthermore, there is no process defined on how to apply RM to an e-Science project and based on the outcome how to define the corresponding DMP. Therefore, this article focus on determine a methodology for

defining a DMP using RM based processes. We believe that this method, initially developed and validated under the motivation of the MetaGenFRAME project [18], can be generalized for the field of Bioengineering, which is a heterogeneous area, comprising biology, medicine, bioinformatics, etc. In this sense, we understand MetaGenFRAME as an e-Science biology project, focused on the definition of workflows for Metagenomics, which focuses on sequence analysis and genome annotation.

This paper is structured as follows: Section 2 outlines the principles of e-Science, scientific workflows, Data Management (DM) and DMP. Section 3 introduces the concepts of DP, RM and Risk Management Plan (RMP). Section 4 identifies a set of typical components of a DMP, according to the actual state of the art, and a respective generic risk analysis. Section 5 describes the MetaGenFRAME project, as well as its specific risk analysis. Section 6 presents hypothesis for the alignment of a DMP and a RMP. Section 7 presents some conclusions and future work.

## **2 Data Management in e-Science**

E-Science is defined as a “*global collaboration in key areas of science and the next generation of infrastructure that enables it*”, being characterized by aspects that deliver “added value” to the scientific enterprise [7], namely allowing researcher’s collaboration, increase use of interconnected computers, enabling high performance visualization of data and the electronic distribution and publication of findings.

### **2.1 Scientific workflows**

Scientific workflows are crucial to e-Science, being the means by which scientists execute, reconfigure and rerun their analysis in a repeatable and verifiable way [12]. These workflows typically involve many steps, and access to large amounts of data [13]. Each workflow lifecycle has phases for (i) generation, where the analysis is defined; (ii) planning, where the resources are selected; (iii) execution, where the computations are run; (iv) the result, metadata, and provenance storing [13].

### **2.2 Data Management**

DM is an integral part of e-Science. It involves cooperative work, data security and archiving. It allows researchers to produce higher quality data, increase the exposure for their research, and protect data from being lost or misused [17]. The scientific community has increasingly perceived concerns about DM, especially with the issues characteristic of e-Science, namely data’s provenance, sharing, access and archival [17]. As a result of these concerns, research funders have been increasingly requesting the inclusion of a DMP as part of the project’s proposals. A typical DMP describes what and how data will be created, stored, and shared, with two major purposes [17]: (i) guide researchers to reuse data; (ii) record the project’s DM decisions.

### 3 Digital Preservation as a Risk Management Approach

DP aims to keep digital objects accessible over long periods of time [10]. For that, it is imperative to be able to verify, at any time that these digital objects are what they claim to be. Hence data must be trustworthy, guaranteeing the content is integral and authentic. Information provenance must be assured to guarantee data's traceability. DP environments might require scalability to face technology's evolution achieved through replacement of technological components, thus implying heterogeneity [8].

RM identifies, assesses and mitigates risks to an acceptable level. It manages risks i.e. the uncertainty associated with events which can affect assets [1] [8]. A risk can be positive (opportunity), or negative (vulnerability) that can be explored by threats.

#### 3.1 Risk Management standards

Standards, methods and tools for RM, vary with the market sector, type of business or organizational activities [11]. However, there are also standards that focus on defining the generic terminology, process, principles, methods and techniques such as:

- *Risk management – Vocabulary*. ISO Guide 73:2009. Geneva, Switzerland : ISO [1];
- *Risk Management - Principles and guidelines*. ISO FDIS 31000:2009. Geneva, Switzerland : ISO [2];
- *Risk management - Risk assessment techniques*. ISO IEC 31010:2009. Geneva, Switzerland : ISO [3];

There are other relevant standards that need to be mentioned in the RM area spread throughout several fields. These are represented in Table 1:

**Table 1.** - RM Standards

Area	Standard
<b>Security</b>	ISO/IEC 27005 [4]
	Risk IT Framework [5]
	Operationally Critical Threat, Asset, and Vulnerability Evaluation (OCTAVE) [6]
	NIST, Risk Management Guide for Information Technology Systems [9]
<b>Banking</b>	Value-at-Risk (VaR) [26]
<b>DP</b>	Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) [27]
<b>Enterprise frameworks</b>	AIRMIC, ALARM, IRM (AAIRM) [28]
	Management of Risk (M_o_R) [14]
	Committee of Sponsoring Organizations of the Treadway Commission Enterprise Risk Management (COSO ERM) [16]

### 3.2 The Risk Management process according to ISO - FDIS 31000

The ISO 31000 proposes a reference process to execute RM properly [2] (Fig. 1).

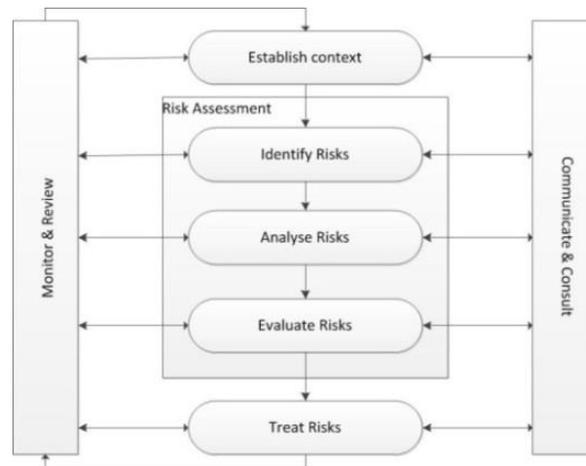


Fig. 1. ISO 31000 RM Process

That process proposes that first the context, strategic objectives and the risk criteria of RM must be defined [8]. The next step is Risk Assessment, which is composed by three distinct phases: risk identification, which generates the list of risks [8]; risk analysis, that considers the impacts and probabilities to determine the risk level; and risk evaluation, to determine what risks need treatment or can be only controlled. The next step is risk treatment, where one or more options are selected to treat risks. Communication and review take place in all previous stages.

### 3.3 Risk Management Plan

A RMP defines the scope and process for the identification, assessment and treatment of risks. The objective of this plan is to define the strategy to manage risks of an organization or project with the minimal impact on cost and schedule, as well as operational performance. The RMP is considered a living document, being updated as needed. A typical RMP comprises the following steps: (i) Planning, (ii) Risk Assessment, (iii) Risk Treatment, (iv) Risk Control and Monitor;

## 4 Risk aware Data Management

The policies that are described in a DMP are specific to the performed research. Researchers might have diverse purposes, resulting in different instances of DMP. However, some issues are common allowing the definition of general DMP guidelines. Table 2 shows a comparative analysis of DMP guidelines, namely the Digital Curation Center (DCC), the Australian National University (ANU) and the

National Science Foundation (NSF) for the fields of Engineering and Biology. In that table we already point where we believe a RMP can be relevant.

**Table 2.** Comparison of DMP Guidelines (present(x), not present (-))

Typical Sections in a DMP	Organizations				To delegate to a RMP
	DCC	ANU	NSF (Eng)	NSF (Bio)	
<b>Ethics and Privacy</b>	X	-	-	X	X
<b>Resourcing (Budget)</b>	X	X	-	-	X
<b>Data Dissemination, Sharing and Licensing</b>	X	X	X	X	X
<b>Data Storage, Preservation and Security</b>	X	X	X	X	X
<b>Data Owners and Stakeholders</b>	-	X	X	X	-
<b>Responsibilities</b>	-	X	X	X	-
<b>Data Formats and Metadata</b>	X	X	X	X	X
<b>Products of Research and Documentation</b>	-	X	X	X	X

These sections and their relation with a potential RMP refer to the following:

- **Ethics and Privacy:** Ethical requirements apply particularly when the research involves animal subjects. These include the purpose and nature of the research, the nature of consent needed and what data must be safeguarded and destroyed. This raises risks that need to be attended, and therefore a RMP should be useful to do so. Examples of ethical risks of the use of genomes can be found in [21];
- **Resourcing:** Appropriate funds should be allocated to DM, namely in time allocations, management of technical aspects, training requirements, storage and backup, etc. In case of lack of funding, DP risks will arise<sup>1</sup>, namely, the entire research might not be developed due to lack of the appropriated infrastructures or personal, and the data generated might no longer be stored and maintained.
- **Data Dissemination and Licensing:** Requirements for dissemination of the research results, which involves licensing measures; Dissemination incurs risks that must be dealt with a cost-benefit analysis to determine if the value outweighs the liabilities. A RMP would be appropriate to perform such a task. Examples of risks associated with data sharing can be found in [22, 23];
- **Data Storage, Preservation and Security:** Requirements for storing data, in a short or long term as well as maintaining the data secure from intentional or non-intentional attacks. RM would be helpful through a RMP; Examples of risks associated with data storing and security can be found in [23];
- **Data Owners and Stakeholders:** List of the owners and stakeholders of the data;
- **Responsibilities:** Lists who will be responsible for ensuring each item in the DMP, as well as, who is responsible for reviewing and modifying it;

<sup>1</sup> <http://www.dlib.org/dlib/july04/lavoie/07lavoie.html?pagewanted=all>

- **Data Formats and Metadata:** Describe the file formats used, as well as, the metadata form used in the project. RM would be helpful in achieving these goals through a RMP. Examples of risks associated with metadata can be found in [24];
- **Products of Research and Documentation:** Determines all data collected, how and the amount. Indicates what needs to be documented, when and how. RM would be helpful through a RMP. Examples of risks associated with lack of documentation, namely data provenance, can be found in [25];

Associated with the former sections, namely those who were delegated to a RMP, there are some typical risks and challenges of e-Science and its workflows regarding the typical DP vulnerabilities and threats [8,23,24]:

- **Workflow's traceability:** Scientists need to know how a workflow was executed and what produced. The dependencies among tasks in a workflow must also be known. The workflow's input parameters must be saved so it can be rerun. Users and workflows must interpret information produced by unfamiliar workflows.
- **Data reuse:** Collaboration between systems that are controlled by different entities, leads to the need for a unique means of data's representation and communication. In e-Science, if a field of study is relatively new, it may be plagued by a lack of standards, which makes data reuse much more challenging.
- **Intentional attacks:** Unauthorized access can lead to stolen or modified data. In e-Science workflows that allow the automatic configuration of script-invoked software, the applications may communicate data to remote locations.
- **Financial changes:** Lack of financial resources to store and maintain all the data or bankruptcy of the organization responsible for managing the data.
- **Organizational changes:** Political or management changes in the entities responsible for data preservation can disturb data storage, sharing and reuse.
- **Web content erosion:** Online reference's small life spans, due to failed or broken links, turn this data's storage and dissemination for future reuse, more difficult.
- **Data sharing:** In e-Science sharing data raises several challenges:
  - Integrity of data: Scientific data must be used and reused retaining its characteristics and content.
  - Data confidentiality: Sharing personal and confidential data can compromise privacy. The gravity varies between processing an individual's personal information or a collective entity's activity details.
  - Intellectual property: Scientists can take the credit for other's research.
  - Data management: Sharing data in a distributed environment hardens its physical management causing difficulties identifying the data set's location.
- **Data set's size:** Applications can generate petaByte-size data sets. Lack of storage space can be catastrophic, since some experiments can only be executed once.
- **Data obsolescence:** It's necessary to access what data becomes outdated. It's also vital to address how or who should delete this information.
- **Infrastructure obsolescence:** The software and hardware supporting workflows can become outdated. Legacy infrastructures may not have metadata support.
- **Natural disasters:** Data must survive events like earthquakes, floods or fires.
- **Human errors:** Data can be unintentionally deleted or modified.

- **Infrastructure faults:** Hardware, software or network faults can compromise the storage, communication or analysis of data and the workflow’s execution.

DP techniques [8] can be applied to mitigate the former risks “shielding” digital objects, which represents the main goal of RM, meaning RM should be used to help attend the former risks. With this in mind, an alignment between DP, RM and e-Science is needed, from which DP represents a main concern [20]. As DP is a concern in a DMP, and, as stated before, DP can be seen as a RM problem, RM principles already identified in DP should be used to help define a DMP.

## 5 A case in Bioengineering: The MetaGenFRAME Project

The MetaGenFRAME project is presented as a use case. It is a Metagenomics project whose description and workflow are detailed next, along with several specific risks associated with it, consisting of an initial risk analysis of the use case.

### 5.1 Bioengineering and Metagenomics

Metagenomics enables the study of populations of microorganisms, namely metagenomes (samples containing the DNAs obtained from uncultured microorganisms) [19]. The MetaGenFRAME project is focused in the study of relatively-controlled environments (possibly composed by several types of different bacteria, with each type being present in different quantities), whose chemical reactions may be influenced and enhanced. The project has a particular focus on prokaryotic organisms. The origin of a metagenome typically consists of a closed environment (for example, a sample containing human gut bacteria) or an open environment (like the open ocean). The tools used for task’s execution are pre-selected via script. The project’s main tasks are shown in Fig. 2. in more detail.

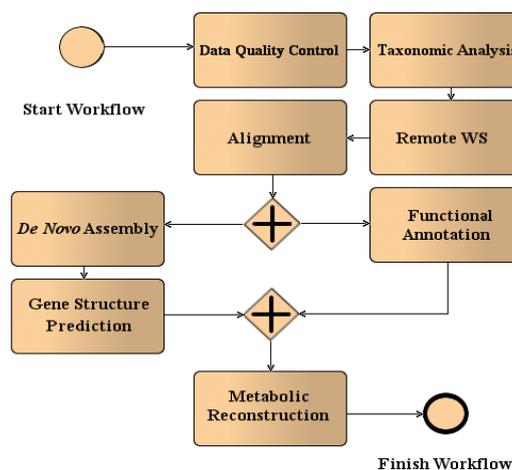


Fig. 2. The MetaGenFRAME workflow.

The following points describe each task in more detail:

- **Data Quality Control:** Before a data set is processed, the information needs to respect certain quality thresholds. This step is executed using NGS QC Toolkit. It's a local task. The inputs are composed of a text file with the sequences that are going to be analyzed, a string with the format used, a variable to filter the OTUs (Operational Taxonomic Units) by identity percentage and, a variable to filter sequences by size. The output is a filtered data set and the removed sequences;
- **Analysis of Taxonomy:** Determine the sample's microbial diversity, to determine the different organisms that are present and, if possible, their resolution levels (species, kingdom, etc). The tool used is MetaPhlAn, being a local task. The input is the filtered data set produced previously and the output consists of several lists of organisms present in the sample, with respective resolutions and identity percentages (converted to query format);
- **Remote WS:** A sequence of web service invocations that use the National Center for Biological Information (NCBI) data base. A first WS uses as input the lists obtained in the former step and returns a set of corresponding IDs. A second WS then consults the NCBI using as input the former IDs returning a list of reference sequences associated to the existing taxonomic results, in .fasta format;
- **Alignment:** Establishment of an order between the studied sequences by comparison with reference sequences. This step uses the TAPyR tool and is performed locally. It receives as input the former list of sequences and generates as outputs a set of aligned sequences in .SAM format, a set of non-aligned sequences in a .fasta file, a set of aligned sequences also in a .fasta file and also another .fasta file with consensus sequences;
- **Functional Annotation:** The set of consensus sequences are submitted to a functional annotation procedure. It is a remote task. The tool used is Blast2GO;
- **De novo assembly:** Sample identification by reconstruction. The tool that supports this task is not determined, but the main contenders are Newbler and MetaVelvet. It's a remote task. As input, it receives the set of non-aligned sequences and as output it returns a set of contigs, which represent junctions of several sequences;
- **Gene structure prevision:** Used to obtain some information about the sample's associated genes and to find if genetic structures are present. One tool that can execute this step is BG7. It is a local task. As input, it receives the set of contigs generated in the de novo task and the output is not yet defined;
- **Metabolic Reconstruction:** One of the aims will be to produce results associated with the metabolism of the initial input. This will use the gene information produced, comparing it against several databases (KEGG as an example). At this stage, genes are mapped against metabolic pathways present in a number of databases like KEGG.

All tasks have a log file saved locally, and the view in which each task presents its results (using Taverna<sup>2</sup>) is customizable to the operator's needs.

---

<sup>2</sup> <http://www.taverna.org.uk/>

## 5.2 Specific risks in the focus project

The MetaGenFRAME project extrapolates several types of information from the data set it receives as input, such as the composition of the organism community present in the sample. It also aims to produce information pertaining the metabolism and main chemical reactions present in the sample. With a particular focus on prokaryotic organisms, this raises important issues associated with the secrecy and storage of data, as it will potentially convey information that is important to the client or entity's activity. An example of such an activity is the process of analyzing and enhancing biomass decomposition, fuel refinement, crude extraction, among others. Such processes are trade secrets, and their study must undertake the precautions mentioned earlier. The project also uses remote web services, so ensuring that the information and services available remotely will remain active is a key-necessity for biologists and other professionals. A preliminary risk analysis of this Metagenomics project produced the following list of specific risks:

- **Accidental alteration or deletion of digital objects or the potential insertion of wrong input values:** Lead to workflow's erroneous execution and worthless data sets. Caused by human errors. One example is the introduction of the wrong value in variables that indicate the percentage of a sequence's nucleotides that must be of quality regarding the total length of the sequences which are filtered in the data quality control task, therefore influencing all the following results.
- **Purposeful alteration or deletion of digital objects:** Caused by intentional (internal or external) attacks. For example, an attack with alteration or deletion of the external WS, NCBI or local PC used. Network or communication failures or weak security measures in the data bases or the local PC can be exploited.
- **Loss of data:** Media faults, caused by failures in the storage systems. This vulnerability can occur in the NCBI and in the local PC used. PCs and external HD are useful for short term storage, but inadequate for long term storage due to high failure rate. This risk is enhanced with the lack of a long term storage policy. Legislative changes in legal requirements, like the minimum time a digital object is required by law to be stored, can lead to data loss. Economic or organizational breakdowns can also influence the organization running the NCBI, causing its termination. Natural disasters can also endanger the data sets stored in the NCBI.
- **Workflow execution failures:** The remote WS or the NCBI can be unavailable for technical reasons, like communication and network failures. The Scientific Workflow Management System (SWMS) used is Taverna, along with several software tools that execute the tasks represented in Fig 2. Software and hardware failure or obsolesce, in the NCBI or local PC, can lead to this risk. In the case of failure, it's problematic if the SWMS doesn't provide debugging capabilities that show the failure cause. For example, if the WS fails and no information is provided the user is left wondering whether the service failed locally or remotely.
- **Data traceability loss:** Lack of documentation endangers data's recreation. In this project, log file's loss comprises the workflow's recreation, with the same inputs.
- **Data sharing:** It is essential to select carefully what data can be reused by the scientific communities or the one that must be kept confidential.

## 6 Alignment of a DMP and a RMP

As a result of our analysis, we propose that sections of a typical DMP should be delegated to a RMP to be defined using RM principles and good practices.

The RMP would serve as a complement to the DMP, therefore producing a more complete and robust DMP. This means these sections should be complemented through the application of RM principles instantiated by the RMP.

The definition of a RMP for Bioengineering and its application requires a specific method, which will be defined in our future research. The use case presented is going to be used to evaluate that method. Evaluation metrics are also going to be applied to the obtained results. These metrics must correspond to the successful achievement of the four main steps that compose this method: (i) Identification of 70 to 90% of the risks, (ii) Successful analysis of the risks, with a quantitative level attributed to each risk, (iii) Successful risk evaluation with the determination of which risks need treatment, through the level of risk, (iv) Successful determination of risk treatment measures for each risk that needs it;

## 7 Conclusions and future work

RM has applications in different areas, organizations and projects. In e-Science, collaboration and DM are crucial. New challenges and risks are raised and must be assessed, so the project's data can be preserved and reused.

To answer to these dilemmas, DMP are developed before a research project takes place. The problem we identified resides in the fact that these DMP don't take in consideration that DM is a RM activity, making them potentially incomplete and immature. In order to assess that, the need for a method for developing a DMP using the good practices of RM is proposed. This method will serve as a starting point to identify a more generic method that can be applied in most e-Science projects in the field of Bioengineering.

We have identified a number of risks from the MetaGenFRAME use case. This identification step correspondes to the first stage of a RM analysis. In future work, the analysis on the same case will continue with the analysis, evaluation and treatment of the identified risks. We expect that from that the proposal for a generic method for RM will emerge.

**Acknowledgments.** This work was supported by national funds through FCT (Fundação para a Ciência e a Tecnologia), under projects PEst-OE/EEI/LA0021/2013, TAGS PTDC/EIA-EIA/112283/2009, and DataStorm EXCL/EEI-ESS/0257/2012, and by the project TIMBUS, co-funded by the European Commission under the 7th Framework Programme for research and technological development and demonstration activities (FP7/2007-2013) under grant agreement no. 269940.

## References

1. ISO (2009) Risk management – Vocabulary. ISO Guide 73:2009. Geneva, Switzerland.
2. ISO (2009) Risk Management - Principles and guidelines. ISO FDIS 31000:2009. Geneva, Switzerland.
3. ISO (2009) Risk management - Risk assessment techniques. ISO IEC 31010:2009. Geneva, Switzerland.
4. ISO/IEC (2011) Information technology. Security technologies – Information security risk management. ISO/IEC 27005:2011. Geneva, Switzerland.
5. The Risk IT Framework (Exposure Draft), IT Governance Institute (2009).
6. Caralli, R. A., Stevens, J. F., Young, L. R., Wilson, W. R. (2007) OCTAVE Allegro: Improving the Information Security Risk Assessment Process, Software Engineering Institute at Carnegie Mellon University, 2007.
7. Jankowski, N. W. (2007) Exploring e-Science: An Introduction, Journal of Computer-Mediated Communication.
8. Barateiro, J., Antunes, G., Freitas, F., Borbinha, J. (2010) Designing digital preservation solutions: a Risk Management based approach. The International Journal of Digital Curation, Issue 1, Vol. 5, 2010.
9. NIST Special Publication 800-30 (Revision 1): Risk Management Guide for IT Systems, 2012.
10. Lipkis, T., Morabito, S., Reich, V., Robertson, T., Rosenthal, D. S. H. (2005) Requirements for Digital Preservation Systems: A Bottom-Up Approach, D-Lib Magazine, Number 11, Vol. 11, 2005.
11. Ramirez, D. (2008) Risk Management Standards: The Bigger Picture Information Systems Control Journal, Vol. 4, 2008.
12. Braga, R. S. (2007) Automatic capture and efficient storage of e-Science experiment provenance, Wiley InterScience.
13. Deelman, E., Chervenak, A. (2008) Data Management Challenges of Data-Intensive Scientific Workflows, USC Information Sciences Institute.
14. Office of Government Commerce (OGC) (2007) Management of Risk: Guidance for Practitioners (M\_o\_R), United Kingdom, 2007.
15. Paul A. David, Spence M. (2003) Towards institutional infrastructures for e-Science: the scope of the challenge, Oxford Internet Institute.
16. Committee of Sponsoring Organizations of the Treadway Commission (COSO) (2004) Enterprise Risk Management — Integrated Framework, Jersey City, NJ: AICPA, 2004.
17. Fernandes, D., Bakhshandeh, M., Borbinha, J. (2012) Survey of data management plans in the scope of scientific research, Inesc-ID, Timbus Timeless Business.
18. Coimbra, M. (2012) Metagenomic Frameworks, Project Report, Universidade Técnica de Lisboa, Instituto Superior Técnico.
19. Wooley J. C., Godzik, A., Friedberg, I., (2010) A Primer on Metagenomics, University of California San Diego.
20. Barateiro, J., (2012) A Risk Management Framework Applied to Digital Preservation, PhD Dissertation, Universidade Técnica de Lisboa, Instituto Superior Técnico.
21. Kaye, J., Boddington, P., Vries, J., Hawkins, N., Melham, K. (2010) Ethical implications of the use of whole genome methods in medical research, Europe Journal of Human Genetics.
22. Bimholtz, J. P., Bietz, M., J. (2003) Data at Work: Supporting Sharing in Science and Engineering, GROUP '03 Proceedings of the 2003 international ACM SIGGROUP conference.

23. Schmitt, C., P., Burchinal, M., (2011) Data management practices for collaborative research, *Frontiers in Psychiatry*.
24. Day, M. (2004) *Preservation metadata initiatives: practically, sustainability, and interoperability*, University of Bath.
25. Miles, S., Deelman, Ewa, Groth, P., Vahi, K., Mehta, K., Moreau, L. (2007) *Connecting Scientific Data to Scientific Experiments with Provenance, e-Science and Grid Computing*, IEEE International Conference.
26. Holton, G. (2003) *Value-at-Risk: Theory and Practice*, Academic Press, 2003.
27. McHugh, A., Ruusalepp, R. Ross, S. & Hofman, H. (2007). *The Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)*. DCC and DPE, Edinburgh. 2007.
28. Association of Insurance and Risk Managers (AIRMIC), ALARM (National Forum for Risk Management in the Public Sector), Institute of Risk Management (IRM) (2002) *A Risk Management Standard*, London, 2002.