PROTOTYPING PHASE 2
FINAL 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)
Xavi Tarrés (Easy Innova)
Víctor Muñoz (Easy Innova)

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

1 INTRODUCTION

During the PREFORMA Prototyping phase, suppliers are expected to provide software prototypes that fulfil the requirements of the PREFORMA project, to demonstrate the results of their development work, and to provide explanations and documentation (manuals) on how the developed software can effectively be used in archiving scenarios at memory institutions regardless of their size and the file type they make use of.

Following the same approach used last year, during the Second Prototyping Phase the plan for releases is as follows:

- Frequent releases: monthly;
- Intermediate releases: end of July 2016 and end of October 2016.

The intermediate release shall contain two parts:

- A functionally stable release
- A report which
    - Describes
        - In more detail the respective release;
        - The time line along with the current position (on time, delayed, ahead)
        - How suppliers managed to provide the required functionality (so far);
        - What is still missing compared to the original specifications and what is the plan to implement it.
    - Provides basic information to be used by PREFORMA WP8 in their deliverables to be submitted to the EC, reporting the work done by both suppliers and PREFORMA consortium members during the prototyping phase.

2 PROTOTYPING PHASE 2 - FINAL REPORT

1. Details
Type of Organisation: SME
Registered Name of Organisation: Easy Innova, S.L.
Registered Address: Emili Grahit, 91 (Parc Científic i Tecnològic de la UdG)
Town/ City: Girona
Postcode: 17003
Country: Spain
Report Author: Miquel Montaner
Telephone Number: +34 972 41 88 54
E-mail Address: miquel@easyinnova.com
Project Name: DPF Manager - Digital Preservation Formats Manager
Report Type: Prototyping Phase 1 – Final Report
Total Contract Price [euro]: 699.475 €
Start Date: 14/04/2015
End Date: 31/10/2016
Sub-contractors: University of Girona (Spain), University of Basel (Switzerland), ID Law Partners (Spain), Bas Van Leeuwen (Netherlands) , Legalment (Spain), Pep Boix (Spain)

1. Description of the release and progress compared to the last intermediate release

The final prototype release of the DPF Manager includes all the features specified in the functional requirements and technical specifications described in the Easy Innova phase 1 documentation and tender proposal. Moreover, we included some not planned functionalities and improvements, most of them proposed during the redesign phase.

One of our challenges was to build an innovate application based on the modern developer techniques. Now we can claim that DPF manager is using the latest developer technologies. It uses the latest java versions (the second most used language in GitHub). The JavaFX, the new framework of java for graphical user interfaces (GUI) coming as a successor to Swing. A modular architecture using modern event driven communication approach, where the modules are independent, improving the maintainability and providing an easy way to expand the application with new functionalities. The use of java ensures portability, java provides the simplest and most familiar form of portability -- source code portability. A given Java program should produce identical results regardless of the underlying CPU, operating system, or Java compiler (including legacy systems). We provide the installers and executables for Windows, Mac OSX and Linux (Ubuntu, Fedora, Devian and Suse) and it is very easy to build the application for other systems using Maven. Maven is the latest tool solution to simplify the build process providing a uniform build system with all the code dependencies and also including quality project information.

The DPF manager all-in-one application includes all the interfaces to be deployed and integrated with different infrastructures in one executable. DPF manager can be initiated with a Graphical User Interface (GUI), a command-line interface or as a Client-Server application. All these different interfaces share and deploy the same implementation, allowing an easy maintainability. This means that you can create conformance checks using one interface and these immediately appear in the other instances of the application.

The application can communicate with other applications in order to provide interoperability with other conformance checkers. The Shell of the DPF Manager distribute the files to the appropriate conformance checkers. Moreover, it can manage multiple conformance checkers for the same file format providing high availability and scalability. There are preservation actions that require to be done in a reasonable time, like the ingestion of thousands of files. In this scenario multiple instances of the DPF manager working as a server can be deployed in different hosts. The client can distribute all the files to the different server instances or specify multiple folders from different hard drives avoiding the I/O bottle neck and improving the performance. The shell is also able to create periodic checks in order to monitor a specific folder of the system.

In addition to the application interfaces, an online web conformance checker is also provided. This website is a frond-end that interacts with a server instance of the DPF Manager. The on-line validator can be tested at this webpage.

The DPF manager not only can be used as an application but also as a framework, ready to be integrated with other applications or frameworks via API. In order to facilitate the integration, DPF manager has been included in the Maven package repository.

The DPF Manager application includes the build in TIFF conformance checker. The TIFF conformance checker includes; the Implementation checker, policy checker, metadata fixer and

a report module.

Implementation checker

The approach used in the implementation checker is a crucial aspect of the project. The TIFF writers are those who follow the implementation and generate TIFF according the specifications. But, implementing a specification is always subject to interpretations, used as convenience for performance reasons o even misunderstood.

In order to prevent future problems, our approach creates an implementation as much accurate to the ISO specifications as possible, giving the same errors, warning and including the same information that the specification provides. We do not want to adapt the implementation to the main interpretation done by the TIFF Writers, at the end, the only model that will prevail will be the ISO documentation. TIFF writers may become obsolete or disappear, the ISO documentation will rest.

This approach would conflict with few TIFF writers which not strictly follow the specifications, where our implementation checker would report errors in every TIFF file generated. Knowing this possible issue, we implemented a new functionality preventing that files generated by writers used for memory institutions o digitalization software fail again and again. In the final release we let the users "create" a custom specification based on our implementation. The users could "relax" some rules of the specifications to let them pass their acceptance criteria. Approving files that not fully fulfil the specifications, is not ideal but we need this kind of flexibility to adapt the implementation to the memory institutions' needs. This is more necessary when the implementation is a guide or a recommendation, like our TI/A initiative where some recommendations will come into conflict with some restrictions of the memory institutions; for example: not enough disk space to store uncompressed files, enough resources to migrate files or preserving files from a specific domain.

Our implementation checker is able to check the following specifications:

- Baseline 6.0: the baseline 6.0 document is divided in two parts. The Part 1: Baseline TIFF is the core of TIFF, the essentials that all mainstream TIFF developers should support in their products. And the Part 2 which contains extensions to Baseline TIFF. Files that use such features shall be designated "Extended TIFF 6.0" files, and they are not recommended for general data interchange, since Baseline TIFF 6.0 readers are not required to support any extension. Our final release contains both implementations, being able to validate the Baseline 6.0, as well as, the Extended TIFF 6.0.
- Tag Image File Format/Electronic Photography (TIFF/EP): ISO 12234-2, based on a subset of the Adobe TIFF standard, and a subset of the JEITA Exif standard, with some differences and extensions.
- Tagged Image File Format/Image Technology(TIFF/IT): ISO ISO 12639:2004, based on TIFF standard includes features for exchange of digital adverts and complete pages. The versatility of the original TIFF/IT specs soon proved to lead to compatibility issues between the software of different vendors. To avoid these compatibility issues, a more limited version of the standard was developed, which is called TIFF/IT P1 (also known as ISO 12639). The P1 stands for Profile 1. When most people talk about TIFF/IT, they mean the P1 standard, because TIFF/IT P1 has a number of limitations. A group of companies joined forces to develop an extended format, to be called TIFF/IT P2. P2

add more functions and extend the P1 profile. Our implementation checker covers the 3 implementations: TIFF/IT, TIFF/IT-P1 and TIFF-IT/P2

- In the final release, we also included the first TI/A Draft specification, a guideline for TIFF archival purposes. This is still a draft but we think it is important to include this specification in order to disseminate the initiative as well as to help improving the draft specification getting feedback from memory institutions using the application. Our purpose is updating the application every time a new version is distributed in the TI/A initiative website.

Moreover, from the feedback received from memory institutions that do not base their acceptance criteria in any ISO specification, we developed the TIFF file identification functionality. When a conformance check is created, even if the user has not selected any ISO in their acceptance criteria, the file is checked across the implementations in order to identify if a file is ISO compliant. This information is not relevant for the acceptance criteria but is useful in terms of digital preservation, knowing that a TIFF file is compliant with a concrete ISO specification increase the sustainability of the file.

All the documentation about the implementation is now included in the website. The documentation includes all the rules used to validate the implementation organized by sections with detailed information about each error: title, description, the reference to the ISO and the error message and code that the application produce. Since each implementation contains thousands of rules, we have included a search box where the user can search by error code or text on the title or description. The documentation directly come from the latest implementation in the GitHub, if any change is made in the implementation the documentation will be automatically updated.

Policy checker

Our policy checker is able to define an acceptance criteria based on technical aspects of the file. We could develop filters for every TIFF tag or metadata inside the file, but this approach would become in a mess, with a lot of selectors useless or never used. We have provided filters to create useful and used acceptance criteria. In order to decide which are the most useful filters, we asked and received feedback from memory institutions, official organizations and we also included governmental public acceptance guides for digital preservation. Moreover, now we also provide the option to generate warnings, the file will pass the acceptance criteria but a warning will be created in order to report a desired feature into the file. In this new release we have also provided new TIFF features that can be evaluated.

Metadata fixer

The TIFF file can incorporate metadata from different containers, using the default Image File Directory with private tags, XMP information, IPTC (legacy IPTC/IMM, IPTC4XMP), EXIF or ICC Profile. We are able to handle with the different formats and add, edit or remove metadata consistently in all the containers. Moreover, not all the TIFF writers are able to detect and interpret all the containers, editing information only in the containers they know. This process create metadata inconsistences. Following the Metadata Working Group guidelines, we are now able to report metadata inconsistences among containers and include this information in the final report.

We also include the ability to incorporate metadata fixes. Metadata fixes are modules that can be easily added into the DPF Manager in order to perform a concrete action with the metadata. Now, the metadata fixer includes 3 modules. One for removing personal private metadata from the file in order to create a Dissemination Information Package, without any information that could contain private data from the author (for example: their house location coming from the GPS information, device used, etc...). This module was already available in the previous releases. Now we have included two more modules, one to fix the metadata inconsistences found in the file and another that tries to solve as much errors as possible (without altering the image data structure) in order to make the file TIFF Baseline compliant.

Report module

In our opinion, the most important part of the DPF Manager is the reporting module. We have created different kind of reports in order to fully adapt our tool to the different environments and use cases. All the reports generated contain the same information:

- The structure of the TIFF file with all the tags and information about the metadata or files embedded in the TIFF.
- The results of the implementation checker and policy checker.
- A metadata analysis looking for metadata incoherence.
- A history from the file extracted from the metadata containers.
- The file identification information.
- And, if the metadata fixer has been used to create a new file, a report of this new image.

We have two human-readable versions: one in HTML format that we use as main output in the GUI interface. This report contains all the information with a nice look and feel and includes interactive features not possible in other formats. This report is thought by non-experts users, showing the most relevant information. But, it can be also used by experts that want to see all the information inside the TIFF activating the expert mode. This report, like the others, can be exported and read by any current browser. The other human readable report format is PDF, containing the same information as the HTML report although less interactive.

We also have two machine readable versions of the report, one in XML and another in JSON. These reports are ready to use by computers, which can extract the report information and act according to its information.

In addition, we designed a new report with the collaboration of the memory institutions, available in this new release, with the aim to create a robust, standard, sustainable, integrable and interoperable report. We thought that it is important to report and present the information in an archival language using the current standards as well as provide a way to fully ensure the integration in the OAIS model. The METS standard fulfil all these requirements.

The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed with XML. This standard is maintained in the Network Development and MARC Standards Office of the Library of Congress, and it is being developed as an initiative of the Digital Library Federation.

METS standard can be used in the role of an Archival Information Package (AIP), Submission

Information Package (SIP) or Dissemination Information Package (SIP) within the Open Archival Information System (OAIS) reference model, which ensures the DPF Manager integrability with Memory Institutions applying the OAIS model in their data object lifecycle. The Mets report only provides a base, therefore, we also included into the METS report the PREMIS data dictionary. PRIMES is used to report the DPF Manager conformance check information; including when the check was made, which version was used, the configuration applied and the resulting XML report (essential information for the process could be repeated producing the same output). In the METS descriptive metadata section the NISO and Dublin Core standard are used to describe all the technical and descriptive metadata about each image contained inside the TIFF.

2. Datasets used to test the release

During all the project we have been applying different tests to our releases trying to ensure the proper operation of the DPF Manager. First of all, we apply unitary tests. They are responsible to test every relevant function in the code and ensure they produce the same outcome in every code contribution. Then, we also apply Graphical user interface tests. All these tests evaluate that every functionality in the application is still working after any source code contribution. Travis, our continuous integration tool, is properly configured in order to test all the modules including: interface, implementation checker, policy checker, metadata fixer and reporter, detecting whether a contribution conflicts with the current implementation or breaks any previous functionality. Although we internally still use unitary tests, in this new release all the unitary test has been removed from the source code while we are waiting for a final answer form the PREFORMA consortium in relation to licenses.

Regarding our TIFF reader/writer and implementation checker, we have generated out an image dataset in order to test the application. This dataset is distributed in our repository using the Attribution-ShareAlike 4.0 International licence. Creating our own image dataset is not enough because the testing process could be self-referenced and influenced about our knowledge about our implementation. To avoid this inconvenience, internally we use another dataset to test the application. This dataset is not public for licence reasons since the origin of those images come from different places (libtiff test images, file format samples, google code repository to test tiff lib implementation, jhove, jhove2, artificial data sets…).

During our research in the definition of TIFF error classes, we found another source of TIFF malformed images for testing, the TIFF read/writers forums and issue trackers. Many users when they found malformed TIFF that produce an error in a TIFF reader, report this issue in the issue tracker or forum. Then, we can obtain directly malformed TIFF (i.e., here) or reproduce the error reported by the TIFF writer (i.e. here). This is a way to generate malformed TIFF that we could found in the real life.

Then, in the previous releases of DPF Manager we asked early adopters to share their reports results in order to test the application. These reports are not only used to test the application but also to found new private tags, private vendor compressions or photometric interpretations, uncommon tiff structures, …

Finally, our most intensive test come from Basel university in collaboration with the Swiss Federal Archive and its Coordination Committee for sustainable archival of digital documents (KOST, kost-ceco.ch). During their research of the state of art about which kind of TIFF are preserving memory institutions, 3.9 Million existing TIFF files in Swiss archives have been checked. Some results of this analysis can be found in the last TI/A draft specification (annex 6). Soon they will be published as a whole report.

All these image datasets helped us to create a robust application able to handle with different kind of TIFF in many different situations.

3. Dissemination and community building

We are moving forward in our strategy of creating archivist awareness around TI/A. A first draft of the specification was published on September 30th and the community of experts started to discuss some of the decisions, that helped us to achieve a consensus for the second draft. Although this version was sent to the ISO Working Group to be discussed during the meeting in Sydney next end of November, this specification is still a living document, where decisions can change as more experts give their opinions.

One of the main dissemination activities from now is to communicate that DPF Manager is the unique tool that validates the first specification of TI/A, helping memory institutions to guarantee the preservation of their TIFF files.

Until now, the impact of our dissemination around TI/A has been:

- The TI/A website has received 6.500 visits of more than 4.500 unique visitors in 18 months, with an increment around 25% after the publication of the 1st draft.
- The twitter account has 410 followers and we have published 850 tweets related to the TI/A initiative and digital image preservation with around 170.000 impressions. The growth of the followers is around 20% during the last month.
- In the TI/A Intranet we have 80 experts registered from 17 different countries.
- In the last 3 months, TI/ Initiative news have been published in external blogs like Digital Meets Culture, Open Preservation Foundation, Digital Bevaring, Twicopy, Europeana Blog, Beel En Geluid, MediaConch Blog, Samradsgruppen, Digital Preservation CZ, Digital Preservation News, Digital Preservation Matters and Kulturreste.

To raise awareness amongst the scientific community, we have presented the TI/A and DPF Manager initiatives in the following conferences:

- De la Rosa, Peplluís, "Preservación digital como servicio (SaaS)", **Congres d'arxivistica de Catalunya**, 28th to 30th May 2015, Lleida (Spain)
- Fornaro, Peter, Rosenthaler, Lukas, "*Archiving Digital Image Data and Motion Picture: Concepts and Solutions*", **International Conference on Advanced Imaging**, 17th to 19th June 2015, Tokyo (Japan)
- Fornaro, Peter, Rosenthaler, Lukas, "*Long-Term Preservation and Archival File Formats: Concepts and Solutions*", **IS&T Archiving 2016**, 20th to 22th April 2016, Washington (USA)
- Fornaro, Peter, Rosenthaler, Lukas, "*File Formats for Archiving: Stability and Persistence Issues*", **Digital Humanities Conference 2016**, 12th to 18th July 2016, Krakou (Poland)
- Zbinden, Erwin, "*Quality Standards for Preserving Digital Cultural Heritage*", Workshop on **iPRES Conference 2016**, 3th to 6th October 2016, Bern (Switzerland).

All these events helped us to introduce our initiative to TIFF experts and potential users, especially memory institutions and to create a community of people willing to use the DPF Manager to analyse the long-term preservability of their image digital assets.

We have to stress that some institutions are giving us valuable support and collaboration to achieve our objectives: Adobe Systems Inc., the Swiss Coordination Centre for the Long-Term Preservation of Electronic Documents (KOST-CECO - http://www.kost-ceco.ch/) and some of the biggest and most important Swiss Archives.

Besides to all the dissemination performed for the TI/A initiative, the advanced current status of the DPF Manager development allowed us also to disseminate the features and advantages of this tool. During the last months, our communication campaign has been oriented to hook early adopters and, what it is more important, to get feedback from them. We also obtained feedback from the automatic reports that we received from the DPF Manager and from the PREFORMA committee. All this feedback helped us to improve our tool to finally achieve the current functional and robust version of the DPF Manager.

The monitoring of the dissemination activities revealed us that as the final release deadline approached, more and more activity from the community we observed (between 35% and 75% of increment depending on the channel in the last 3 months).

From October 2015, the DPF Manager website received 10.000 visits (35% of increment last 3 months) of more than 4.200 unique visitors and the twitter account has over 230 followers from different countries in Europe (increasing 75% only in the last month), mostly archivists and developers belonging to memory institutions. Our @dpfmanager tweets have a total amount of 96.889 impressions from October 2015 (+50% in the last 3 months).

We have also released 9 specific newsletters (1 in October 2015, 1 in November 2015, 2 in April 2016, 2 in May 2016, 1 in July 2016, 1 in September 2016 and 1 in October 2016) containing only news around DPF Manager with a target audience of 1.545 experts in digital preservation, developers, researchers and memory institutions. We also published several blog posts in the DPF Manager Blog mainly related to the new software releases, which have also been communicated via Twitter. We also prepared a webinar about DPF Manager organized by OPF Foundation and some of our presentations are available on Youtube.

Now, the final release is out and starts a new stage: communication focussed on achieve as many final users as possible. This is the moment that everybody was waiting for. Sometimes some institutions do not want to use a tool that it is still in development trying to avoid initial issues. Now there is no excuse for this institutions and it is time to start using it.

In the following days/weeks we are going to enhance the communication in the different channels talking about the final release, the process followed to arrive until the current version of the DPF Manager, attracting final users, contacting associations, communicating next events, engaging more experts discussing about TI/A recommendations, publishing new versions of the TI/A specification, even publishing in the regional press and, of course, more and more activity on the social networks, always looking for the maximization of the impact.

Now also starts our communication campaign to attract developers that want to contribute in the development of the project. The main strategy is simple, if memory institutions start using DPF Manager, probably they will require some modifications/improvements/new functionalities. Developers (internal or external) supporting these institutions will be requested to help them and then, we have to give developers as many facilities as possible to do it. During the last 2 months we have dedicated a lot of resources to this objective, documenting everything, modularizing as much as possible the source code, creating guidelines, … and publishing all this documentation on the website and GitHub.

4. Open Source approach

Provisioning the source code

The University of Skövde reported some contributions made under other licences different than the specific licences required by PREFOMA ("MPL v2.0 or later" and "GPL v3 or later"). We immediately solved this issue and we updated all the files with the required header license information.

Provision of 'roadmap' on the development platform

Since we manage the DPF manager roadmap in GitHub, we tried to make is as much generic as possible where external contributions were always welcome. Of course some tasks were specific assigned to our internal collaborators but most of them were unassigned. We have updated the GitHub roadmap and milestones plans for the time period until 2020 with a release plan which targets external contributors. We expect that the release of the TI/A draft implementation will promote the collaboration of external contributions aiming to integrate the DPF Manager in their software.

Identical software under both "GPLv3 or later" and "MPLv2 or later"

In order to fulfil the requirements meanwhile we wait for a final answer to our question to the University of Skövde, we removed our unitary test that use a testing dependency with Junit framework under EPL licence. If needed, we could easily redistribute again the unitary tests.

The Skövde partner reported some dependencies from libraries under Apache 2.0. In Easy Innova deliverables of the phase 1 we included a list of our dependencies that wolud be used during the Prototyping phase 2. The document named "Intellectual property Rights report" included the review and analysis from the perspective of IP law including:

- Authorship and rights in the proposed solution
- Listing of proposed third party components and identification of the relevant licenses
- Analysis of the architecture of the proposed solution and interrelations between software components
- Analysis of the license obligations and their compatibility with the required license for the solution
- Compliance with the licensing obligations of the different components of the solution with the required dual free software license GPL3+/MPL2+
- Advice on the implications and process for delivering the proposed solution to PREFORMA under the dual free software license GPL3+/MPL2+
- Advice on free documentation license (Creative Commons)

This report, written by Malcolm Bain, a reputed IT law, with a focus on legal issues of free software, concludes that libraries under Apache 2.0 licence are compatible with GPL3+ and MPL2+ licenses, both in compiled (static linking) or non-compiled/dynamic linking scenarios.

We know that software under Apache 2.0 licence, a permissive FOSS license (approved by Open Source Initiative), is incompatible with the use of libraries under GPL3+ or MPL2+ (Strong copylefts licencies) but, software under GPL3+ or MPL2+ can use libraries under Apache 2.0 and be freely and legally distributed.

We are still waiting and answer from the consortium. As soon as we have a final decision we will work to solve these inconveniences.

Provision of executable of the software on the open source portal

The executables distributed since now included the java runtime in order to avoid java dependencies. In this release the runtime is no longer included in the executables, in order to fulfil the PREFORMA licenses requirements, now the user have to resolve the dependency installing the java oracle JDK 8 with comes with JavaFX or compile the OpenJDK and OpenJFX.

Provision of detailed documentation concerning interpretation of the technical specification of each file format

Now we have explicitly published all the rules used to validate the implementations in our web page documentation. This information directly comes from the GitHub folder that contains the current files used to validate the specifications. These files are in an xml format that could be easy read by any text editor. In order to facilitate the navigation, we included a search box able to search by any text in the rules description, title or error code.

Provision of software which can be redistributed in a cascade

All the issues about provisioning the software which can be redistributed in a cascade are regarding the issues described before. Ones we solve the previous issues the software will be redistributed in a cascade with the PREFORMA licence requirements.

5. Standardisation efforts

As known, we are working on the definition of a technical specification (recommendation) for the use of TIFF in archival environments, called (TI/A http://www.ti-a.org), to be published as an ISO recommendation. It is important to note that the final result is not a specification of a new file format but a recommendation of the proper use of the existing TIF format. During the last 3 months several tasks could be finished and the standardization process progressed to a large extend.

We could analyse a large set of TIF files from different memory institutions. The assets of files are of big importance to understand the history of TIFF regarding the application and selection of tags. It would be of disadvantage to recommend any set of tags only from the point of view of today because of a general technical advance in the field of imaging. It is important to have a deep view of real data stored by memory institutions as objective as possible because those assets do represent our data of interest that shall be tested by DPF manager in future. The outcome of the analysis (a feature histogram) is interesting in multiple aspects: It is clearly visible that the selection and use of technical features of TIFF has drastically changed over time. A couple of years ago the Group 4 Fax compression was a feature that has been chosen very often. Files with a quantisation depth of 16 bit were on the other side very rare, what is a strong contrast to today's best practice. LZW or JPG compression can only rarely be found most likely due to a strong lack of acceptance of such compression schemes since the early years of digital archiving.

We had intensive discussions with various experts in digital archiving, especially on the Intranet but also in Twitter. Particularly, the most controversial topic of discussion was "to compress or not to compress". The most important message and outcome is that the recommendation that we will hand in to ISO will not go into technical details of image compression nor will it force any institution not to use image data compression as long as the technology applied can be regarded as technologically stable, well documented and supported by typical software. The compression scheme that is not acceptable by any means is "old JPG". As a conclusion it can be said that the recommendation shall not contain any "questions of faith", but only conclusions based on existing data and facts coming from the community of archivists and users of real applications.

As a consequence, the Draft of the recommendation has been brought to a pre-final stage. The final draft is available at the TI/A intranet. This document was already submitted to the ISO TC171 by October 28th 2016, to start the official process of standardisation. This deadline was imposed by the Technical Committee if we wanted to include the discussion about TI/A in the next working group meeting to be held in Sydney from Nov 28th to Dec 2nd. Thus, from the ISO point of view, the project clock already started.

6. Impact assessment, sustainability, future use and exploitation

As stated in the previous report, we have an important handicap regarding sustainability, future use and exploitation: the low technical knowledge about TIFF format leads memory institutions to be not aware about the lack of conformance of their TIFF assets to the standards and about the most probably unsuitability of their tagged image files for long-term archival. Therefore, our first activity has been to create awareness. The TI/A initiative precisely aims to build a set of recommendations for memory institutions certified by the International Organization for Standardization (ISO) with the aim of guarantee the long-term preservability of their TIFF files. This ISO Recommendation indeed is our ultimate weapon to create awareness among archivists in the same way that PDF/A did some years ago.

Taking into account that DPF Manager is now the first tool to validate these ISO Recommendations, we ensure sustainability and future use of our software.

Memory institutions are increasingly understanding that they have to validate their TIFF files conformance to the standards and to the TI/A ISO Recommendations. We had a large list of early adopters using DPF Manager before the final release. Some of them are National Archives of Sweden, Aquaforest Limited, Oregon State University Libraries, bj institute, MIT Libraries, MoMu, Hochschule der Künste Bern, Royal Museums of Fine Art of Belgium, Technical University of Viena, National Archives of Denmark, City Council of Stockholm Archive and University of Pittsburgh. The last ones to use DPF Manager have been the Swiss Federal Archives and the cantonal archives of Basel City and Saint Gallen. We have to say that not all the users that download our tool are registered, thus we have a lot of unknown early adopters.

These early adopters allowed us to receive 4.081 reports of **11.603** files analysed (5.808 with correct baseline, 513 with correct TIFF/EP and 1.409 with correct TIFF/IT) where we could discover 18 private tags and 16 typical errors in the baseline. The increment of use by early adopters in the last months is notorious, they analysed more or less the same number of files in 2 months than in all the year before.

It is important to mention that DPF Manager has been used intensively in two cases:

- Packed validated the conformance of around **40 thousand** TIFF files of scanned paintings for *La Fundació Tàpies* (http://www.fundaciotapies.org/).
- The University of Basel analysed almost **4 Million** TIFF files from 3 big memory institutions in Switzerland in order to understand which variants and tags have been used in the past to create TIFF files.

This means that until now, more than 4 Million TIFF files have been analysed using DPF Manager, a crucial test for us in order to validate the robustness of the tool.

For us it is also very important the collaboration with associations or entities protecting the interests of memory institutions, as in the case of KOST-CECO in Switzerland, which is managing the digital preservation issues of their 30 members.

We are also collaborating with Europeana Technical Space project, which is using our DPF Manager to validate the TIFF files uploaded by memory institutions. The reports generated from these analyses are really helpful for us to improve our tool if necessary.

The stage where we were looking for early adopters finished. The final release is out and all the functionalities (including the TI/A first specification) are available. Now we are looking for final users that want to take profit of all the amazing functionalities of the DPF Manager. Our communication strategy is now totally oriented to this objective.

The long term success of the DPF Manager depends on establishing a successful community around the open source project and also on developing a set of commercial services to ensure the project doesn't end up like JHove that died as soon as the public funding stopped. In the documentation we submitted at the end of Phase 1 we included a brief business plan where we outlined that our exploitation plan is based on offering services like Cloud-based SaaS, on premise deployments, technical support and maintenance contracts, consultancy services and training courses to developers, integrators and end-users. Although we have simplified the use of the conformance checker compared to our main competitor JHove, memory institutions would need the technical support and knowledge of a specialised company to integrate, maintain and evolve their long-term preservation systems. In order to offer our services in the near future, we already registered the domain www.dpfmanager.com.

The plan for the next months is to improve and complete the initial business plan. For us it is very important to have a very realistic and comprehensive business plan that will guide our strategy going forward. Our main goals are:

- To elaborate a market study to understand our customers (libraries, archives, memory institutions) and survey end users to estimate more precisely the demand.
- To define the best commercial strategies to reach the target customers
- To identify potential partners
- To make an exhaustive financial plan that will define a realistic roadmap for the exploitation stage.
- To develop a complete business plan.

7. Gap analysis and next steps

We successfully finished all the tasks described in the Phase 2 proposal and we also finished successfully some new functionalities and improvements not initially planned.

Regarding the last period, we developed the interoperability among conformance checkers (Tasks 1.3.1.2.3) and (Task 1.3.1.2.4), we updated the global report in order to incorporate the other conformance checkers reports (Tasks 1.3.1.3.1) and we updated our configuration file using an xml format (Tasks 1.3.1.4).

Apart from finishing all the implementation checkers (Tasks 1.3.2.1), including the file identification (Task 1.3.2.1.5), we divided the Baseline 6.0 specifications according the two parts described in the document and we also developed a new functionality to select which rules are going to be evaluated in each implementation. Then, we also created new policies (1.3.2.2.1) to be evaluated by the policy checker module.

We then created the new METS report (Task 1.3.2.3.2) in order to provide an integration with organizations using the OAIS model.

Regarding the metadata fixer, now we are able to read and write all metadata containers inside TIFF (Tasks 1.3.2.4.1 and task 1.3.2.4.2). This task was required in order to complete the metadata inconsistency report and fix (Task 1.3.2.4.3).

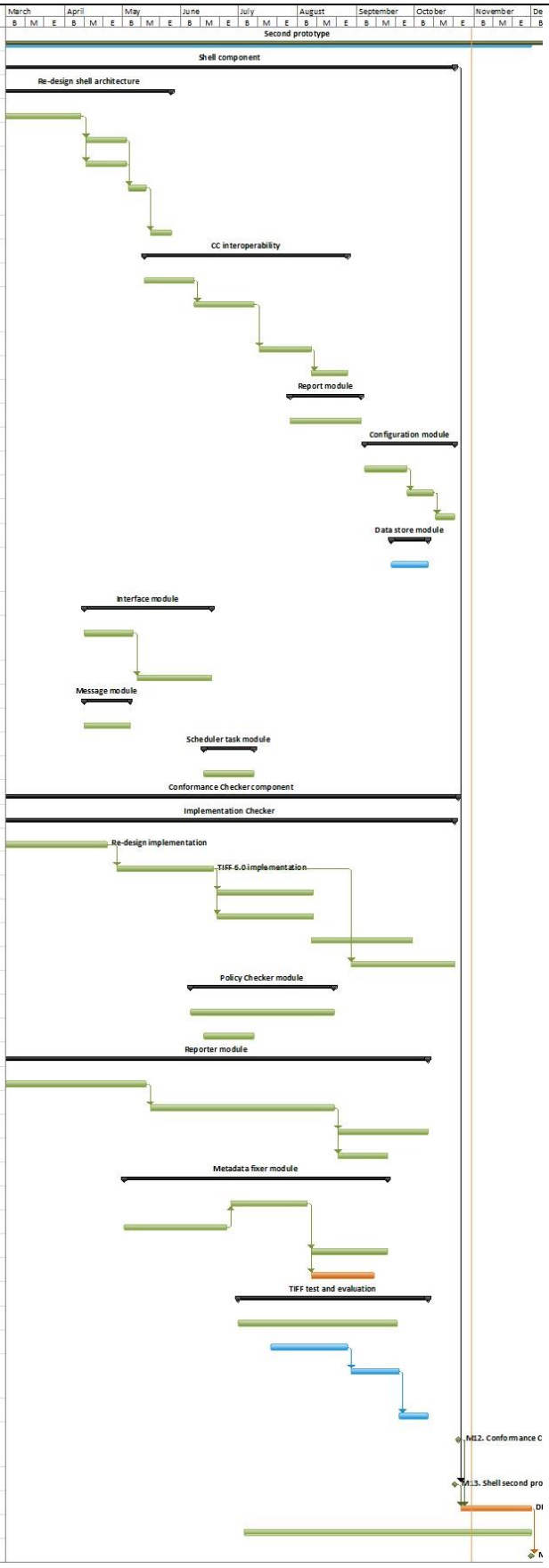
We finally updated all the documentation and generate the documentation from the current implementations (Tasks.1.3.5).

Although our report now includes a file provenance section (Task 1.3.2.4.4), we are more ambitious in this functionality. We want to use the metadata inconsistence analysis to detect changes not explicitly reported inside the TIFF metadata. This is something that we want to develop in the coming releases.

There is only one tasks that has not been developed yet, the integration on the DIRECT system. We did not receive a concrete documentation about the DIRECT infrastructure but, taking into account the modular architecture of the DPF manager, this integration should be ready in a few days. Then, we delay the construction of a website to analyze the performance of the DPF manager compared with other conformance checker tools. It is important to say that this task was not initially planned during this phase, although we thought during the re-design phase that was a nice to have. We think it could be an interesting development for the next testing phase, thus we are going to include it in our tender proposal.

Just to overview the status of the development, in the following Gant diagram we show the pending tasks in blue, tasks in progress in orange and tasks done in green. All the task required for the Prototyping Phase 2 have been developed and we have included many functionalities not required in the tender specification but required form the feedback received from memory institutions during this phase.

Name	Start
1.3 Second prototype	**01/03/16**
1.3.1 Shell component	**01/03/16**
1.3.1.1 Re-design shell architecture	**01/03/16**
1.3.1.1.1 Modular architecture	01/03/16
1.3.1.1.2 Multithread application	12/04/16
1.3.1.1.3 Event-driven	12/04/16
1.3.1.1.4 Common architecture for interfaces	04/05/16
1.3.1.1.5 Multi instance app	16/05/16
1.3.1.2 CC interoperability	**13/05/16**
1.3.1.2.1 Multiple CC	13/05/16
1.3.1.2.2 Command line interoperability	08/06/16
1.3.1.2.3 API interoperability	12/07/16
1.3.1.2.4 Server interoperability	08/08/16
1.3.1.3 Report module	**28/07/16**
1.3.1.3.1 Global report	28/07/16
1.3.1.4 Configuration module	**05/09/16**
1.3.1.4.1 Configuration controller	05/09/16
1.3.1.4.2 Get Conformance	27/09/16
1.3.1.4.3 XML validation	12/10/16
1.3.1.5 Data store module	**19/09/16**
1.3.1.5.1 Implement DIRECT system key value data store	19/09/16
1.3.1.6 Interface module	**11/04/16**
1.3.1.6.1 Graphical user interface redesign and extension	11/04/16
1.3.1.6.2 Server interface	09/05/16
1.3.1.7 Message module	**11/04/16**
1.3.1.7.1 Log system implementation	11/04/16
1.3.1.8 Scheduler task module	**13/06/16**
1.3.1.8.1 Implement scheduler task	13/06/16
1.3.2 Conformance Checker	**01/03/16**
1.3.2.1 Implementation Checker	**01/03/16**
1.3.2.1.1 Re-design implementation	01/03/16
1.3.2.1.2 TIFF 6.0 implementation	28/04/16
1.3.2.1.3 TIFF/EP implementation	20/06/16
1.3.2.1.4 TIFF/IT implementation	20/06/16
1.3.2.1.5 TIFF Identification	08/08/16
1.3.2.1.6 TI/A implementation	29/08/16
1.3.2.2 Policy Checker module	**06/06/16**
1.3.2.2.1 New policy rules	06/06/16
1.3.2.2.2 ICC Profile policy rules	13/06/16
1.3.2.3 Reporter module	**01/03/16**
1.3.2.3.1 Report Re-design	01/03/16
1.3.2.3.2 Report XML	16/05/16
1.3.2.3.3 Report PDF	22/08/16
1.3.2.3.4 Report JSON	22/08/16
1.3.2.4 Metadata fixer module	**02/05/16**
1.3.2.4.1 Detect metadata	27/06/16
1.3.2.4.2 R/W embeddedmetadata	02/05/16
1.3.2.4.3 fix metadata inconsistency	08/08/16
1.3.2.4.4 Detect file provenance	08/08/16
1.3.2.5 TIFF test and evaluation	**01/07/16**
1.3.2.5.1 prepare TIFF classes	01/07/16
1.3.2.5.2 Test and evaluation	18/07/16
1.3.2.5.3 DPF Manager implementation comparison	29/08/16
1.3.2.5.4 TIFF test web page	23/09/16
1.3.2.6 M12. Conformance Checker second prototype release	24/10/16
1.3.3 M13. Shell second prototype	21/10/16
1.3.4 DPF manager test	25/10/16
1.3.5 DPF manager documentation	04/07/16
1.3.6 M14. DPF final prototype	30/11/16



Next steps from the development point of view are focussed on our proposal for the phase 3 tender, promote the collaboration of external contributions to ensure the long-term sustainable of open source communities as well as promote the integration of the DPF manager into Digital Asset Management software.

We created a new roadmap specially designed for external contributions. This roadmap proposes new functionalities to be developed for the time period until 2020. The roadmap is published at GitHub using Projects, one of the latest functionalities added in Github. Projects let to apply Kanban methodology inside GitHub organizing new features in tasks to be developed by contributors.

Some of the new features that we will like to promote for the external contributors and collaborate with them are the following ones.

- DNG implementation checker. DNG is an open format from adobe to store RAW information. DNG use the same structure and share tags with TIFF, so it is easy to develop a DNG implementation checker using the DPF manager TIFF implementation
- There are some ISO's TIFF-F (RFC 2306) and TIFF-FX (RFC 3949) about TIFF out of the scope of PREFORMA project but that can be easy integrated in our implementation checker, we want to promote their implementation in order to be able to identify this kind of TIFF.
- Implementation checker rules based on Xpath and XQuery selectors. Now our implementation checker user custom rules, we want to include different kind of selector in order to provide more flexibility.
- Integration of the DPF Manager into the Archivematica: open-source digital preservation system.

Regarding TI/A Initiative, our aim is to convert our current proposal to a ISO Recommendation. This is a long travel but we are very close, at the end of November we present the proposal to the ISO Working group in a face-to-face meeting in Sydney. We hope good acceptance (we have Adobe on board) to follow with the process.

Finally, from the community point of view, now all our communication will be oriented to attract new DPF Manager users (memory institutions and software developers) and get feedback from them. Having in mind that at the end we want to earn money from the service behind the DPF Manager, understanding the memory institutions needs and reacting to their requests is crucial.