END OF PHASE 1 REPORT

Project Acronym:	**PREFORMA**
Grant Agreement number:	**619568**
Project Title:	**PREservation FORMAts for culture information/e-archives**

DPF manager

Revision: final

Authors:

Miquel Montaner (Easy Innova)
Robert Sallo (Easy Innova)
Josep Lluís de la Rosa (University of Girona)
Lukas Rosenthaler (University of Basel)
Peter Fornaro (University of Basel)

Dissemination Level		
P	**Public**	**X**

1 INTRODUCTION

The purpose of the end of phase report is to ensure that contractors have performed the procured R&D services as specified in the framework agreement. Please describe the work undertaken during phase1, including what work was completed and why this was important. Please complete this form as fully as possible.

This report must be submitted within 14 days of the completion/ termination of the phase. You are advised that satisfactory completion of this report forms part of the contract.

Reports should be submitted by email to the following email addresses:

- Peter Pharow: phw@idmt.fraunhofer.de
- Claudio Prandoni: prandoni@promoter.it

The objectives of reporting:

- To create an understanding of the work undertaken and its success in meeting the projects agreed objectives.
- Also, to provide the company with a comprehensive report to share with stakeholders and those that may help further commercialisation.

The report should be completed by the contractor, with input from any sub-contractors or project partners as appropriate. Please answer, wherever possible, on behalf of the business units, divisions, companies or other legal entities involved in the work. If this is not possible, please specify the organisation to which your answers refer.

Please answer the questions in the spaces provided. Try to answer fully, but keep your answers succinct and no longer than necessary to clearly explain them. When describing technical solutions, please regard your audience as being someone familiar with the technology, but not an expert. The report may be done in narrative alone. However, diagrams or pictures may be added where these aid clarity within the restriction on the page limit of a total of 20 sides of A4.

Because the true impact of an R&D project often takes several years to emerge, we may approach you for up to six years after project completion to follow up on the questions in this report. Your co-operation with any such follow up work is greatly valued.

2 END OF PHASE 1 REPORT

1. Details
Type of Organisation: SME
Registered Name of Organisation: Easy Innova SL
Registered Address: Emili Grahit, 91 (Parc Científic i Tecnològic de la UdG)
Town/ City: Girona
Postcode: 17003
County: Girona
Country: Spain
Report Author: Miquel Montaner
Telephone Number: +34 972 41 88 54
E-mail Address: miquel@easyinnova.com
Project Name: DPF Manager
Report Type: End of Phase 1 Report
Total Contract Price [euro]: 57.993€
Start Date: November 4th 2014
End Date: March 13th 2015
Sub-contractors: University of Basel, University of Girona and ID Law Partners

2. At the outset of this piece of work, what were your aims and objectives?

At the outset of this piece of work our objective was to create a tool, DPF Manager, which would be specially designed to address present and future challenges in the digital preservation and standards compliance areas. Our objective was for it to be an open platform with a high modularity that would allow memory institutions and developers to easily add, modify and adapt specialized modules in order to personalize the analysis of still-image files to their processes.

We wanted to develop a full conformance checker to validate TIFF files against TIFF/EP (ISO 12234-2:2001) and TIFF/IT (ISO 12639:2004) specifications.

We also wanted to define a subset of variants of the TIFF 6.0 suitable for long-term archival purposes, since although TIFF is an appropriate, open and well documented standard for archival, this does not guarantee long life preservation.

We wanted it to have the following features to guarantee the suitability of DPF Manager to address future challenges:

• Modularity: the whole platform will be designed to be highly modular. Specific modules with specific tasks will be created, which is a really important feature that allows memory institutions to highly personalise the platform.

• Integration capabilities: thanks to its modularity and the Low-Level API, it will be really easy to integrate other open source projects or proprietary software to offer new formats validation or new functionalities.

• Easy to install: no expert knowledge will be needed to install the DPF Manager.

• Easy to develop with: our experience in open platforms and open source projects ensure that with simple technical documentation and/or short trainings, developers will be totally capable to develop over the platform.

• New formats: whenever new formats appear in the future (e.g. TIFF 7.0), creating a new conformance checker for this format will be very easy. Developers will just need to create a new module that programmatically inherits from the current TIFF 6.0 module and implement the new specific validations.

• Scalability: the platform will be designed having always in mind system performance on high workload conditions.

• Personalisation of automatic tasks: the platform will offer memory institutions the capability to schedule certain tasks like periodical batch analysis of files obsolescence with personalised parameters in order to facilitate archivists and content curators work.

We wanted to create a community around the project that would be continuously creating new modules to validate new formats, to add new functionalities and to guarantee the sustainability of the platform for a long time.

We also wanted to create a sustainable business around the DPF Manager, so we created an early version of a business model and business plan based on offering the following services on top of the free open source DPF Manager: cloud/SaaS platform, on-premise installations, 3rd party marketplace, consultancy services and training courses.

3. Please provide a summary off the outputs of this piece of work and relate these to the original objectives. How do the outputs address the challenge of this PCP?

Easy to extend implementation and policy checker, and new TIFF Standard proposal for long term digital preservation, TIFF/A
We have designed an easy to extend implementation checker, and a reference implementation for TIFF baseline, TIFF/IT, TIFF/EP and the new standard we have proposed called TIFF/A. This fits well with our original objective to create a modular implementation that could be extended to support additional existing and future formats, and the reference implementation requirements in the challenge brief (both in terms of standards and internal criteria).

Easy to use GUI, comprehensive reports and three APIs for the conformance checker interoperability
We have provided our conformance checker with an easy to use GUI, reports for expert and non-expert users alike, and three different APIs that allow for a wide range of deployment and use scenarios. This is related to our objective of creating a solution that was easy to use, integrate, and that could be scaled to fit user needs. It also meets the challenge of creating a conformance checker that could implement the 3 OAIS functions, 4 use cases, and be used in the 5 scenarios described in the challenge brief.

Open source best practices and dissemination plan
We will bring in an expert in the creation and growth of online communities around open source software products. We have defined the DPF Manager open source development approach, including

- licensing with the advice of a PREFORMA recommended IPR lawyer,
- open software development and releases,
- pre-compiled executables and compiling tools for end users and developers for multiple operating systems
- action plan to create and grow the community around the project.

This fits with our goal to create a community that would extend and reuse the DPF Manager to ensure the long term survival of the open source project. That fits with the requirements in the challenge brief related to open source development, stakeholders and adoption.

Joined standards organization to promote the TIFF/A specification to a new official standard, and to help improve the TIFF/IT and /EP specifications.
We have joined AENOR (Spanish standards organization) AEN/CTN50 committee as full members. That gives us the right to vote and propose new standards and modifications to existing ones. That fits well with the challenge requirement to work towards maintaining and improving the specifications related to the TIFF format.

Created a business model around DPF Manager, a business plan and engaged with Hewlett-Packard to form a partnership
DPF Manager will always be free. We have defined six services to commercially exploit DPF Manager: Cloud-based/SaaS preservation platform, on premise deployments, technical support and maintenance contracts, market place, certification for service providers, consultancy services and training courses. This closely matches our original objectives, and also matches the requirement to create business opportunities and sell professional services to generate enough revenue to make it a sustainable business.

4. Describe any changes to the original plan in the tender. What was the reason for these changes? Please include any circumstances that aided or impeded the progress of the project and the actions taken to overcome them.

Aside from the changes described below, we have followed the original plan as intended, as it was designed to help us deliver a solution that matched the requirements of the PREFORMA project.

TIFF/A specification and new standard proposal

We have developed a specification for a new standard named TIFF/A, that will make the TIFF format more suitable for long term preservation. This was not in the original plan, because the need for such specification arose during the 1st design phase, when we took a deep look at the TIFF format and realised some of its features where not suitable for long term preservation.

Implementation checker

Added TIFF/IT-P1 (ISO 12639:1998) and TIFF/IT-P2 (ISO 12639:2004), to give memory institutions a more granular control of the standards and sub-standards they can validate against.

Policy checker

In the original plan we said the policy checker would include a "configurable natural language metadata analyser to identify inappropriate content, e.g. sexually explicit or offensive content". We decided to not implement this feature, as some artworks do contain sexually explicit and/or offensive content (e.g the 1987 "Piss Christ", by American photographer Andres Serrano), and that may be reflected in the metadata. We have replaced this feature with a grammar builder described in point 5.

Metadata fixer

We have introduced a new feature, the ability to remove sets of metadata from a file, as requested by some memory institutions of PREFORMA during the bi-weekly groups calls.

Reporting

We have added a global report (generated when more than one file is checked at the same time), and also the possibility to generate PDF reports and to customise them with the memory institutions logos and other corporate image elements.

APIs

In the original document we defined one API for the DPF manager, a web service. To increase the possibilities of integration with other software, and to provide a wider set of use scenarios as described in the document "EASY INNOVA - 1. Functional Specifications.pdf", we have added two new APIs, a command line API and an internal API.

5. Please provide a short factual summary of the most significant outcomes of your work.

Fully modular conformance checker
We've designed a modular loosely coupled conformance checker that will be easy to maintain and extend by other developers. We have created a plug-in like system that will make it very easy to add new features when required.

Grammar builder for policy checker
To give as much flexibility as possible to memory institutions, we've developed a grammar with logical and arithmetical expressions that will allow memory institutions to fully define their own internal acceptance criteria.

Really easy to use Graphical User Interface
It is not unusual that highly-specialised tools like a TIFF conformance checker come with user interfaces that are complex and difficult to use. DPF Manager will have an easy to understand/easy to use interface that will guide users through the whole conformance checking process.

It will follow the "progressive disclosure" principle, where expert users will be able to configure the tool up to the last detail, while non-expert users will be able to use it by just clicking "Next, Next, Finish".

Actionable and easy to understand reports
The end-user reports will come in HTML and PDF format, with a customisable layout to adapt it to the corporate image of the memory institution.

Each report will have a summary with information for non-expert users, and a significant amount of detail for expert ones, including a list of all the metadata in the file, the internal file structure, errors found during the implementation & policy checking process, and suggestions on how to solve those issues.

Multiple API interfaces for different integration scenarios
We have provided DPF Manager with 3 different APIs (command line, web service, internal Java API) that allows for a wide variety of integrations with legacy software, external producers and digital asset management/image & metadata editors.

TIFF/A
We have reviewed the current TIFF specification and have selected a sub-set of its features that our experts together with an important number of memory institutions believe are the best suited for long term preservation. We have created a draft specification and we have started the process with ISO to create a new standard based on the proposed specification.

6. Describe the innovative aspects of the work, including any new findings or techniques.

Protocol for configuration sharing between conformance checkers
We have developed a protocol based in the exchange of XML/JSON data that allows a conformance checker from supplier 1 to query the capabilities of another conformance checker from supplier 2. Capabilities include: list of file formats and standards to validate, accepted criteria for defining the policy checker rules, report formats and options, and capabilities of the metadata fixer

With this protocol, integrating multiple conformance checkers becomes as easy as giving conformance checker 1 the location of conformance checker 2. Then the two conformance checkers will configure themselves in a way that is transparent to the end user.

Global report and file conformity score
When a user requests a conformance check on multiple files, DPF Manager is able to collate the individual reports for each file into a global report, that allows the end user to see at once the results of the conformance checking process.

We have also introduced a scoring system, where each file is given a conformance score. The higher the score, the closest the file is to being conformant. This score can give a visual indication to the end user of the evolution of the conformity of a whole archive.

Easy to use GUI
Unlike some of the existing solutions for conformance checking (such as JHOVE2), DPF Manager provides an extremely easy to use Graphical User Interface, that guides users through the whole conformance checking process, and makes it easy to understand the reports generated by the conformance checker.

Multi-API
As shown in the document “EASY INNOVA - 1. Functional Specifications.pdf”, we have designed 3 different APIs for the DPF Manager (command line, web service, internal Java API) that allows its use in a multitude of different scenarios.

Valid for memory institutions of any size
Related to the point above, we’ve worked hard to create a solution that can be used by very small memory institutions to very large ones with millions of digital assets in their archives.

TIFF/A
Together with the DPF Manager, we have also researched and developed together with some memory institutions a new standard proposal, that would make TIFF files more suitable for long term preservation.

7. Describe where the R&D and other operational activities have been performed.

All the activities related to the PCP contract have been performed in 2 different countries:

Spain (EU member state)

- Easy Innova (supplier) settled in Girona (Emili Grahit, 91 17003 Girona)
- University of Girona (subcontracted) settled in Girona (Universitat de Girona 17006 Girona)
- ID Law Partners (subcontracted) located in Barcelona (Passeig de Gràcia, 81-1 08008 Barcelona)

Switzerland (FP7 associated country)

- University of Basel (subcontracted) located in Basel (Bernoullistrasse, 32 CH-4056 Basel)

8. Please provide complete and clear information about the allocation of monies paid by the Authority with consideration to the R&D service contract minimum requirement (that more than 50% of the contract value is attributable directly and exclusively to legitimate R&D services).

Around 95% of monies paid by the Authority have been dedicated to R&D services. Such services are:

- Own R&D tasks: Easy Innova dedicated to the project 6 persons, in particular, Josep Lluís de la Rosa, Miquel Montaner, Víctor Torres, Antonio Manuel López, Robert Sallo and Xavier Tarrés. All the tasks performed by these persons are R&D activities.
- R&D subcontracted tasks: we subcontracted the University of Basel and the University of Girona to perform some specific tasks.
- Legal advice: we subcontracted ID Law Partners, experts on IPR and open source licenses to support us about the licenses requirements of the PREFORMA project.

The expenses not considered as R&D services are travel and accommodation costs: in particular 2 travels have been carried out during Phase 1:

- Travel to Basel (from Spain to Switzerland): 2 days trip to work with the University of Basel.
- Travel to Brussels (from Spain to Belgium): 2 days trip to present or design proposal to the PREFORMA consortium in the End of Phase 1 Workshop.

Following you can find the detailed monies allocation:

Concept	**Costs**	**%**
Easy Innova Personnel	33.790 €	51,2%
Subcontracting of the Universities	28.000 €	42,4%
Subcontracting of ID Law Partners	1.000 €	1,5%
Travel and Accommodation	3.181 €	4,8%
Total Costs	**65.190 €**	

These costs have been financed as follows: 57.993 € from the PREFORMA Project and 7.978€ as Easy Innova investment.

9. Describe any potential long-term collaborations/ partnerships entered into. Please list the organisation/s and the role they played in the project.

Digital Humanities Lab at the University of Basel, Switzerland:

The Digital Humanities Lab are the TIFF format experts in our team. They have been doing research in long term archival of still-images and video since 1995. For this project they have drafted the TIFF/A specification, and have provided us with access to numerous memory institutions in Switzerland.

We plan to carry on working with them in other digital preservation projects, as we've found that their knowledge and our capability to build technical solutions are an excellent match.

Memory Institutions

During the design phase of the PREFORMA project we have got in contact with several Memory Institutions from over 14 different countries. They have helped us better understand their needs and the way they work.

We intend to take this collaboration further by working with them on improving the TIFF/A specification, and also to find areas of R&D where we could collaborate and deliver solutions to improve their technical capabilities.

AENOR

Part of our proposal is to create the new standard TIFF/A. To do so we have joined AENOR (Spanish standards organisation, ISO member) and become full members of the AEN/CTN50 committee.

This membership will be useful not only for the PREFORMA project, but also for other projects as well, as it will put us in contact with other member organisations. These contacts may prove to be useful in the near future, as we intend to establish new partnerships with some of these organisations to expand our reach.

It will also give us the right to vote and propose modifications to emerging standards as part of the AENOR/ISO standardisation process.

id law partners

Malcolm Bain from id law partners has provided us with legal advice on FOSS licensing during the design phase, and will continue to do so if we are one of the selected suppliers.

As open source and open data projects become more commonplace, we plan to continue collaborating with him in future projects, as his expertise has proved to be invaluable in a field as challenging as licensing and intellectual property rights management.

Hewlett-Packard CDS

After presenting DPF Manager to HPCDS, they sent us a letter of intent to state their interest on a potential commercial partnership to distribute & integrate DPF Manager into existing document and digital asset management applications.

10. Please describe how your organisation has gained from this project. What new business opportunities have been created? Do you expect your organisation to grow as a result of this project?

First, and thanks to the thorough work PREFORMA has done in regards to the analysis of the needs of memory institutions, and our own analysis via online questionnaires, we have gained valuable insight about this market, which will be a welcomed addition to what we already know thanks to other digital preservation projects we have worked on.

Second, during the development of the design phase we have had the chance to collaborate with other brilliant institutions and companies (namely the Digital Humanities Lab at the University of Basel and the law firm id law partners). We plan to continue working with them in other projects as their knowledge and experience have made a difference in the outcomes of this project.

Third, thanks to the contacts we have made with some memory institutions, we have already started discussing potential business collaborations with a few of them. These opportunities would not have arisen if we had not participated in the PREFORMA project.

Fourth, if our bid is selected and we are asked to continue to the prototyping phase, we have already identified multiple new business opportunities. These opportunities are described in more detail in the document “EASY INNOVA - 7. Business Commercialization Plan.pdf” delivered on the 3rd of March.

In order to carry out this business plan, we foresee to grow our technical team by 2 to 3 people during the prototyping phase (additional people will be added if necessary), and 3 to 4 additional people during the first two years of the commercialization phase after the project.

Moreover, our potential partnerships with big players in the industry (e.g. Hewlett-Packard CDS) may offload some of the work to them, as they already have a well-established network of customers and technical support agents.

11. Describe the potential for exploiting the work. Please identify any new intellectual property which has been filed or for which filing is anticipated.

This is a summary of the document "EASY INNOVA - 7. Business Commercialization Plan.pdf" delivered by Easy Innova on the 3rd of March. The document outlines the market opportunity and how we plan to create a set of services around the DPF Manager to generate enough revenue to make it profitable and therefore self-sustaining in the future.

We do not believe in offering DPF Manager under a dual license model (free version with basic features, paid version with extra functionality), as this inevitably creates an unhealthy environment with 1st class and 2nd class "citizens", where most of the efforts end up focused on the clients that generate revenue rather than the overall user base.

We think that DPF Manager should always be free for everyone, and that revenue should come from services built around the product. We have identified and quantified 7 different ways in which we can monetize the work developed during the PREFORMA project:

1. **Cloud-based/SAAS:** hosted and maintained by Easy Innova, for those clients that want to simplify DPF Manager's adoption process and maintenance.

2. **On premise deployments:** for those clients that want to keep the full control of their IT infrastructure, but need help setting the platform up.

3. **Technical support and maintenance contracts:** for those clients who want to outsource the technical maintenance and support.

4. **Market place:** for developers interested in creating and commercializing modules for the DPF Manager, Easy Innova will offer a centralised marketplace. A fee will be applied to each sale.

5. **Consultancy services:** for those clients that need technical and functional advice on what is the most appropriate way to fulfill their needs.

6. **Training:** for end users, IT departments and developers

7. **Certification for service providers:** for those companies/institutions that want to offer the service of installing the software on premise or offer a SaaS solution.

The strategic objectives set in the business plan are:

1. A portfolio of over 300 memory institutions in 3 years
2. Be present in 35 different countries in 2020
3. A community of 2.000 members by the end of 2020
4. A network of 50 service providers in 2020
5. Over 1.5 million Euros in revenue by the end of 2020

At this point there are no plans to file any intellectual property.

12. Describe the suitability of the project results for: (a) developing a prototype, and (b) development of test series – in order to facilitate assessments of progress into next phase.

The documents “EASY INNOVA - 1. Functional Specifications.pdf” and “EASY INNOVA - 3. Technical specifications.pdf” delivered on the 3rd of March, that include the architecture of the system, a detailed set of mock-ups of the Graphical User Interface, and the reference for the APIs amongst other things, are complete enough to be considered a valid blueprint to build the DPF Manager tool. In this respect we consider the results of the project suitable for developing a prototype of the tool.

High level software and test data development plan

1. WP1 & WP2: Development of the implementation checker + creation of test data
2. WP3: Development of the remaining conformance checker components (policy checker, metadata fixer and reporter)
3. WP4: Development of the shell: APIs, data persistence (DIRECT framework)

In parallel to the first 3 work packages we will start working with the other suppliers to create the specifications of the APIs that all the conformance checkers have to share to be interoperable.

First steps

Once into the prototyping phase, we will start by working on the implementation of the reference standards, the implementation checker. We will start implementing the TIFF format parser and structure and metadata extractor. Then we will implement the validators of the different standard file format specifications (TIFF/EP and TIFF/IT).

Parallel to that, our partners at the University of Basel will start creating the series of test images, as described in “Section 15. Provision of data”. These images will contain a variety of test scenarios (valid images and invalid images with different errors in different places).

The objective of this piece of work is to be able to produce, as quickly as possible, a “standalone” implementation checker that can be tested with the data provided by the University of Basel and the memory institutions in the PREFORMA project.

This is the core of the conformance checker and is the first part we want to complete, working iteratively as the results of testing the data reveal shortcomings or malfunctions in the code.

This early “standalone” implementation checker can also be distributed to the memory institutions that may be interested in testing it.

Next steps

There are two main areas where we would like to dig deeper to get an even better understanding of the features that the solution has to offer: the types of policies and metadata fixes that users may want to define. We have defined the current functionality based on a number of examples from PREFORMA and other memory institutions we have contacted, but we would like to collect even more real-life cases to make sure DPF Manager suits a variety of different scenarios.

13. Open Source approach

This is a summary of the document "EASY INNOVA - 6. Open Source Best Practices and Dissemination Plan.pdf" delivered by Easy Innova on the 3rd of March. The document outlines how we plan to create and grow an open source community around the DPF Manager, as well as describing the software release cycle.

We plan to create an open source community with two main points of contact: the project's Github repository page and the project's website.

The Github repository is aimed at other developers and technical end-users. By visiting the Github repository page, users will be able to:

- Get the source code and test files (both in a single file or by forking/cloning the repository) for the nightly-builds, beta, release and Long Term Support (LTS) branches.
- Get all the tools needed to compile the source code for the most popular operating systems (MS Windows 7, Mac OSX, common Linux distributions including Ubuntu, Fedora, Debian, and Suse).
- Download pre-compiled binaries for the most popular operating systems (MS Windows 7, Mac OSX, common Linux distributions including Ubuntu, Fedora, Debian, and Suse).
- Report issues with the application and the test file using the issue manager/bug tracker provided by Github
- Access the project documentation for developers (API specifications, etc.) and also for end users (tutorials, cookbooks, etc.)

The project website is aimed at other developers, technical and non-technical end-users, other members of the open source community, and other stakeholders interested in the project.

The content made available through the project website will be:

- Links to all the content already provided in the Github repository (source code, pre-compiled binaries, documentation and bug trackers)
- IRC channel and forums for technical and non-technical visitors
- Access to the mailing lists
- Additional documentation in multimedia formats, such as videos presenting the project and video tutorials
- Additional information on how to join the community and collaborate with the project
- Additional information aimed at media, bloggers and other communicators (PDF leaflets, etc.)

In order to create and grow the community, we have created a dissemination plan, by segmenting the project stakeholders in 6 groups (memory institutions, developers, standardisation bodies, researchers, other projects, others), and defined 50+ actions with start and end date, with the aim to raise awareness and participation in the project.

Finally, in order to reinforce the open source project, we will bring into the team an international expert in the creation and growth of online communities around open source software products, as explained in section 1.2 of the open source best practices document.

14. Standardisation efforts

Easy Innova and the Digital Humanities Lab at the University of Basel have created a specification draft for a new standard TIFF/A that will make the TIFF format more suitable for long term preservation.

In order to get this specification recognised as a new standard by the International Organization for Standardization (ISO), we have taken the following actions:

Joined AENOR

AENOR is the Spanish national standards organisation and ISO representative in Spain. We have already become a full member of AENOR/CTN50 SC1 subcommittee (Spanish mirror of ISO/TC 171). The membership gives us rights to propose and vote on new ISO standards.

As soon as we finalise the draft with the feedback received from memory institutions and other stakeholders, we plan to submit it for formal review to start the standardisation process.

The whole process can take up to three years, so it may be some time before we can see the results of our standardisation efforts.

Informed ANSI

ANSI holds the secretariat of ISO/TC 171. We got in contact with Betsy Fanning, the nominated secretary. After explaining our proposal, she invited us to a Technical Advisory Group meeting in San Jose, CA at the end of April to discuss the initial draft of the TIFF/A specification.

Although this step would not be required if we used our status as members of AENOR to push our proposal, ISO requires 5 member countries to support any proposed standard before it is officially recognised, so the support of ANSI would make our proposal more likely to be accepted.

Started the process to join SNV

SNV is the Swiss national standards organisation and ISO representative in Switzerland. Our partners at the University of Basel have already started the process to become full members. As with AENOR, becoming a full member of SNV will give the University of Basel voting rights on new proposals.

By becoming members of both AENOR and SNV, we have given ourselves the possibility to choose which organisation is better suited to push a new standard proposal. We will make a decision as soon as we know which channel would be more effective in helping us achieve our goal.

15. Provision of data.

The best test files are those that represent reality as closely as possible. Besides synthetically generated test files, such a real set of image data files is the perfect complement for any test collection.

Real test files:

In most cases it is only possible to generate test files that are expectable (fake real data), being that files with or without errors. Real errors are often especially difficult to capture because they represent the unexpected as well. To get a real set of test files we will collect a large variety of natural born image files. The image files will be requested from the numerous memory institutions we have already been working with. The files we will ask for are:

1. of different size and format
2. of different color spaces
3. made for different purposes
4. have different sets of metadata
5. come from different institutions (archives, museums, …)
6. they fulfill the requirements of open access Creative Commons (CC), ShareAlike 4.0 International (CC BY-SA 4.0)

The owners of the files will be asked to confirm us all and especially point 6 of the above list.

The Digital Humanities Lab of the University of Basel has done and guided many different digitization projects. Therefore we have a very close relation to those institutions offering us the data, which is the key to be able to get access to such files. The memory institutions asked will be invited to be beta testers of the DPF Manager and therefore they have an advantage to be part of the test group.

Artificial test files:

We will generate a set of artificial test files that are different in regards of:

1. different size and format (TIFF, TIFF/IT, TIFF/EP, TIFF/A) with and without errors
2. different color spaces
3. have different sets of metadata
4. they have a set of random and intentional errors like:
    a. bit errors
    b. structural errors
    c. errors in components like color profiles, EXIF or IPTC containers
    d. in the compression scheme
    e. in the approach for CFA interpolation (in case of TIFF/EP)

The artificial test files will enrich the collection of natural test files, as both must compose the complete set. The test files will be generated by unprotected (open) raw material (bitmaps) and by self-developed software based on open libraries, free of any license, written in c++.

16. Please insert additional information that may be pertinent. This may be in the form of text, pictures, diagrams, data, graphs that support the work.

17. Describe what ethical aspects you have identified and how this may influence your solution.

Ethical: in accordance with principles of conduct that are considered correct, esp. those of a given profession or group. Ethics or morality has to do with the principles, standards, rules, norms of conduct that make cooperation, justice, and freedom possible

According to these definitions of ethics, during the design phase we have identified issues concerning the preservation of the integrity of the original content of TIFF files, and issues concerning the chosen open source license:

Automatic fixes by the metadata fixer

With the metadata fixer we had the issue of deciding whether we would allow it to fix issues that also required a change in the image data itself, or we would restrict ourselves to metadata only.

From a technical point of view, automatically fixing the image data is relatively straightforward (e.g. un-compressing a TIFF with compressed data), as there are plenty of well known and tested libraries that deal with compression, etc.

But we felt that automatically modifying an image that was meant for long term preservation was not appropriate, as the modification may change the representation of the image without the user taking notice (specially when DPF Manager is used to perform periodical checks). There are better ways to modify image data, such as using image editors, where the user is aware and in control of the changes being made.

In the end we decided that the metadata fixer would only fix issues that did not require a change in the image data.

Licensing requirements

Choosing a license is not easy, as one has to balance two opposing forces: the rights you want to keep and the rights you want to give away. In the case PREFORMA, an open source project financed with public funds, we believed the licensing requirements did not achieve the right balance and were at odds with the requirement to create a successful open source community.

The requirement to limit third party libraries to those licensed under GPL3+ and MPL2+ may have fulfilled the needs of the main stakeholders, but that license is not commonly used in the open source community[(1)], and therefore it may have limited the amount of open source developers interested in being part of the community.

In the end, PREFORMA relaxed the licensing requirements a little bit. While the change was welcomed, we would still have preferred (from an open source point of view) that the original requirement was to use a more well-know license .

(1) List of most used licenses in Github: MIT 44.69%, GPLv2 12.96%, Apache 11.19%, GPLv3 8.88%, BSD 3-clause 4.53%, Unlicense 1.87%, BSD 2-clause 1.70%.

Source: https://github.com/blog/1964-open-source-license-usage-on-github-com

3 FINANCIAL REPORT

	Unit price	Quantity	Quoted price (€)	Total Price (€)
Labour Price			**31.593 €**	**33.790 €**
1. Josep Lluís de la Rosa	*7.756 €/PM*	*0,20 PM*	*1.939 €*	*1.551 €*
2. Miquel Montaner	*7.266 €/PM*	*1,25 PM*	*9.083 €*	*9.083 €*
3. Albert Trias	*4.873 €/PM*	*0 PM*	*7.310 €*	*0 €*
4. Víctor Torres	*5.221 €/PM*	*0,25 PM*	*1.305 €*	*1.305 €*
5. Anna Bosch	*5.221 €/PM*	*0 PM*	*6.526 €*	*0 €*
6. Antonio Manuel López	*3.620 €/PM*	*1,50 PM*	*5.430 €*	*5.430 €*
7. Robert Sallo	*5.886 €/PM*	*1,25 PM*	*0 €*	*7.357 €*
8. Xavi Tarrés	*4.532 €/PM*	*2,00 PM*	*0 €*	*9.064 €*
Sub Contract			**24.000 €**	**29.000 €**
University of Basel			*12.000 €*	*16.000 €*
University of Girona			*12.000 €*	*12.000 €*
ID Law Partners			*0 €*	*1.000 €*
Travel and accommodation			**2.400 €**	**3.181 €**
TOTAL PRICE (excluding VAT)			**57.993 €**	**65.971 €**

Comments about the difference with the initial budget:

- The End of Phase 1 Workshop was not planned and the efforts to prepare it and the costs of the travel and accommodation increased our budget.
- The preparation of the End of Phase Report was not planned and the efforts to prepare it increased our budget.
- We planned to subcontract a legal expert during the prototyping phase, but the requirements of the project on IPR and licensing forced us to advance part of the budget to the design phase.
- Finally we subcontracted more tasks to the University of Basel, but this not increased our budget, just redistributed it.