Guidelines for standardized practices in the use and production of images in research data management context

Orientação para práticas normalizadas na utilização e produção de imagens no contexto da gestão de dados de investigação

Joana Rodrigues

Faculdade de Engenharia da Universidade do Porto and INESC TEC
jsrodrigues@fe.up.pt

Abstract

This paper identifies guidelines to help researchers using or producing images in a research context. Throughout the various experiences carried out during the last years, it became apparent that, although they considered images to be important parts of their projects, most researchers had never reflected on the subject more consistently, nor had they thought about formalizing practices to ap- ply to their projects. The aim of these guidelines is precisely to offer some tips, based on practices mentioned in the literature and consistent with the principles of research data management. The recipients are intended to be researchers, regardless of their field of research and their scientific maturity and experience. I believe that there is no rigid standard of guidelines, but establishing some habits can benefit the scientific process and open up greater debate and scientific production on the subject.

Keywords: Images, Research Data Manangement; Research Lyfe Cycle

Resumo

Este artigo identifica orientações para ajudar os investigadores que utilizam ou produzem imagens em contexto de investigação. Ao longo das várias experiências realizadas ao longo dos últimos anos, tornou-se evidente que, apesar de considerarem as imagens como partes importantes dos seus projetos, a maioria dos investigadores nunca tinha refletido sobre o assunto de forma mais consistente, nem tinha pensado em formalizar práticas para aplicar aos seus projetos. O objetivo destas orientações é precisamente oferecer algumas dicas, baseadas em práticas mencionadas na literatura e coerentes com os princípios da gestão de dados de investigação. As orientações destinam-se a investigadores, independentemente da sua área de investigação e da sua maturidade e experiência científica. Acredito que não exista um padrão rígido de orientações, mas o estabelecimento de alguns hábitos pode beneficiar o processo científico e abrir maior debate e produção científica sobre o tema.

Palavras-chave: *Imagens, Gestão de Dados de Investigação; Ciclo de Vida da Investigação*

1. Contextualization

Images in research can be important parts of the research process, whether they are data or metadata. Images should not be captured or chosen in such a way as to be manipulated and used inappropriately, even if it is to serve the research objective a priori [2, 6].

Depending on the assigned purpose, images can assume different roles and importance, can be produced or used at different times in the research process, and can be associated with different collection, editing, and analysis methodologies [2].

These guidelines were designed taking into account the practices highlighted in the literature, personal experience and always taking into account the principles that govern the management of research data. It is important to emphasize that these are not general guidelines for the use and production of images, but rather guidelines for the use and production of images in a research context, based on data management.

The guidelines will be presented along with some examples. It's important to note that these are only some of the possible examples to elucidate the guidelines given at the time.

2. General Guidelines

There is no detailed standard for the production and use of images in research. However, this work made it possible to determine which guidelines should be followed to guarantee the integrity of the images and the investigative process.

Images can take an analog or digital format. Most of the images currently produced are digital (very motivated by technological development), however analog images continue to be produced in different contexts and their importance should never be overlooked.

In a scientific context, when creating images it is of great importance to guarantee their quality. In the case of digital images, it is necessary to check that the image capture instrument is properly calibrated and aligned with the object. In the case of analog images, it is necessary to ensure that they follow the defined production requirements, taking into account the context for which they are intended [5].

It is possible to say that digital images have challenges that are different from analog images, as the immense possibilities of processing, analyzing and editing images, based on increasingly advanced tools and software, call attention to other precautions. For this reason, in the use and manipulation of images, especially digital images, there are some ethical considerations that should guide their use [4]. According to Cromey [4], these considerations can be seen as assumptions or guidelines, namely:

- Scientific images can be compromised by improper manipulations;
- Manipulations of digital images should always be performed on a copy of the original image;

- Simple adjustments to the entire image are acceptable;
- Image cropping is usually acceptable, as long as it does not alter the content of the image;
- Images that are to be compared to each other must be captured under identical conditions and undergo identical processing;
- Manipulations performed only on specific areas of the image are questionable;
- The use of filters to improve image quality is not recommended;
- Cloning or copying objects from other images and composing a new image is very questionable;
- Avoid using image formats with compressions that affect image quality;
- Intensity measurements should be performed on a uniformly processed image calibrated to recognized standards;
- Resolution and magnification are important;
- Caution is required when changing the size, in pixels, of the digital image.

3. Guidelines for Research Image Management

In addition to the more technical and general issues, such as those presented above, the investigation process that involves the use and production of images must consider other guidelines for the management of these images, especially in the sense of following the guidelines indicated in the scope of Research Data Management [2].

It is really crucial to standardize processes, establish methods, and ensure uniformity in image management, as this will make it possible to avoid failures and redundancies [1].

Throughout the work developed in the last years, the research life cycle has been used several times as a tool for analysis. In the case of these guidelines, it was considered that this life cycle could be an ally in order to systematize the information that is intended to be conveyed since it groups the various guidelines by life cycle stages. Figure 1 presents the research life cycle used to establish and organize these guidelines.

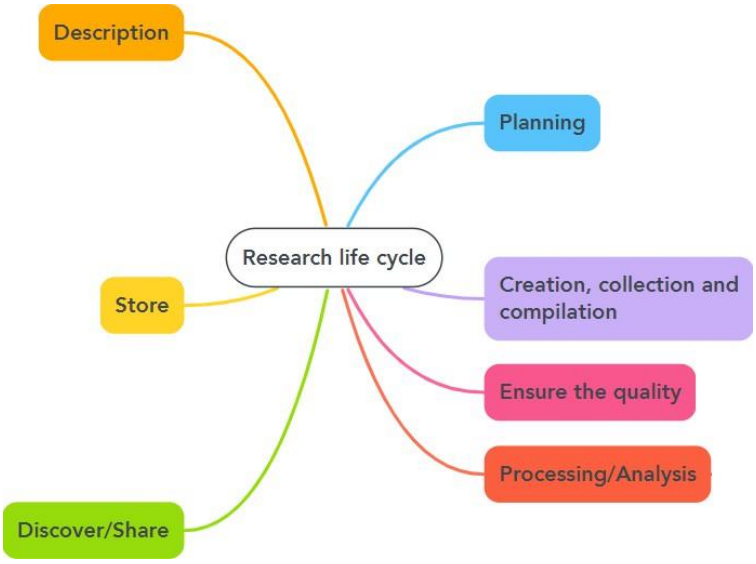

Figure 1) Research Life Cycle

In order to create a set of properly structured guidelines based on good practices, I followed suggestions from the Consortium of European Social Sciences Data Archives (CESSDA)[1], in its 2019 document "Adapt your Data Management Plan: A list of Data Management Questions based on the Expert Tour Guide on Data Management" [3]. This document offered some general ideas that allowed me to organize these guidelines in the context of the images that we consider important to mention.

3.1. Planning

The planning phase is essential whenever we talk about data management, regardless of the type of data [1, 6]. As far as guidelines for image management are concerned, this should be supported by the use of a planning document. This should be seen as a fundamental tool for organizing and structuring all the procedures inherent in the use and production of images throughout the research process, i.e. from the moment it is planned until it is completed.

There is no need for a document that applies exclusively to images. The document can be generic for the project but include a section dedicated to image management. A good research practice is the use of Document Management Plans (DMP), where the entire process relating to the project and its data is documented. The DMP can include a section where image management is explained.

This planning document (whether it is a DMP or not), in addition to explaining the role of data, including images, should contain a set of good research practices. The planning document should take place at the beginning of the project to guide all the following stages. By carrying out this procedure, the planning phase of the life cycle is guaranteed.

Specifying the component of the document relating to the images, it is important to note that, in addition to all the information on the phases of the life cycle of the images, this document must guarantee the presentation of general information about the images. This information is detailed in Table 1.

Table 1) Guidelines for the planning phase

Guidance *General information about the images*	**Example**
Origin of the images	Captured in the laboratory via microscope
People involved	Principal researcher and research assistants
Definition of the project where the images are inserted	Experience "2.0" within the scope of the "XPTO" Project
Copyright issues	Copyright: exclusive right of the author to reproduce the resource. License: CC BY
Financing	Funded by the research foundation "XYZ"
Contacts	example@project.com
Costs	200 €
Dates	May 2022 - September 2022
Purpose of images	Elaboration of a case study using direct observation of a phenomenon via microscope, with image capture
Type of images	Microscopic images
Format of images	*.png

[1] http://www.cessda.eu/

Edition	The images were cropped to measure 210mmx148mm
Quality	Resolution: 1080p
Sample	45 images

The most specific aspects of each image or set of images must be specified later in the detailed description of the project (e.g., collection methodology, analysis techniques, processing tools, among others).

3.2. Creation, Collection and Compilation

Issues relating to the creation, collection, and compilation of images should also be considered and followed according to certain guidelines.

In addition to the more technical aspects of images, this phase of the life cycle must include image organization and storage issues. When one or more images are taken, various details need to be documented. Table 2 shows the guidelines that should be followed in this regard:

Table 2) Guidelines for the creation, collection and compilation phase

Guidance	Guidance description and examples
Formal characteristics of images	name, type, size, dimensions and format of images (e.g., "Figure A", photograph, 68,1 KB, .*jpg, W: 538 pixels x H: 473 pixels)
Organization of images	- Location where the images will be stored, - Structure of folders and files related to the organization of the images
Collection of images	*In the case of image use:* e.g., obtained through the "XPTO" data repository *In the case of images production:* e.g., obtained by researcher X, in Furnas do Enxofre, on Terceira Island, between April 22, 2019 and April 24, 2019, using the Canon EOS Kiss X2 Reflex 12.4 camera, in rainy weather conditions, with a temperature of 11ºC, wind 18k/h and relative humidity of 52%).
Purpose of images	Specify the purpose of the images in the study. e.g. “The images will be used to verify the facial expressions of the patients involved in the study”.
Reuse of images	In the case of reusing images, specify where the images are located, how their search is planned, what the access terms and conditions are, what the expected process and time span for obtaining access to the images are, and what the expected cost relative to this access.

3.3. Processing/Analysis

The way an image is processed and analyzed is decisive in obtaining reliable and trustworthy results. It is not enough to collect an image, it is necessary to determine the procedure and method inherent in the way it will be processed and analyzed. Table 3 sets out the guidelines on this topic.

Table 3) Guidelines for the processing/analysis phase

Guidance	Example
Definition of a strategy for the development of image versions	All images must be backed up, assumed as "backup_figure x". Each time an image is edited, it must be saved as a copy (always keeping the original) and renamed, by including a version identifier (example: "sceloporus magister_version 2")
Who will be responsible for their maintenance and elaboration of criteria for the distinction of versions	Principal researcher
Determination of issues relating to interoperability, namely the use or not of ontologies and terminologies	Use of the OpenCyc ontology
Use of software, and coding issues	Use of the Pix4D (professional photogrammetry and drone mapping)
Determination of image analysis approaches, methods and procedures	Content analysis using qualitative methods

3.4. Ensure the Quality

Quality issues have very specific characteristics in images. If, on the one hand, the content of the image is essential, more technical characteristics, such as the format in which it is stored or even its dimensions, can be essential in guaranteeing the quality of an image. Quality, with the use of editing, can be related to trying to solve some problem with the image or to improve some characteristic. Quality information must be taken into account. Table 4 presents the guidelines for this topic.

Table 4) Guidelines for the ensure the quality phase

Guidance	Guidance description and examples
Editing and image quality issues	- Identify if there were problems with the image that require editing if these problems are not identified, and explained the reason for the editing; - Mention important quality issues for the study (e.g. resolution, color, brightness, saturation, etc.)

Determine the minimum quality required for images	The minimum quality required for images must be determined in terms of origin, content, size, format and methods
Explain how the quality of the images will be evaluated	Determine in advance which criteria will define whether the image has sufficient quality (example: clear identification of all elements present in the image, correspondence of the colors present in the image with reality, specific image format, etc.)

3.5. Store

There are various data storage solutions, but the way and method in which data is stored is decisive for aspects such as organizing data, retrieving information and safeguarding resources, for example. With this in mind, it is important to determine the practices to follow when it comes to storage. Table 5 presents guidelines with corresponding examples.

Table 5) Guidelines for the store phase

Guidance	Example
Define where and for how long the images will be stored	The images will be stored in the cloud on the project's resource management platform. Backups will be stored on the project's external disk. The images will be stored for up to 6 months after the total completion of the project, except those present in publications, communications, and platforms such as repositories and directories (for these see the last line of this table).
Specify how and for what periods backups will be carried out	Image backups must be carried out at the same time as the original images are stored in the cloud. Every time a new version of the image is made, a new backup must be made, to be included on the external disk where the other backups are located.
Determine an image recovery strategy in case of loss	If the original image is lost, it must be backed up and a new copy made to the backup disk.
Clarify how access to images will be managed and how data security will be ensured (contingency plan if necessary)	Access to each of the project elements will be determined. The principal investigator and research assistants will have access to all the images and will be able to manage the sharing of some of the resources with other elements. The other elements will have access restrictions. Access will be guaranteed using passwords.
Location and method of storing the images after the project has been completed	After the completion of the project, images that are not part of project reports or publications/communications arising from the project will be deleted (6 months after the total completion of the project). Images that are present in publications, communications, datasets or platforms such as repositories and directories will be kept, always indicating the copyright and licenses for access and use.

3.6. Description

For this specific topic (description), within the guidelines, it was realized that it is essential to explore the practices to follow a little more. The specification of what documentation will be produced in relation to the images could be used as a guideline, as well as what description characteristics will be applied, namely the use of metadata and metadata models. However, it is important to go beyond this and reinforce the idea of the dual role of images: as data and as metadata. This information is described in Table 6.

Table 6) Guidelines for the description phase

Images as data	Images as metadata [7]
The guidelines to be followed should be guided by the good practices standardized by the principles of the RDM	Images should be seen as the agent of description. That is, the value to be included in the descriptive fields.
The metadata model proposed for the description of images, although not a fixed and watertight tool, can be fundamental in the description of image data, especially when it comes to specific characteristics of this typology, such as issues associated with quality, editing, and scale, for example.	If, for textual and numerical values, the indication is a description oriented towards due interpretation, using clear language, focusing on the main information to be presented and without interpretative duality, the same applies to image values. It must be ensured that the images or set of images allow a true reading of what is intended. It is not intended that the image is just a decorative element that accompanies the data, it must be able to be "read" and understood
The use of controlled vocabularies should be seen as an important guideline, as it makes it possible to establish a set of vocabulary elements that restrict the description values, making it possible to standardize the description and improve interpretation.	The image must be accompanied by other values, such as text, whenever necessary
When it comes to images, the description of the data must go beyond what is visible in the present content of the image, i.e. the formal analysis must be accompanied by the analysis of the semantic expression.	The person performing the description must be able to distance themselves from their condition of an unequivocal connoisseur of the image and put themselves in the role of someone viewing it for the first time. They must therefore be able to check that the image contains all the elements necessary for its descriptive function and add other elements whenever they consider that the imagery is not sufficient

3.7. Discover/Share

Increasingly, data discovery and sharing practices are a concern for researchers. Scientific pro- duction aims to disseminate it to raise new hypotheses and test ideas. Today, properly cited data or publications are fundamental to the reputation of researchers, their projects, and their research centers. The dynamics of open science promote publication practices that take into account ethical issues, copyright, and transparency, for example, through the use of DOI's and specific reuse and citation conditions. Table 7 shows the guidelines identified as important for this topic.

Table 7) Guidelines for the ensure the discover/share phase

Guidances for discover	Guidances for share
Guarantee that the images will be archived in a reliable data repository	Ensuring the use of a persistent identifier for images
Specification of how the images will be made available (open access or not)	Specifying the formats in which the images will be shared
Number of images to publish (all or just a few?)	Determining the need to use specific software to process the images? If yes, define whether this can be deposited/shared together with the images,
Identify the licenses required for images	Determining how images should be cited when reused
Definition, if necessary, of an embargo period for (all or some of) the images	Specify the existence of agreements or access restrictions that need to be considered
Determine whether there is a risk of delaying the publication/availability of images (in whole or in part). If so, explain what needs to be done to avoid this.	Specify the existence of legal/ethical restrictions that prevent the publication and share of images. Explain whether these restrictions will require action to be taken before images are made available.

3.8. Image Protection

It became clear that the life cycle used should include a phase dedicated to the protection and privacy of data in image format. As this is an emerging issue and the subject of many regulations, it was decided to include a set of guidelines to guarantee this component in image management:

- define whether or not it is mandatory for an ethics committee to approve the use and production of images,
- determine whether it is necessary to request informed consent for the use and production of the images, if so how this permission will be obtained, and how the images will be organized and stored. In the case of informed consent, it must follow the model: "Informed, Free and

Clarifies Consent to Participate in a Research Project in accordance with the Declaration of Helsinki[2] and the Oviedo Convention"[3]. This consent allows the participants to attest that they agreed to participate in the study and that they were guaranteed confidentiality, the exclusive use of the data collected, and anonymity, promising never to publicize the identification of participants.

- specify how access to images that constitute personal and sensitive data will be controlled,
- define how the security of this sensitive data will be guaranteed,
- provide for the existence of a requirement to notify the privacy officer of images that contain confidential information or special treatment,
- clearly specify if there are images that, due to their privacy conditions, require special treatment or access limits during and after the project,
- how images that constitute personal or sensitive data will be protected during and after the project,
- determine how permissions and restrictions will be reinforced,
- specify whether there are any ownership/copyright issues associated with the images, including clarifying what permission is needed to collect and reuse images and whether, in the case of reuse, rights will be transferred to another organization,
- provide for any restrictions that may occur in terms of privacy.

[2] http://epidemiologia.med.up.pt/pdfs/Helsinq.2013.pdf
[3] http://dre.pt/pdf1sdip/2001/01/002A00/00140036.pdf

4. Conclusion

Images in scientific research play a fundamental role in communicating and analysing data, and it is essential to ensure that their production, use, and management follow well-established guidelines. Standardising these practices contributes to scientific results' transparency, reproducibility, and integrity, ensuring that images are used ethically and rigorously.

Throughout the article, guidelines have been presented that aim to standardise the treatment of images in the context of research data management, considering not only their collection and editing but also the way they are stored, archived, and interpreted. Adopting these practices reduces ambiguities, avoids inappropriate manipulation, and ensures that the images are a reliable reflection of the data obtained. In addition, standardization helps researchers and institutions to establish clear criteria for the production and reuse of images, promoting the credibility of research and facilitating scientific collaboration.

It is important to emphasise that the implementation of these guidelines must be accompanied by continuous reflection on the challenges and technological advances in the field of data management. The evolution of image processing and analysis tools requires constant updating of standards and recommended practices, ensuring that they continue to meet the needs of the scientific community. Thus, standardising the use of images in research should not only be seen as a technical requirement but as a commitment to the quality, ethics, and reliability of scientific data. The application of these guidelines reinforces the importance of a more transparent and accessible science, contributing to the advancement of knowledge responsibly and sustainably.

5. Bibliographic References

[1] AKERS, Katherine and Doty, Jennifer (2013) “Disciplinary differences in faculty research data management practices and perspectives”. *International Journal of Digital Curation,* n 2

[2] ARCHARYA, Tinku and Ray, Ajoy (2005) “Image Processing: Principles and Applications. *John Wiley Sons, Inc.* New Jersey, ISBN: 978-0-471-71998-4

[3] CONSORTIUM OF EUROPEAN SOCIAL SCIENCES DATA ARCHIVES (2019) “Adapt your data management plan: A list of data management questions based on the expert tour guide on data management”. Retrieved from https://dmeg.cessda.eu/ Data-Management-Expert-Guide/1.-Plan/Adapt-your-DMP-Part-1

[4] CROMEY, Douglas (2010) “Avoiding twisted pixels: Ethical guidelines for the appropriate use and manipulation of scientific digital images”. *Science and engineering ethics,* n 4, pp 639-67

[5] CROMEY, Douglas (2013) “Digital images are data: And should be treated as such”. *Methods in molecular biology (Clifton, N.J.),* n 1, pp 1-27

[6] PRYOR, Graham (2018) “Why manage research data?” *Managing Research Data.* Cambridge University Press, pp 1-16

[7] RODRIGUES, Joana and Lopes, Carla Teixeira (2023) “Images as metadata: A new perspective for describing research data”. *Journal of Library Metadata,* n 23, pp 87-101