



DPF Manager

The open source
COMMUNITY

DPF MANAGER

Surveys and conclusions

Project acronym: PREFORMA

PREFORMA - Future Memory Standards

PREservation FORMAts for culture information/e-archives

EC Grant agreement no: 619568

EC Call ID: FP7-ICT-2013-11



INDEX

1. PREFACE	3
1.1. SURVEY	3
1.2. STATISTICAL SAMPLE	4
1.3. OVERALL RESULTS	6
2. DETAILED RESULTS	7
2.1. DEFINITION OF DEFAULT MODELS	7
2.1.1. SMALL MEMORY INSTITUTIONS	7
2.1.2. MEDIUM MEMORY INSTITUTIONS.....	8
2.1.3. LARGE MEMORY INSTITUTIONS	8
2.2. COMPARATIVE	8

1. Preface

The PREFORMA project did conduct a series of workshops to determine the needs of the memory institutions, producing a set of requirements for the suppliers to deliver on.

To take that work further, and to get a better understanding of how different memory institutions tackle the digital preservation challenge, we designed and distributed a questionnaire with specific questions about type and volume of digital assets, current capabilities and limitations, and future needs.

The results from the analysis of the answers have informed our technical and functional design decisions.

In this survey we wanted to collect some data from the memory institutions to be able of define the current state of preservation in real environments:

- Amount of information processed.
- Type of information.
- How they deal with the known problems related to the files they used to manage.
- Techniques that they use.
- Software used.
- Knowledge in image preservation.
- And other relevant information.

The aim of this document is to present the results of those surveys, for that purpose this document contains the following parts:

- First of all we show the survey sent to every memory institution.
- Second, we expose the statistical sample.
- Then, we present the overall conclusions extracted from the results.
- Fourth, we define a generic small/medium/large memory institution thanks to the data extracted.
- Finally we compare results taking in considerations the size of every memory institution, where the size of every memory institution has been set by itself.

1.1. Survey

This survey has been sent by e-mail to memory institutions all over the world, including the PREFORMA partners, and as of March 3rd 2015 we have collected more than 50 answers.

The questionnaire will stay online during the prototyping phase of the project, not only because each answer gives us a bit more information about how memory institutions operate, but also because every participating memory institution can be a potential early adopter of our solution.

This is the URL where the memory institution can answer and collaborate with the development of the DPF Manager:

<https://docs.google.com/forms/d/1YiXrN70xae9JnEPxsvGNMofF8N8iMxfKvtuHU6mXORM/view/form>

1.2. Statistical sample

Below is the list of over 50 memory institutions from 15 different countries that have completed the questionnaire:

Country	Memory Institution
Austria	Kärntner Landesarchiv
Belgium	KIK-IRPA
	Archives Générales du Royaume
Bolivia	Fundación Cinemateca Boliviana
Canada	Bibliothèque et Archives nationales du Québec

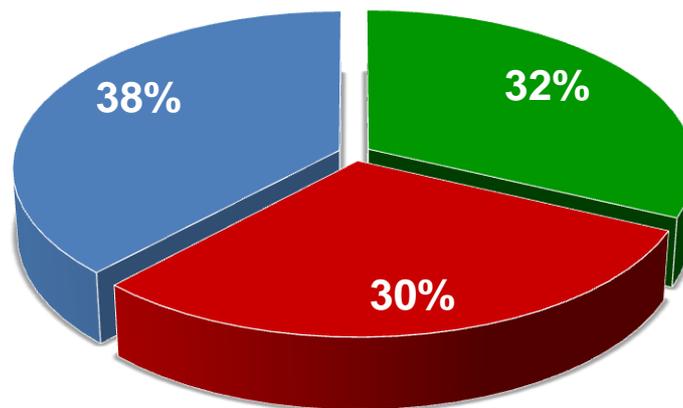
	City of Regina Archives
Chile	Seremi de Educación Región del Biobío
Estonia	National Library of Estonia
Hungary	Budapest City Archives
Israel	National Library of Israel
Ireland	Irish Jesuit Archives
	Local Government Management Agency
Luxemburg	Archives nationales de Luxembourg
Peru	Archivo Histórico de Límites
Spain	Archivo Real y General de Navarra (AGN)
	Archivo Real Cancillería de Valladolid
	Archivo Central del Ministerio de Educación, Cultura y Deporte
	Arxiu de Tarragona
	Arxiu Municipal. Ajuntament de Girona
	Universitat de Girona. Biblioteca
	Archivo nacional Cartagena
	Archivo Municipal de León
	Biblioteca Armand Cardona Torrandell
	Universitat Autònoma de Barcelona
	Arxiu municipal de Salt
	Universitat de Barcelona
	INSPAI, Centre de la Imatge de la Diputació de Girona
	Archivo General de la Universidad Complutense de Madrid
	Instituto Español de Oceanografía
	Archivo General ACdP-CEU
	Biblioteca de Castilla y León
Sweden	National Archives of Sweden(RA)
	National Library of Sweden
	Riksarkivet (Nat Arch of Sweden)
Switzerland	BJ institute
	Dokumentationsbibliothek St. Moritz
	ETH Zurich, ETH-Bibliothek
	Fachlabor Gubler AG
	Kunsthalle Basel
	Sammlung René Gardi
	Schweizerische Gesellschaft für Volkskunde
	Staatsarchiv Basel-Stadt
United Kingdom	Bedfordshire Archives
	Houses of Parliament
	Research Support Librarian Goldsmiths, University of London
	Royal Northern College of Music
	Swansea Libraries

The Stanley & Audrey Burton Gallery
The Theatres Trust
WS Society

These 50 memory institutions have been classified by themselves as small, medium or large, with this almost equal distribution result:

Memory Institution Size

■ Large ■ Medium ■ Small



1.3. Relevant results

The majority of the Digital objects that the memory institution curate comes from their own digitalization or external digitalization. Consequently, the integration in the OAIS model should be done in digitalization time and transfer time respectively in order to validate a file for a long term preservation purpose.

Even though all the memory institutions know the OAIS reference model and its benefits, the majority of them don't apply this model. The main reason seems to be a problem of resources. Therefore as much big a memory institution is, with more resources, more they apply the OAIS model.

Little and medium memory institutions that cannot afford the cost of a proper digitalization process, adequate software or hardware and employees with the sufficient knowledge in preservation to use these tools cannot apply the OAIS model.

Almost 60% of the memories polled don't apply OAIS. A way to help them is providing tools to validate their process of ingestion and migration with fills with their requirements; cheap software that could be used for a non-expert users.



One of the most alarming conclusions extract from the results of the survey is the lack of knowledge about the sustainability factors of digital formats even in large memory institutions.

Most of them are not concerned about the influence of storing the metadata outside the file, storing compressed images or other technical aspects like tonality or color space.

With the proper validation tool and working to develop a suitable standard for preservation process, those problems could be corrected.

Another relevant aspect is the ignorance in the impact of patents inside the file, small and even medium memory institutions don't see this aspect as a danger for their archives. But this can be a huge problem when we talk of large memory institutions, with many digital objects to preserve.

Similar to with the OAIS model, almost 80% of the memory institution polled are aware of the importance of well formatted file structure, but approximately only 30% of them use software to validate file format structures.

One of the conclusions is that actually there is not a tool to validate files suitable for these memory institutions who now are avoiding this process. Therefore, the existence of a free or just cheap tool, with a good documentation and simple use, adapted to the knowledge of the users will push these good preservation practices.

2. Detailed Results

2.1. Definition of default models

Thanks to the surveys received from the memory institution we are able to define three different models of memory institution for the three sizes defined: small, medium and large.

2.1.1. Small Memory Institutions

Based on the surveys done, small memory institutions have these following characteristics:

- They locate approximately 14.000 digital images.
- Incorporate 2.500 images per year.
- They manage photographic prints, negatives, paintings/drawings and also books and texts.
- The majority of the images have been added by their own digitalization.
- Even though they know the OAIS model they don't apply it.
- Even the costs in general are important, they are more aware of the cost of preservation and storage than the cost of the software.
- They handle TIFF and JPEG images .
- They use to store metadata with simple mechanism like stored in the same file o just the filename as the important metadata.

- They are not aware of the consequences of some technical aspects for the preservation of the images like color reproduction or the impact of patents.

2.1.2. Medium Memory Institutions

Based on the surveys done, medium memory institutions have these following characteristics:

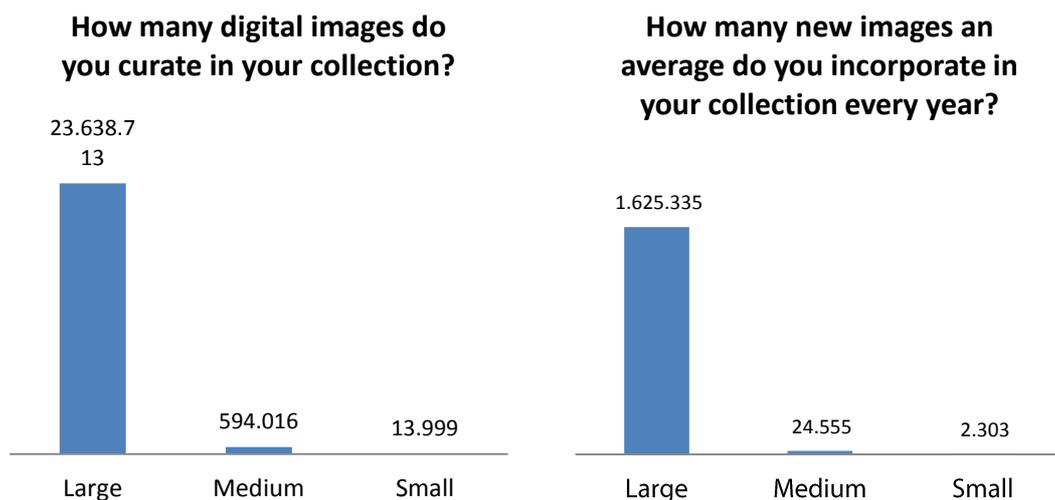
- They locate approximately 600.000 digital images.
- Incorporate approximately 25.000 images per year.
- They manage photographic prints, paintings/drawings, and also books and texts.
- The majority of the images have been added by extern digitalization.
- They know the OAIS model and a significant part of them apply it.
- They are more aware of storage and also the software cost.
- They handle TIFF mostly.

2.1.3. Large Memory Institutions

Based on the surveys done, large memory institutions have these following characteristics:

- They locate more than 23,5 million digital images.
- Incorporate more than 1,5 million images per year.
- They manage books/texts, paintings/drawings and photographic prints.
- The majority of the images have been added by their own digitalization.
- They know the OAIS model and they apply it.
- They are more aware of the cost of digitalization infrastructure cost, the cost of preservation processing and the color reproduction in the future.
- They handle TIFF mostly.

2.2. Comparative

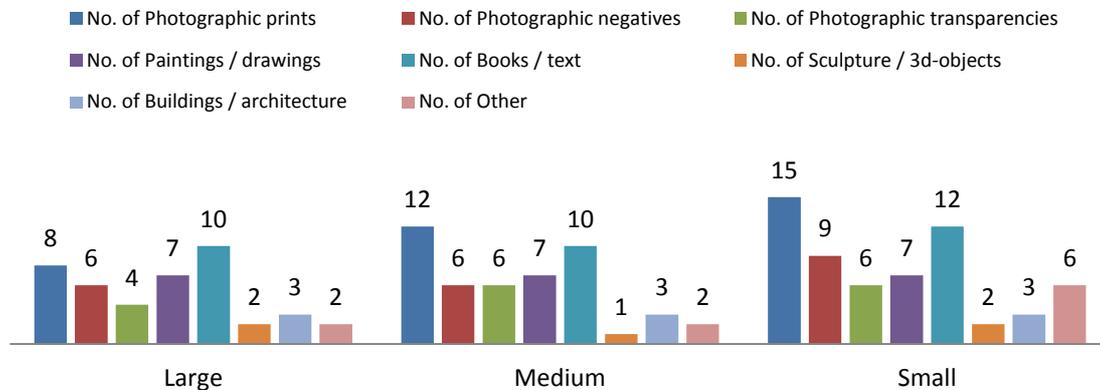


There is a huge difference between the three sizes, while medium are more than 40 times bigger than the small ones, the big ones are almost 40 times bigger than the medium.



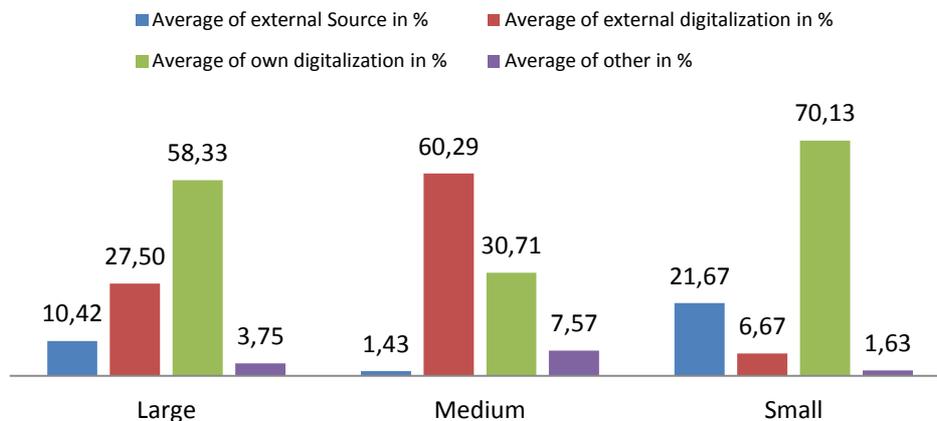
These differences make three scenarios with different needs each one. If we take a look at the speed their collection grows, the bigger the memory institution the faster they grow.

What kinds of objects do you depict?



In the three cases, the most popular objects are Photographic prints and Books/texts digitalization doesn't matter the size of the memory institution, the behavior is similar in the rest of objects.

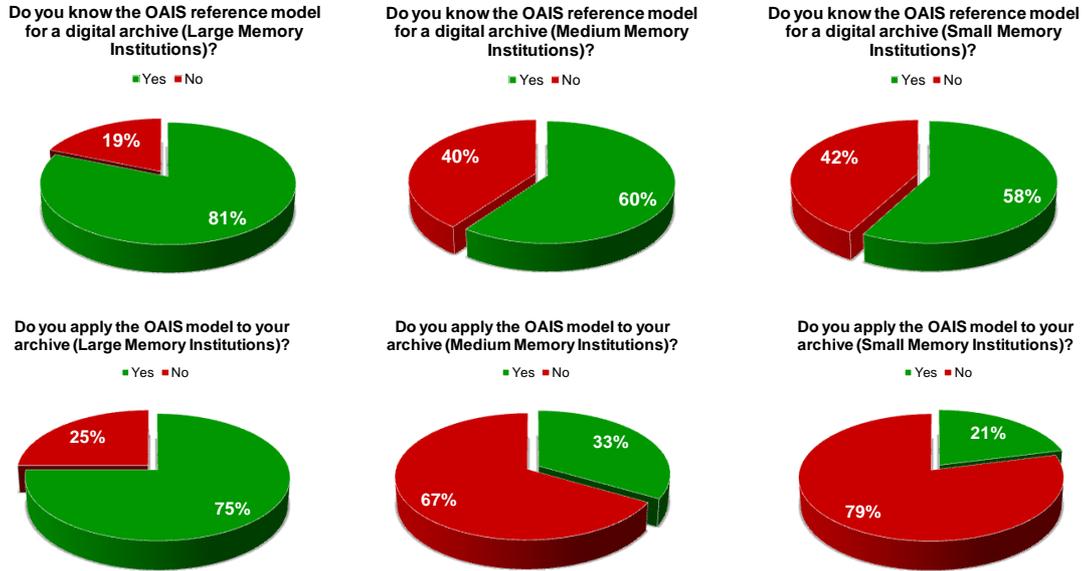
What are the origins of your image archive and in which percentage



Here we have one of the first significant differences: the evolution of the origins of the images as the size increases. For small institutions the most important part of the images are obtained by themselves, but on medium size memory institutions the used to get the images by external digitalization, but finally for large size memory institutions the start again to digitalize by themselves.

It is caused because of the need from the large memory institution to control the quality of the inputs, and for the cost of the preservations procedures. In the case of the small memory institutions it is because of the costs in general, they cannot afford to pay external services.

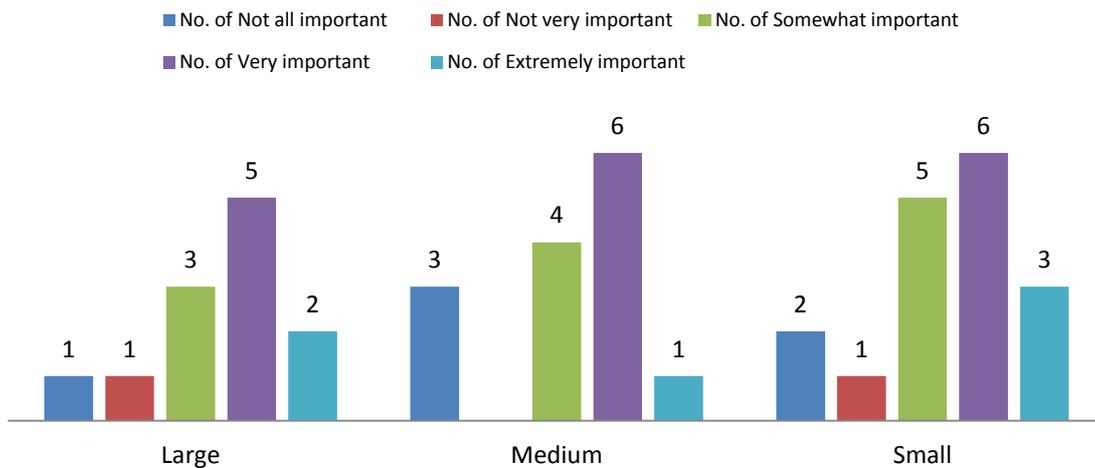
The case of the medium memory institutions is special, they have enough budget to outsource some process and they don't have such a huge number of images that make this outsource process unviable.



Here arises one significant evolution while the size of the group grows. For small memory institutions, despite they know the OAIS model, they don't apply it in their archives, but as the size of the institution it's increased, they are more aware of the importance of its application.

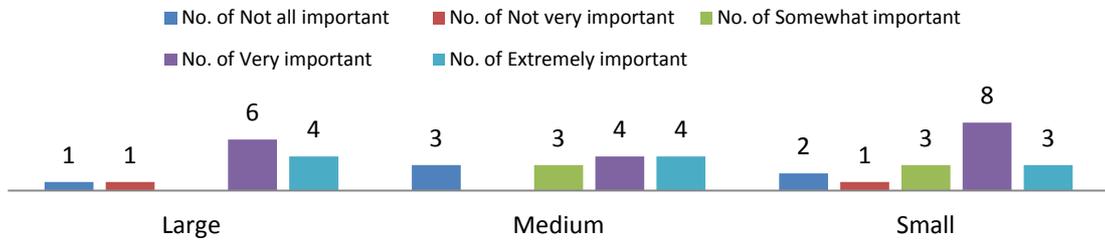
Therefore, for medium size institutions there are more not applying than applying it, even though the difference is less. But when we take a look at the large memory institutions, we realize that the answers has been turned around, the institutions that apply the OAIS model are more than the double than the rest.

How important is the software cost for your institution (being 1 not at all important and 4 very important)?



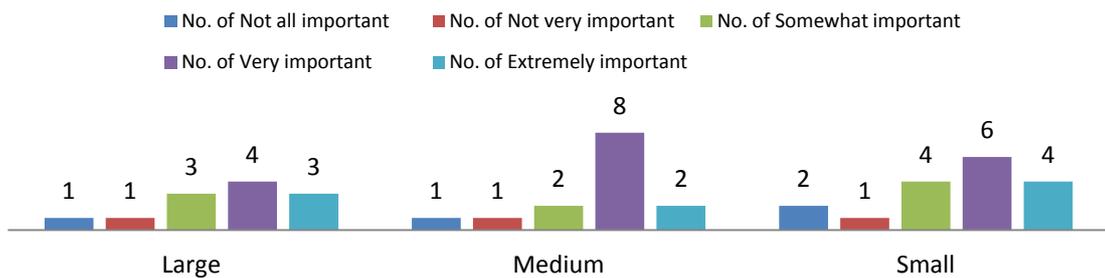
In the three cases the software becomes a very important cost, therefore an open source solution to help them in the archiving process has to be welcome.

How important is the digitalization infrastructure cost for your institution (being 1 not at all important and 4 very important)?



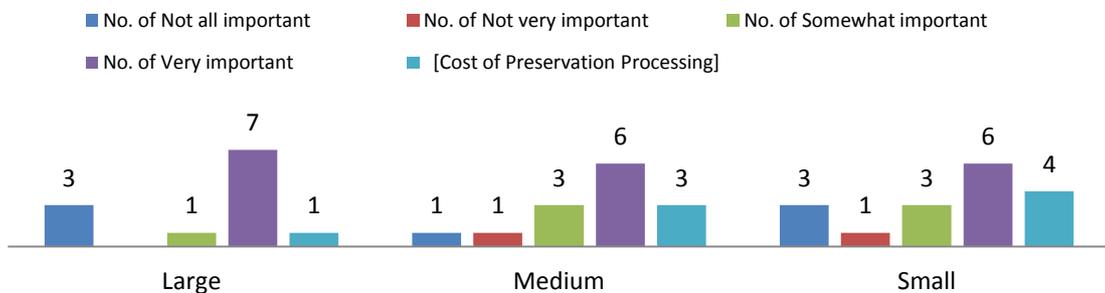
The cost, under the point of view of the memory institutions, is a point in common regardless from the size. It will be good to highlight that even though the three sizes have their own problems, in this case, the answers are very similar. In this case, the three of them agree on the importance of the storage cost, infrastructure and preservation processing cost.

How important is the storage cost for your institution (being 1 not at all important and 4 very important)?



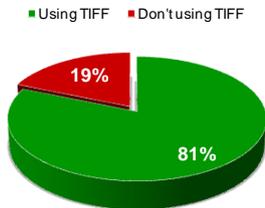
Only if we refer to the storage cost, large institutions are not so firmly as the medium and the small ones.

How important is the preservation processing cost for your institution (being 1 not at all important and 4 very important)?

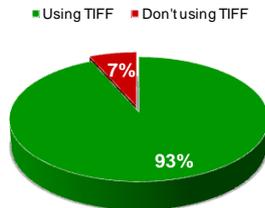


Most of the memory institutions consulted are public organization, with limited budget, for that reason the control of the costs is very important, in this case, using an open source platform without costs of purchase and without maintaining cost will be a good solution to apply better preservation methodologies without increase the costs of the general processing.

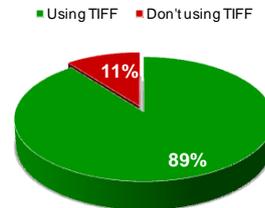
Which file formats are you using for archival purposes (Large Memory Institutions)?



Which file formats are you using for archival purposes (Medium Memory Institutions)?



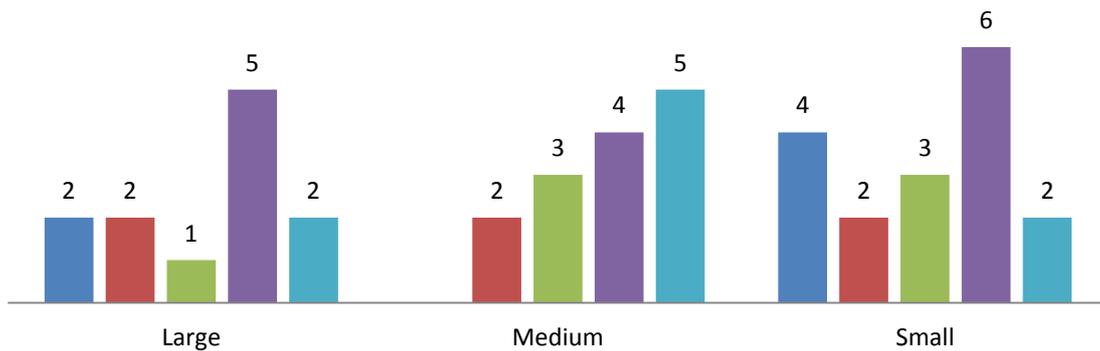
Which file formats are you using for archival purposes (Small Memory Institutions)?



The results are very similar for all the three sizes, being the TIFF and the JPEG the most common used format for archival purposes. Typically, they use TIFF as a production and preservation format and JPEG as a publication format, to share their archive through a web platform i.e.

How important is the technical documentation and wide range of tools for this format for your institution?

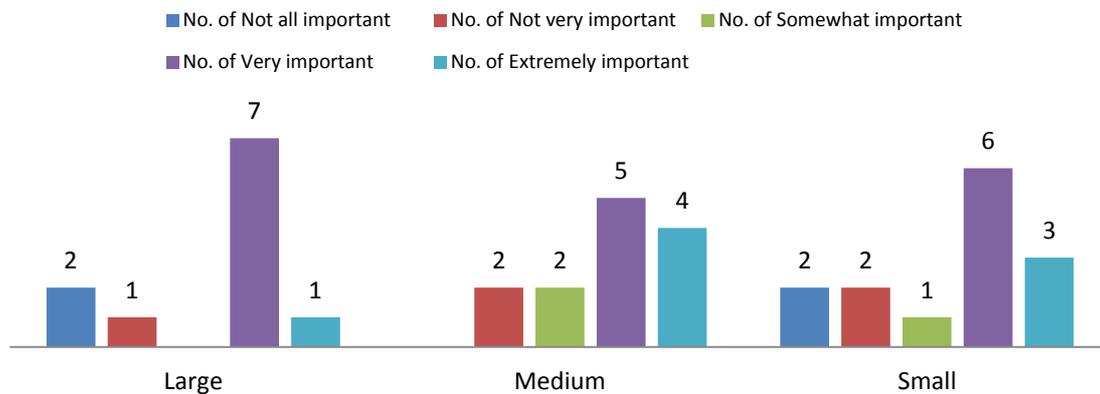
■ No. of Not all important ■ No. of Not very important ■ No. of Somewhat important
 ■ No. of Very important ■ No. of Extremely important



Like any other specialist, memory institutions have the need to access to good documentations for every tool they use, this need links perfectly with the requirement established by PREFORMA to offer with the tool developed a high quality set of document to make easier the use, develop and integration of this tool.

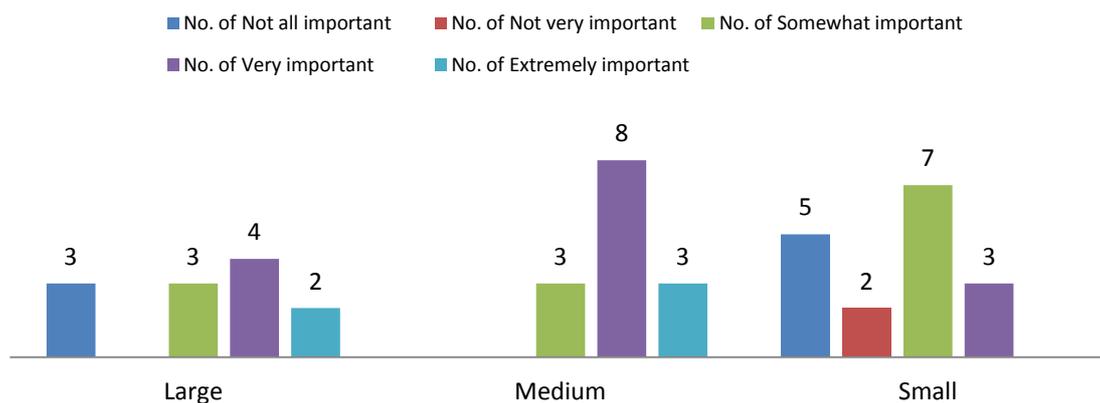


How important is the wide adoption for your institution?



With these numbers it is easy to confirm the importance of building a community, as more users, developers and stakeholders has the community the easier new users, developers and stakeholders will have.

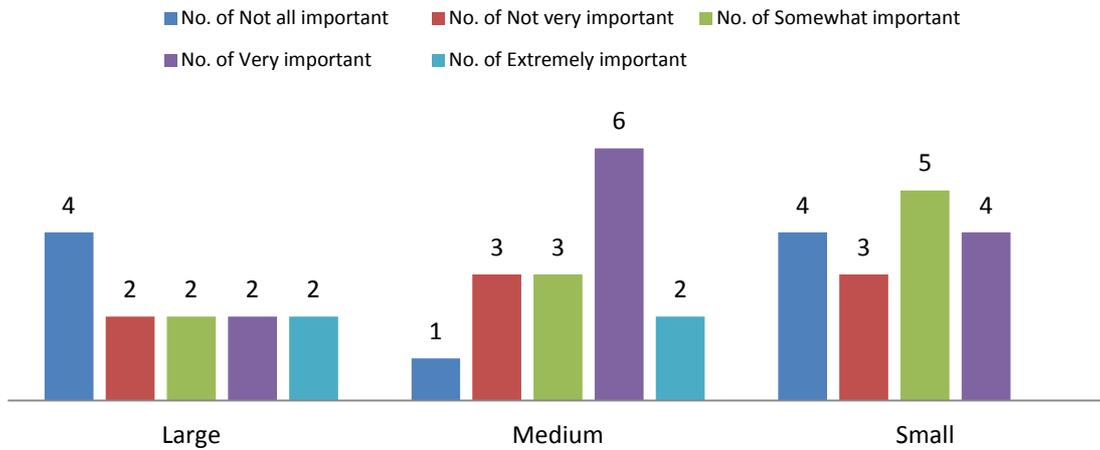
How important is the integrated metadata for your institution?



As higher is the size of the memory institution more important is the integrated metadata, and therefore the metadata management and validation, probably small institutions don't have enough resources and time to handle those metadata, but with the proper tool they can incorporate more and better information to their archives, handling those metadata more easily and without extra job.

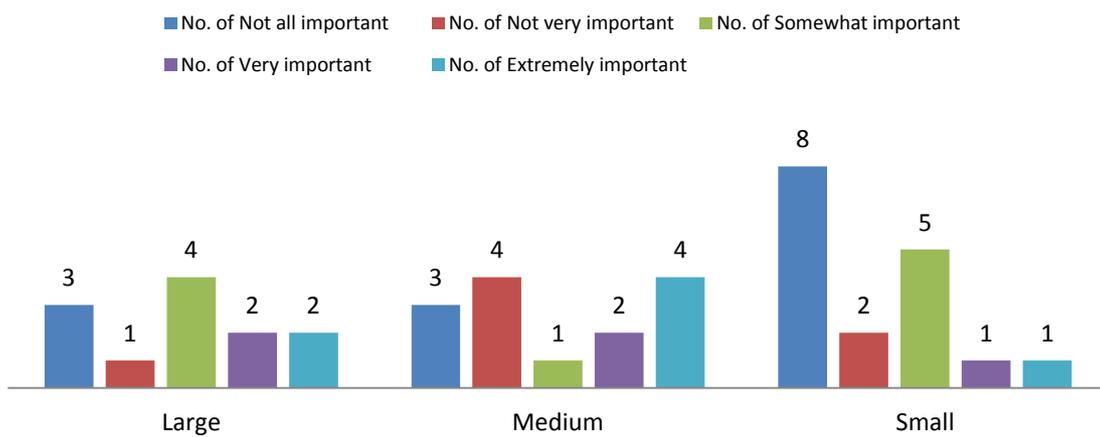


How important is the compression for your institution?



From this information we can extract that the memory institution are not aware if the importance of the compression in an image and the risk related to its compression.

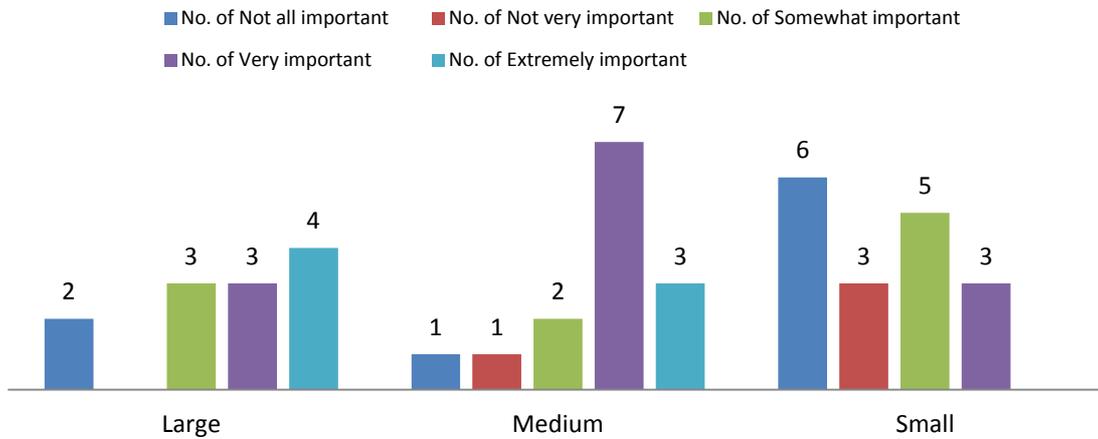
How important is the impact of patents for your institution?



Similar to other questions, this topic is more important whenever the memory institution becomes bigger. This is an important point to control costs and preserve intellectual property.

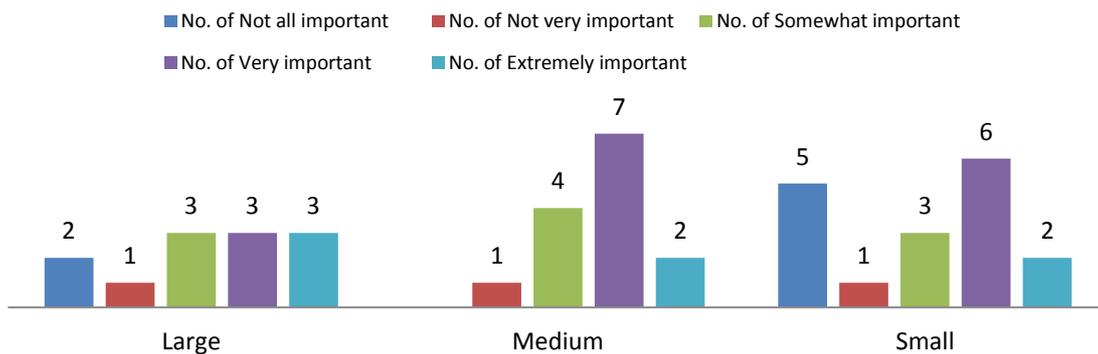


How important is the tonality quality for your institution?



Similar to the compression, memory institutions are not aware to the importance of the tonality. If in the case of compression the risk was to lose information, in this case the risk is to lose quality of the image and therefore endanger the preservation of the image.

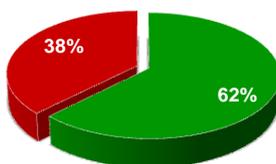
How important is the resolution quality for your institution?



This question was proposed to highlight the evidence of how the memory institutions give more importance to not relevant properties like resolution instead of tonality or compression.

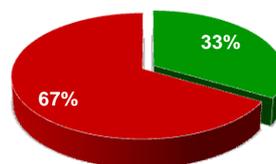
Do you apply any metadata validation process during the archival (Large Memory Institution)?

Yes No



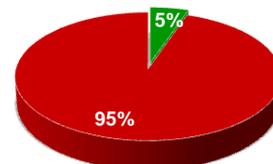
Do you apply any metadata validation process during the archival (Medium Memory Institution)?

Yes No



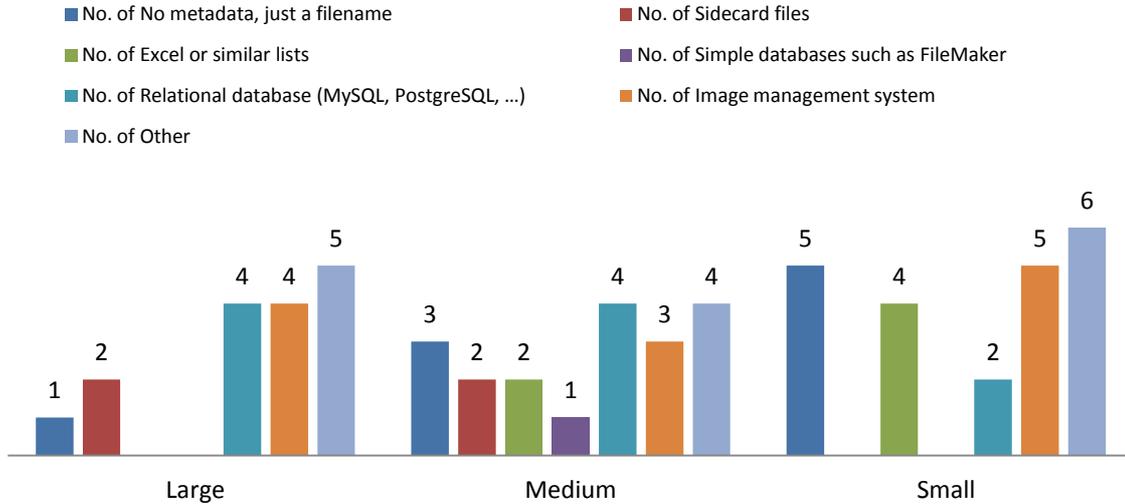
Do you apply any metadata validation process during the archival (Small Memory Institution)?

Yes No



As we can see, as larger are the memory institution more important is to validate the file liable to be archived, while in small institution no one is applying any validation, for the large institution more than the 60% are doing this validation.

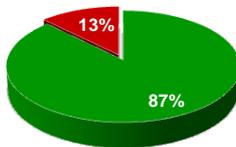
How do you store the metadata of your images?



Here appears two different conclusions, in one hand it seems necessary to have a tool able to analyze metadata from other sources than the same file, but on the other hand we can see that some memory institution don't realize the risk of having metadata information separately from the image file.

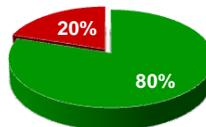
Are you aware of the importance of well formatted file structure following the standards and how it influences in his preservability (Large Memory Institution)?

■ Yes ■ No



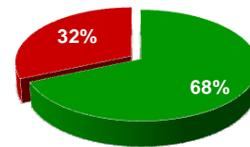
Are you aware of the importance of well formatted file structure following the standards and how it influences in his preservability (Medium Memory Institution)?

■ Yes ■ No



Are you aware of the importance of well formatted file structure following the standards and how it influences in his preservability (Small Memory Institution)?

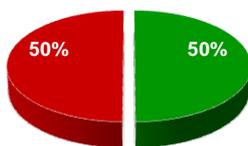
■ Yes ■ No



The three sizes have the same opinion about the importance of the standards and their application. These are a good news, encouraging to maintain and also develop new standards to provide better solutions to these institutions.

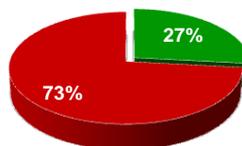
Do you use any software, like JHOVE, in order to check a well formatted file or its standard compliance (Large Memory Institution)?

■ Yes ■ No



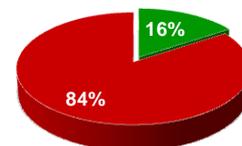
Do you use any software, like JHOVE, in order to check a well formatted file or its standard compliance (Medium Memory Institution)?

■ Yes ■ No



Do you use any software, like JHOVE, in order to check a well formatted file or its standard compliance (Small Memory Institution)?

■ Yes ■ No



Similar to the question about the validation of the metadata, here we can see that while the larger institutions are more used to use software to check their file, the small ones, in majority, don't use those software.

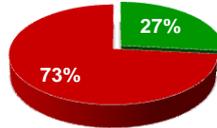
Would you like to be informed and/or participate in a community devoted to digital archival of images in order to discuss about image standard features suitable (Large Memory Institution)?

■ Yes ■ No



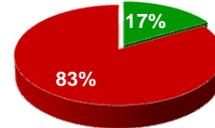
Would you like to be informed and/or participate in a community devoted to digital archival of images in order to discuss about image standard features suitable (Memory Institution)?

■ Yes ■ No



Would you like to be informed and/or participate in a community devoted to digital archival of images in order to discuss about image standard features suitable (Small Memory Institution)?

■ Yes ■ No



Finally we can see a high interest by all the memory institutions, no matter their size, of being informed and/or participate within the community we are starting to create, that encourage us in order to develop and promote new standards more suitable for preservation purposes.