Research data management in the image lifecycle: a study of current behaviors

Joana Rodrigues[1,2][0000−0002−1309−2122] and Carla Teixeira Lopes[1,2][0000−0002−4202−791X]

[1] Faculty of Engineering of the University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
[2] INESC TEC, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
joanasousarodrigues.14@gmail.com, ctl@fe.up.pt

Abstract. Research data management (RDM) practices are critical for ensuring research success. Data can assume diverse formats and data in image format have been understudied in RDM. To understand image management habits in research, we have conducted semi-structured interviews with researchers from four research domains. Most researchers do not formally manage their images, nor do they develop RDM plans. They assume that image management is not a topic discussed at project meetings. In turn, they tend to perform some individual practices, depending on the context and their own opinion, such as creating captions to describe the images and organizing and storing the images in specific locations. However, they see these habits as necessary and admit that they will start to do so in a formal and collaborative way with the working group. These results provide valuable information on practical aspects of the use and production of images in research.

Keywords: Research data management · Images · Research · Image life cycle

1 Introduction

Images are dominant in communication, and their abundance, diversity of origins, and variety of holders create a multiplicity of practices regarding description, interpretation, and systematic use [15]. Many authors talk about the role of human behaviors in the life cycle of images [7]. The awareness and education of others for the practices of image record description is essential, commenting on the importance of eliminating description mistakes associated with automatic systems [4].

Research Data Management (RDM) is gaining the attention of researchers, who often ask for support in the process [2]. Tasks as an adequate data organization, rigorous description, storage, and sharing allow data to keep their meaning and facilitate their interpretation [7]. Data management is also fundamental in the communication between researchers and the scientific community, where images can play a central role in the "teaching-learning" process [10].

When we talk about research data, a dictionary in the area of RDM tells us that they are "Facts, measurements, recordings, records, or observations about the world collected by scientists and others, with a minimum of contextual interpretation. Data may be in any format or medium taking the form of writings, notes, numbers, symbols, text, images, films, video, sound recordings (...)"[6]. In its turn, image is the visual representation of something. It can be captured through various instruments or obtained through manual techniques (such as drawings, paintings, engravings and illustrations). Images can be in an analog or digital format. Examples of images are photographs, microscopic images, medical images (e.g., x-ray image), paintings, illustrations, videos (moving images), computer-made images, engravings, graphics, drawings, maps.

The lack of knowledge about image management in research motivates this study. Images are fundamental pieces in the research process, facilitating interpretation and analysis. For this reason, it is essential to know how researchers deal with images and to make them aware of the value of images, motivating its adequate management.

The capture and production of images are increasingly facilitated by the emergence of technologies that allow you to obtain an image in seconds. With a simple smartphone, we can capture or produce an image quickly and easily. In addition, technological devices also allow, in a fraction of a second, an image to be broadcast to any part of the world. This scenario of image empowerment leads to more images being used and produced in the context of research. However, the formal management practices of these images are not aligned with their exponential growth. How do researchers deal with their images? Images are organized, analyzed, described, shared according to previously established methods? This work arises to clarify what happens with data in image format. We conducted interviews with researchers from different research domains addressing how they organize and manage the images they use or produce in their projects. Results contribute to the definition of a current scenario, leading to proposing the conditions and the necessary steps to be followed.

2 Literature Review

An unprecedented growth characterizes the volume of data produced in research, as powerful computational capabilities are available to even small research groups: the so-called long-tail of science [9]. Usually, small groups or individual researchers have minimal resources to ensure the long-term availability of their data [5]. As such, they need good research data management practices supported by practical tools so that the datasets they produce can be made available to others [18]. Make data available is especially important as more research funding agencies adhere to the European Commission's Guidelines on FAIR Data Management in Horizon 2020, which advocate for a set of principles to make data Findable, Accessible, Interoperable, and Reusable [14].

Palmer et al. [13] analyze the importance of creating a structure of principles and processes that help articulate and support data description. One point of

consensus was the crucial role of images, a typology that constitutes a social object, with materiality and associated ideas [1, 12]. For researchers, images have a double purpose: they function as metadata, providing context, and as a vital medium to record the object of study [1]. This approach leaves room for positioning an image in RDM as a fundamental asset in describing and interpreting specific domains.

To the best of our knowledge, only one work studied the habits of researchers managing images in the research process. Fernandes et al. [8] created a questionnaire for this purpose and analyzed the collected responses. The authors concluded that researchers do not have the practice of managing and organizing the images they use or produce in their projects. Unfamiliarity, lack of standardized practices, and time constraints are the main reasons for these gaps in image management.

3 Methodology

To further investigate the responses and conclusions of Fernandes et al. [8], we have conducted interviews with researchers from different domains. Although the questionnaires allow reaching a more significant number of people, the interviews allow more profound analysis and validation of the results, provided by the direct contact that the interviews allow. With the results obtained in both works, we can develop a holistic view of the conclusions, facilitating a comprehensive and general understanding of the phenomena.

The methodology used in this work is a qualitative method of investigation through the case study, namely exploratory research.

Exploratory research aims to improve familiarization with the case under study, so this method will investigate the topic so that it is possible to reflect on it and improve its understanding [16]. The exploratory research aims to generate ideas and hypotheses that complement this theme and help the work be developed, namely in identifying a set of habits and behaviors related to data management in image format. The case study is an integral part of the qualitative method of investigation, which aims to specifically analyze the topic in question and, in this way, create in-depth clarification about it.

In this case, qualitative research relies on primary sources in the field of empirical materials. For this, semi-structured interviews are carried out. We recruited, by convenience, 15 researchers in 4 different domains: 4 researchers in Life and Health Sciences (LHS), 5 in Exact Sciences and Engineering (ESE), 3 in Natural and Environmental Sciences (NES), and 3 in Social Sciences and Humanities (SSH). These are four main research domains used by the national funding agency for science, research, and technology in the country of this study.

For recruitment, the choice fell on some researchers who had already participated in other projects and partnerships, as it is more agile, faster, and more likely to be successful. Table 1 shows the profile of the fifteen researchers interviewed. To maintain the confidentiality of the participants, we do not disclose all the information about the participants.

Table 1: Researchers profile

Research	Specific work area	Academic qualifications	Position in the institution	Age group
ESE1	Data science	PhD	Senior Researcher	30-40
ESE2	Enterprise systems engineering	MSc	Researcher	30-40
ESE3	Human-centered computing	PhD student	Research Assistant	30-40
ESE4	Enterprise systems engineering	MSc	Researcher	20-30
ESE5	Electrical and Computer Engineering	PhD	Senior Researcher	40-50
LHS1	Genome editing	PhD student	Research Assistant	30-40
LHS2	Animal health	PhD	Researcher	30-40
LHS3	Biological an molecular science	MSc	Researcher	20-30
LHS4	Reproductive genetics	PhD	Associate Researcher	20-30
NES1	Biology	PhD student	Researcher	20-30
NES2	Biodiversity	PhD	Post-Doc Researcher	30-40
NES3	Biology	PhD	Researcher	40-50
SSH1	Psychology	PhD	Senior Researcher	30-40
SSH2	Psychology	PhD	Researcher	40-50
SSH3	Anglo-American Studies	PhD student	Researcher	20-30

The interviews aim to give credibility and depth to this study, as they make it possible to indicate results based on real practices. Interviews can be divided into three types: informal conversation, interview guide, and semi-structured [11]. The choice was made for the latter since it does not require a rigid, fixed, and standardized protocol, that is, the questions do not have to be asked in an orderly manner and with a standardized formulation. So, this interview model allows the elaboration of a set of pre-defined questions, however, the guide is adaptable according to the direction of the dialogue between the interviewer and the interviewee [11].

Interviews were semi-structured by the phases of the images life cycle: Planning, Creation/Compilation, Quality assurance, Processing/Analysis, Description, Storage, and Sharing [17, 3, 19], as you can see in Figure 2. This choice was because this life cycle includes all the phases that we consider essential in managing data in image format. We intended to balance the importance given to all stages, showing that the success of image data management lies in the joint use of all these phases of the life cycle.

The interview script includes fourteen questions that are available in the INESC TEC data repository[3]. Each set of questions was developed in order to understand the behaviors in each of the phases of the cycle, so that it was possible to concretely determine the actions, without mixing lifecycle stages. Due to COVID-19, the interviews took place remotely. The sessions were recorded in video and audio for later analysis. Data were anonymized, and all recordings will be deleted when the study is complete. The interviews took place from May to October 2020. For the development of the interviews, all the researchers read and filled out a document called "Informed, Free and Clarifies Consent to Participate in a Research Projet according to the Declaration of Helsinki and the Convention of Oviedo", where they attested that they agreed to participate in the study and that they were guaranteed confidentiality, the exclusive use of

[3] https://doi.org/10.25747/hcrd-ht83

the data collected and anonymity, promising never to public the identification of participants.

To process and analyze the data from the interviews, we began by transcribing the interviews (after all interviews have already been done). After that a table was prepared for each question, where the identifier of each interviewee was placed in the first column (e.g., ESE1), in the second column, it was described o the research domain and in the third column a response summary. The summary is composed of topics that include the main aspects of what was said, in order to make clear the specific behavior in that situation. Below, in Figure 1, we present a summary of our methodological approach.

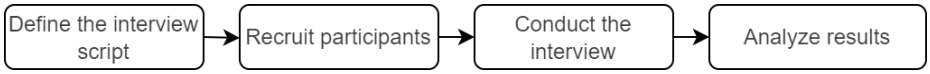

Fig. 1: Our methodological approach

4 Results

In this section, we summarize the findings obtained in the context of the interviews, organized by the different phases of the life cycle of the images. Figure 2 visually represents the topics examined in the results.

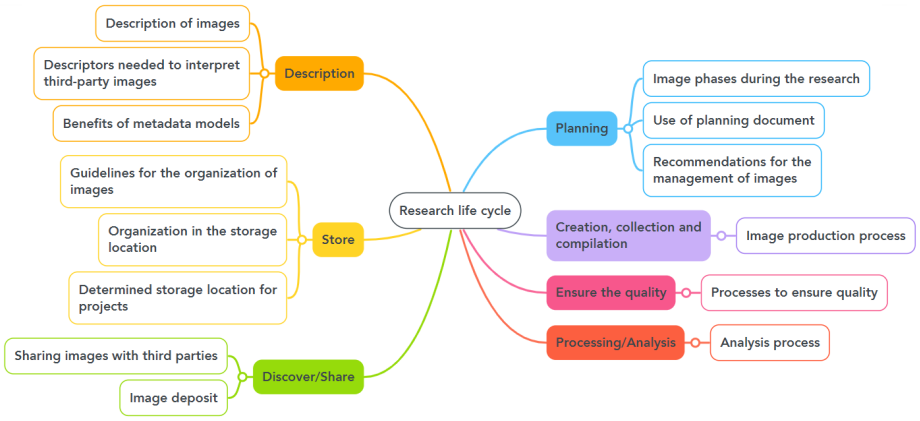

Fig. 2: Mind-map of the topics examined in the results

4.1 Planning

We asked if interviewees plan the phases through which the image goes, produce a document establishing guidelines for the management of the images, and if and what recommendations they would like to have for image management.

Image phases during the research

It was clear that most researchers do not formally plan the management of the images. However, in the research routines, they carry out practices that indicate informal planning. We identified sequences of actions in the interviews. Table 2 shows these sequences, columns identify the researchers, along with their research domains, and rows represent phases of image management. Cells' values indicate the order of the stage in the overall sequence.

Table 2: Image management planning cycles identified by researchers

	ESE1	**ESE2**	**ESE3**	**LSH1**	**LSH2**	**NES1**	**SSH1**
collect	1				1		
select	2						
store	3				2	2	1
create		1		1		1	
treat		2		2			
use		3			3		
share		4				4	3
analyze			1			3	
investigate			2				
design			3				
assess quality				3			
refine				4			
index				5			
publish				6			
back up							2

As previously mentioned, in most cases, formal planning of image management does not exist. The images go through the cycle that the researchers, at that moment, considered appropriate. Sometimes, the image production is planned and described in advance. Other times, it appears in meetings, and then they are used for various purposes. Some researchers refer that when there is planning, the image is produced and worked collaboratively. Data observation is one of the stages that typically occurs when there is planning. In these cases, the image has a prominent role since it allows the interpretation and further processing of the data. For some researchers, if an image is reused, the image quality check is planned, and only then is it used for the research. One of the researchers said that the evaluation of this quality was also carried out through metadata, as it allows to verify the information and better interpret the image. Only one of the interviewees said that planning always takes place.

Of the seven cases presented in the Table 2, it is interesting to highlight the researcher designated by SSH1. This is the one that most significantly differs

from the sequences of the other researchers. This researcher, from the domain of psychology, mentioned that his practices are neither formal nor pre-established, but carried out according to what he considers to be "the most qualified method for treating images". In the absence of a structured procedure, this researcher stores your images, then prepares a back-up, as it believes that "data safeguarding is essential and its eventual loss could constitute a rupture in the research process" and, finally, shares the data with third parties, particularly in data repositories.

Use of planning document

Only two researchers said they always have a planning document, as it ensures that the team knows what has been defined and approved. Yet, the predominance is the absence of a document including standard guidelines for image management. Many say they don't feel the need for it. One of the researchers said that he knows only a few open-access guidelines and tools mainly aimed at metadata. Another researcher stated that the non-use of this document happens because there is no such culture in science, as it is not a habit in the context of research projects. Most researchers assume that they never thought about the possibility of developing this type of guide, but they consider it an advantage. One of the researchers notes everything planned and accomplished, but not in the form of a guiding document. There were no significant differences by research domain.

Recommendations for the management of images

Some researchers answered recommendations about image management are essential but do not know precisely at what stage they may be relevant. One stated that recommendations are important mainly at an early stage of the project. Another said recommendations are vital in the planning phase to affect the quality of the production and use of the images. Another researcher noted that recommendations are not essential and may stop being applicable as projects change. However, many researchers said these recommendations would be helpful. Most of them highlighted that a document with universal recommendations to guide researchers to prepare and publish their data clearly would be advantageous. They also spoke of the importance of these recommendations providing practical examples. Many researchers said that these recommendations could safeguard characteristics of the use of images and their benefits. Also, these recommendations can make researchers able to more quickly and easily assess the images they use or produce.

4.2 Creation/Compilation

At this stage, it was asked how the image production process develops during the investigation.

Image production process

One of the ESE researchers said that he produces many photographs, maps, and graphs. The process involves documenting the code, rerunning the code, and generating the image again. Other ESE researchers said that this process could vary, but, usually, existing images are chosen and treated. The image reference (name that clearly identifies the image) for the future is always kept. The researchers also mentioned that images can come in two forms: 1) they are produced spontaneously when immediate need arises 2) they are planned in advance. In several cases, the images are produced collaboratively in this research domain.

In the domain of LHS, all the researchers said that the production process depends on the type of image. Production processes change with the types of image. Graphics are usually associated with data collection, data organization, image creation, image refinement, and treatment. Graphical abstract and infographics - root creation. Western plots - created photography, simple editing, and comparison of the bands. The creation of histograms, boxplots, funnel plots, and different types of graphs were also mentioned. Researchers also create explanatory diagrams and flowcharts, usually created with software.

The NES researchers mainly produce images in geographic information systems, where they create the map and add informational elements. However, this process depends on the type of investigation. This issue is often seen as a part of the research, where data collection and analysis processes are also considered. Satellite images are usually downloaded and pre-processed to identify defects, apply corrections, and georeference the image.

The SSH researchers say the images can be produced or reused, depending on the type of research and its purpose. In sociocultural projects, as a rule, images are produced. When dealing with historical or conceptual projects, the images are reused from repositories, articles already published, or past projects. Regarding production, all the researchers in the area mentioned the importance of anonymization. For this reason, the production process consists of collecting, blurring the characteristics of identification, storage, and use.

4.3 Quality assurance

In this section of the interview, we asked about the existence of processes to ensure the quality and the activities inherent to this.

Processes to ensure quality

Most researchers say there is little or no activity to guarantee or verify the quality of the images. One of the researchers explicitly stated that there is no such verification, but said that if the image "looks good" and reflects what was intended, it is approved. One of the researchers noted that this verification occurs mainly in western plots. For some interviewees, verification only occurs when they fail to perceive the original image. Some researchers claim that, although images are critical to projects, they are not published and shared most of the time, so a

raw image is sufficient to analyze what is intended. One of the researchers said he checked the quality when it comes to videos and photographs, as the details of these images can be critical. Some researchers mentioned adjustments to the code, cropping of images, treatment of image noise, care with contrasts, verification of the dimensions and source of images, aesthetic validation of the image, and verification of spatial resolution, temporal resolution and georeferencing.

4.4 Processing/Analysis

For this topic, we asked about the image analysis process, namely how information is extracted from the images.

Analysis process

Most interviewees do not have a pre-established method of analyzing/processing the images. Most researchers said they did not analyze the images. Some researchers claim to do this frequently with methods that depend on the images and the project. One researcher stated that this analysis/processing depends on who performs it, as he will check what will be necessary according to the type of image. One researcher that works exclusively with maps said that this type of image is already the final product. Researchers identified the semi-automatic classification of images, the geostatistical processes, the extraction of direct or indirect information from the image, content analysis (especially for videos and photographs), information extraction (for graphics), visualization of the data, and the annotation of the image results as analysis/processing methods. Researchers also mentioned visual analysis, measuring distances, checking details, and cutting out irrelevant parts.

4.5 Description

We asked interviewees how they describe images, which information they need to understand and use third-party images, and the benefits and interest of a metadata model for image description.

Description of images

Few researchers describe their images. The description is not a frequent practice, and its realization depends on the context, the image, and the researcher who produces it. For one of the researchers, "less is more", that is, too much description can affect the purpose of the description and confuse whoever tries to interpret it. We identified the following examples of description: 1) description through technical details (for those who already know the context of the image) and general details (for those who do not know), 2) description through captions (most mentioned), 3) elaboration of metadata (most used to describe components of maps, production periods, dates and the purpose of the image), 4) description through annotations in a laboratory notebook. For researchers who

describe images, the advantages include sharing metadata/captions for better interpretation and dissemination of the study.

Descriptors needed to interpret third-party images

The most mentioned descriptors are the source, rights of use, author, place of origin, title, context, characteristics of obtaining and processing the image, caption, scale, related resources, methodology, description, date, production method, treatment and analysis processes, spatial resolution, temporal resolution, abstract, keywords, and the area of study. Elements of authorship, date, context (abstract, description, keywords), methodology, production techniques, and study area were the most mentioned and are transversal to all research domains. Some elements vary according to the particularities of the research domain.

Benefits of metadata models

Researchers see metadata models as very important. Only one researcher said he was not sure, as he did not know whether the outcomes would compensate for the time spent in the description. The other researchers responded positively to the possibility of using metadata models, stating that they will facilitate the storage of relevant information about an image. In addition, the metadata model would serve as a recommendation to describe, a description standard to follow, and would facilitate description for researchers. One researcher said that the model should be adapted to the type of investigation. Another researcher thinks it is difficult for a model to reach a balance of descriptors. On the one hand, there are general recommendations that have general applicability and are very relevant. On the other hand, a generic model can leave out important details, such as the context and the scientific area.

Almost all researchers mention that it is helpful to have guidelines for describing images. They also claim that it can be very advantageous to have a cross-cutting model for multidisciplinary research groups to harmonize and create common communication points. They highlight the importance of a core of generic and specific metadata depending on the area. Researchers stress that the metadata model can function as a working tool and help members of the research group follow the same working methods, avoiding deviations. The metadata model can also stimulate the publication and sharing of data in repositories and directories.

4.6 Storage

We asked researchers whether they define guidelines for the organization of the images, what is the organization of the images in the storage location and if there is some central location where the images are stored.

Guidelines for the organization of images

No researcher defines guidelines for organizing their images. One said that he does not do it formally but pays special attention to the formats, the way they are organized, and the place where they are stored. Another admitted that this theme is a general failure, mainly because each researcher stores his own images, which is not safe and does not provide access to other research group members. Many researchers admit to having their method of organization and storage, but it is not formal or discussed in the working group. Everyone says they give importance to the way they organize the data, but it is not a procedure stipulated by the project's guidelines.

Organization in the storage location

All researchers claim to organize the images into project folders and subfolders on the computer. Some affirm organize projects folders by content (separate the content into categories that they create). Some researchers divide the raw data from the processed data. Others have folders for analysis and folders for scientific articles, each having different versions of images.

Determined storage location for projects No ESE researcher has a storage location, except one researcher who said that the images are stored on the computer and accessible via the internal network. In the LSH domain, only one researcher claims to have a storage location for the project's images. This researcher said that the existence of this location depends on the project. If it is a small project, all the images are on the computers; if it is a large project, a folder is created in the cloud for everyone to access. Two of the NES researchers say that this central location exists. One said that it is on the computer or the external disk and that, sometimes, it is in both places at the same time. Another stated that they have a Network-Attached Storage (NAS), a device dedicated to the storage of network data, allowing homogeneous access to data. For the SSH domain, only one of the researchers stated that this location exists. He said that they have an external disk where they store all the data. This disk is kept in the group's laboratory and accessible to project members. Researchers also have a copy of this data on their computers.

4.7 Sharing

In this last phase, we asked with who the interviewees usually share their images, what motivates sharing, and if they ever made an image deposit.

Sharing images with third parties

Some researchers share their images with group colleagues in presentations to colleagues, at conferences, in articles, and on posters. Researchers also share the images with people associated with the project for specific tasks. Some share images with colleagues in the group to receive feedback on the image, to know if it is readable if it makes sense and if it is reliable. One researcher never shares

the original images. Another said that it only shares images in articles. Most researchers share images after being treated because others may not easily interpret a raw image. One of the researchers says he is careful to accompany the image of a caption with the context whenever he shares it. The reuse of data is one of the main reasons for sharing and the possibility of new collaborations, sharing new knowledge. Some researchers point to confidentiality and sensitive data as barriers to sharing. The fact that images are not central to the project also limits sharing.

Image deposit

Only some researchers have already deposited their images in a joint deposit that also included other types of data. One of the researchers said it was unusual and that he normally didn't hear about it. Another researcher deposits data, but never deposited only images. This researcher said he did not see any relevance in deposit the individual images. One of the researchers said he knew that depositing is one of the data management requirements but has not yet done so with the images. One of the researchers said that sometimes deposit their data but considers that it is more beneficial to share the information that gave rise to the images and the results than the image itself. A part of the interviewees claim to see no use in depositing the images in isolation; it would only make sense accompanied by textual information. These researchers say that they only make deposits if it is a requirement of the conferences or scientific journals. Deposit normally occurs in repositories.

5 Discussion

We interviewed a group of researchers about how they proceed and see the management of imagery data. We realize that all researchers produce or use images, but these are not always essential parts of the study. Therefore, many researchers are not concerned with its management. The results coincide with those collected by Fernandes et al. [8], where the generality said they did not have a guiding document or formal practices for this purpose. Although images are widely produced and used in research, there are no guidelines to direct researchers to standardized practices regarding images.

In the creation process, it is visible that the methods differ by research domain. In this study, the use of digital instruments is highlighted, prevalent in ESE, LHS, and NES, which was also confirmed in the previous survey [8]. At certain times, researchers pay attention to quality assurance, mainly in the domains of LSH and NES. However, there were no standard procedures to do so, varying with the type of study. In the questionnaire article [8], not all researchers perform this task, which also happens here. Regarding the processing/analysis phase, most researchers admit that it occurs according to the situations and context of the work, without any formal method or common practice in the work group. In the questionnaire article [8], it was identified the same.

Most researchers do not use metadata models. They consider it relevant, but it is not a common practice. The description is usually done through captions. They believe that when the image is reused, metadata is essential for interpretation, and the use of a metadata model for projects would bring benefits. The previous article [8] also shows that researchers do not use metadata models. However, in this work, we gathered more comprehensive answers regarding the descriptors that researchers consider relevant. All the researchers were able to find a justification for choosing a specific descriptor. For example, the keywords were mentioned several times, as they are essential for a preliminary analysis of the relevance of the data. Regarding storage, researchers do not have formal habits or standardized methods of organization. The computer prevails as a storage location, as it is easily accessible. The same was evident in the study of Fernandes et al. [8]. Finally, the researchers showed that no sharing habits. Sharing demonstrations take place mainly in scientific articles and conferences. Internally within the project, sharing occurs frequently. They have also not been shown to have depositing practices in repositories or the like. The same results were obtained in the questionnaire article [8].

This work allowed us to confirm several results obtained previously with a questionnaire. The use of the interviews allowed close contact with the researchers. If the questionnaire makes it possible to collect factual information, the interviews were essential to determine each of the choices, obtain justifications for answers, understand the various topics and understand the researchers' predisposition to initiate image management practices.

We noticed that most researchers are aware of the RDM guidelines, but this is not usually discussed in group/project meetings. Researchers also do not yet assume that imagery data is amenable to management, for many researchers as images are complements to the work and, therefore, their management is not essential. For most researchers, the time spent is still not worth the effort in data management unless that is an obligation. It became clear that the existence of practical and easy-to-implement standards and models (metadata model for image description, for example) is welcomed by researchers, as it is a way of seeing the work done with little effort. It was clear that the type of image is one of the most significant differentiating factors for the kind of management done. The research domain brings particularities, but it does not always determine all differences. Sometimes, the type of images and the conditions leading to their acquisition are more differentiating than other characteristics.

During the research phase in the literature on the research life cycles, we realized that none of the options contemplated data security and privacy procedures as one of the main phases. However, issues related to data confidentiality, namely images produced or used in the research context, must be reflected. In fact, an image can indicate, for example, a place or a person, thus breaking the principles of confidentiality that are often desired.

The General Data Protection Regulation (GDPR), approved on April 27, 2016, launched new challenges to the various audiences that the sphere of privacy affects, from citizens, companies, organizations, educational institutions,

or research centers. The protection of individuals concerning the processing of personal data is a fundamental right that needs to be defended, regardless of the context. With the validity of the Regulation, a solid and more coherent protection framework at the European level was demanded. Personal data are understood to be those that allow the identification of the data subject, such as name, identification number, location data, identifiers electronically, and data related to physiology, genetics, mental health, economic, cultural, or social situation. However, if an image shows or indicates any of these cases, it may compromise the protection of the author's data or the actors highlighted in the image.

In the context of the research, some guidelines have already emerged to prevent this problem. Data Management Plans (DMP) and Privacy Impact Assessments (PIA) are very useful tools in data protection. The DMP allows structuring and organizing data from research projects, playing a central role in the development of good research practices, contemplating privacy actions to be applied to data. The PIA was introduced through the General Data Protection Regulation (article 35). The PIA aims to prepare a document with guidelines that intend to direct and control privacy risks. This document must foresee the risks and establish concrete scenarios associated with the research data, in order to anticipate problems and define strategies for the success of the management.

Therefore, it would be important to start a reflection on the clear introduction of a privacy phase in the research life cycles, so that practices related to this topic are mandatory for researchers who want to fully comply with security and data confidentiality.

6 Conclusion and future work

There is still a way to go in the sphere of research data management. Researchers tend to agree that the RDM is vital in science, but a lack of knowledge and time is often a hindrance to progress. Within the various types of data, the text is the one we see most often associated with data management practices. However, the rest of the typologies should not be overlooked. Images, for example, can be valuable instruments in scientific production. Their capacity for representation can help researchers. Besides, the use or production of images is not limited to a domain of research. Allied to this, the enormous technological evolution and the new capture devices that allow to obtain and spread an image quickly.

It is visible that when confronted with this reality, researchers realize what habits they have regarding images. However, they do not usually set guidelines and plan methods for managing their images. Most of the behaviors highlighted in this work mainly occur informally and without conformity of the working group. The path is to continue to reflect on these issues, confront researchers with these issues and motivate data management practices to become recurrent. Once implemented in the working group, they are easily adapted to the various projects, bringing significant benefits.

The work that we have been developing in the field of research data management in INESCTEC (institution which we belong), leads us to believe that there

is growing knowledge and interest in developing image data management practices, on the part of the research community. First, more and more researchers look to us to help them in the process of organization (through the development of data management plans), of description (through the use of metadata models), and of deposit/sharing (through the choice of repository and assistance in the filing process). Second, more researchers are concerned about reuse issues of their images, especially authorship issues, showing interest in knowledge about license and rights issues. The possibility of attributing DOI to the data has been an incentive to share. Finally, researchers almost always admit that a data management policy would be very helpful. This policy does not arise for lack of time or in-depth knowledge. In fact, some practices are beginning to be applied in projects, in order to initiate, even if superficially, behaviors that lead to good image management. Therefore, we believe that this study is necessary for the definition of best practices of research data management, namely image management, as it allows establishing a scenario and, from there, define a set of guidelines and offer tools for the image management. In addition, this study makes it possible to strengthen the existing state of the art, through analyzes of real practices.

In this sense, future work includes the development of a metadata model for the description of data in image format. This model will present specific descriptors to data in image format and some elements will be associated with controlled vocabularies. We believe the use of these vocabularies will improve the description and facilitate the work of the researchers while describing. This model will emerge from a development and evaluation process that seeks to meet the expectations of the research community, regardless of its domain, always taking into account the good practices of RDM. This metadata model is already in an advanced stage of development and will soon be presented to the scientific community. In addition, we will propose a document with guidelines that will act as a guide for researchers, where they can find tips for managing their images, from the planning to the sharing stage.

Acknowledgements Joana Rodrigues is supported by research grant from FCT - Fundação para a Ciência e Tecnologia: PD/BD/150288/2019. Thanks to Miguel Fernandes who conducted the initial interviews and was involved in the construction of the interview script.

References

[1] Catharine Abell. "The epistemic value of photographs". In: *Philosophical Perspectives on Depiction*. Ed. by Catharine Abell and Katerina Bantinaki. Oxford University Press, 2010, pp. 82–103. ISBN: 9781351517577. DOI: 10.4324/9780203790762.

[2] J. Aguiar Castro et al. "Involving data creators in an ontology-based design process for metadata models". In: *Developing Metadata Application Profiles* (2017), pp. 181–213. DOI: 10.4018/978-1-5225-2221-8.ch008.

[3] Alex Ball. *Review of Data Management Lifecycle Models*. Tech. rep. Bath, UK., 2012, p. 15. URL: https://purehost.bath.ac.uk/ws/files/206543/redm1rep120110ab10.pdf.

[4] Marcus. Banks. *Using visual data in qualitative research*. Ed. by Uwe Flick. 1st. London: SAGE Publications, 2007, p. 161. ISBN: 9780857020260.

[5] Kristin Briney. *Data Management for Researchers: Organize, Maintain and Share your Data for Research Success*. Exeter, UK: Pelagic Publishing, 2015, p. 189. ISBN: 978-1-78427-030-8.

[6] CASRAI. *Data*. Mar. 2022. URL: https://casrai.org/term/data/.

[7] Andrew Martin Cox and Winnie Wan Ting Tam. "A critical analysis of lifecycle models of the research process and research data management". In: *Aslib Journal of Information Management* 70.2 (2018), pp. 142–157. ISSN: 20503814. DOI: 10.1108/AJIM-11-2017-0251.

[8] M. Fernandes, J. Rodrigues, and C. T. Lopes. "Management of Research Data in Image Format: An Exploratory Study on Current Practices". In: *International Conference on Theory and Practice of Digital Libraries - Digital Libraries for Open Knowledge* (2020), pp. 212–226. DOI: 10.1007/978-3-030-54956-5_16.

[9] P. Brian Heidorn. "Shedding light on the dark data in the long tail of science". In: *Library Trends* 57.2 (2008).

[10] Dirk Huds. "The impact of photography on society". In: *Our Times* (2019). URL: https://ourpastimes.com/the-impact-of-photography-on-society-12377030.html.

[11] Hanna Kallio et al. "Systematic methodological review: developing a framework for a qualitative semi-structured interview guide". In: *Journal of Advanced Nursing* 72 (2016), pp. 954–2965. DOI: 10.1111/jan.13031.

[12] A.L. Lacerda. "A fotografia nos arquivos: produção e sentido de documentos visuais". In: *História, Ciências, Saúde* 1 (2012).

[13] C.L. Palmer and et. al. "Site-based data curation based on hot spring geobiology". In: *PLoS ONE* 12.7 (2017).

[14] European Commission; Directorate-General for Research and Innovation. "Guidelines on FAIR Data Management in Horizon 2020". In: (2016).

[15] M. A. Sandweiss. "Image and Artifact: The Photograph as Evidence in the Digital Age". In: *Journal of American History* 94.1 (2007), pp. 193–202. ISSN: 0021-8723. DOI: 10.2307/25094789.

[16] M. Saunders, P. Lewis, and A. Thornhill. *Research Methods for Business Students*. 6th. Pearson Education Limited, 2012.

[17] Structural Reform Group. "DDI Version 3.0 Conceptual Model". 2004. URL: https://ddialliance.org/sites/default/files/Concept-Model-WD.pdf.

[18] Carol Tenopir et al. "Data sharing by scientists: Practices and perceptions". In: *PLoS ONE* 6.6 (2011), pp. 1–21. ISSN: 19326203. DOI: 10.1371/journal.pone.0021101.

[19] UK Data Archive. *Research data lifecycle*. 2019. URL: https://www.ukdataservice.ac.uk/manage-data/lifecycle.aspx (visited on 12/19/2019).