DELIVERABLE

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

Deliverable D8.5 Final Prototype Report

Revision: 1.0 FINAL

-

Author:

> **Borje Justrell, Riksarkivet**

Contributors:

> **Magnus Geber, Riksarkivet**
> **Claudio Pandoni, Promoter**
> **Benjamin Yousefi, Riksarkivet**

Reviewers:

> **Peter Pharow, FRAUNHOFER**
> **Erwin Verbruggen, Sound and Vision**
> **Eva McEneaney, LGMA**

Revision History

Revision	Date	Author	Organisation	Description
0.1	08/12/2016	Borje Justrell	Riksarkivet	Structure of the document.
0.2	05/01/2017	Borje Justrell	Riksarkivet	New content in chapters 1 - 3 and Annex 1.
0.3	16/017/2017	Borje Justrell	Riksarkivet	Changes in chapters 3 and 4.
0.4	18/01/2017	Borje Justrell	Riksarkivet	Included in chapters 1 – 4: new text from Claudio Prandoni and Magnus Geber, and review comments from Erwin Verbruggen.
0.5	07/02/2017	Borje Justrell	Riksarkivet	New content in chapter 5; text from Claudio Prandoni included in chapter 4.
0.6	10/02/2017	Borje Justrell	Riksarkivet	Smaller changes and updates in chapter 1-4 and revised text in chapter 5.
0.7	15/02/2017	Borje Justrell	Riksarkivet	Changes in chapter 4 and 5 based on comments from Magnus Geber, Benjamin Yousefi and Claudio Prandoni.
0.8	18/02/2017	Borje Jsutrell	Riksarkivet	Changes mainly in chapter 5 based on comments from Claudio prandoni. Smaller changes in chaper 2- 4 and the executive summary
1.0	21/02/2017	Claudio Prandoni	Promoter	Final version integrating the comments and revisions by Eva McEneaney

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

PREFORMA - Future Memory Standards
PREservation FORMAts for culture information/e-archives
EC Grant agreement no: 619568

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This deliverable presents the results of the second part of the Prototyping phase in the PREFORMA project, thereby to some extent also summarising the outcomes of this phase in its entity. The document consists of two parts:

The first part (chapters 1 to 3) gives the overall context, including the aims and objectives of the second part of the Prototyping phase. It also reports on formal procedures, like the discussions with the suppliers and internally in the PREFORMA Consortium, and presents specific issues of general nature in more depth.

The second part (chapters 4 to 5) focuses on the results achieved during the Prototyping phase, particularly its second part, but also on progress in the development of the PREFORMA prototypes, which will be further advanced in the Testing and validation phase.

1 INTRODUCTION

1.1 DOCUMENT OVERVIEW

This document presents the final results of the Prototyping phase. It consists of five chapters:

1. **Introduction** – This chapter presents the aims and objectives of the deliverable but also some steering documents for the 2nd part of the Prototyping phase;

2. **Formal Procedures** – This chapter clarifies the starting position of the Prototyping phase as such and the internal process of managing it, with focus on the 2nd part.

3. **Specific Issues** - This chapter discusses some issues of specific nature covered by the objectives of the Prototyping phase;

4. **Results of the Prototyping phase** - This chapter gives an overview of the results achieved during the 2nd part, and of the evaluation of the final results of this major phase;

5. **Points on Progress** - This chapter concludes the progress and results made during the Prototyping phase.

This document also provides eight annexes:

Annex 1: Work plan for the second part of the Prototyping phase agreed at the first supplier meeting during the 2nd part of the Prototyping phase

Annex 2: GitHub guidance document

Annex 3: Template for Intermediate Reports

Annex 4: Template for Final Reports

Annex 5: Template for End of Phase Reports

Annex 6: Template for the evaluation

Annex 7: Roadmap for the second part of the Prototyping phase

Annex 8: Evaluation Report

1.2 AIMS AND OBJECTIVES

In deliverable D8.2 "*Design – First report*", the process of evaluating the bids for PREFORMA major phase 2 (Prototyping) and its outcomes is described. The media types chosen for the Prototyping phase are: texts (PDF/A-1, PDF/A-2 and PDF/A-3), still images (uncompressed TIFF) and audio-visual records (container formats MKV; codec FFV1; audio format LPCM).

Three suppliers were invited to sign a contract for the Prototyping phase:

- veraPDF Consortium (led by the Open Preservation Foundation and PDF association) for media file type text;

- MediaArea for media file type audio-visual;

- EasyInnova for media file type still images.

The Prototyping phase belongs organisationally within Work Package (WP) 6 of the PREFORMA project. The three selected suppliers have, therefore, conducted the prototyping of the PREFORMA software under this Work Package.

Formally, WP6 is composed of two tasks, which in the PREFORMA´s Description of Work (DoW) are named **T6.1 Prototyping Step 1** and **T6.2 Prototyping Step 2**, in this deliverable they are called 1st and 2nd part respectively. The Work Package leader and leader of both tasks has been Riksarkivet.

According to the DoW (Description of Work), suppliers are expected to provide software prototypes that fulfil the requirements of the project, demonstrate the results and provide documentation on how the developed software can be effectively used by and at memory institutions. Furthermore, suppliers are expected to utilise best practices from open source development, which include the use of:

- an open work practice for development;
- frequent open releases;
- promotional activities aiming towards a sustainable community.

Finally, the DoW explains that suppliers are expected to establish a feedback process between relevant standardisation organisations and other stakeholder groups. Such feedback may include implementation notes that detail the interpretation of the standard specification (especially for sections of the standard specification that are unclear) during software development.

Task T6.1 was finished by the end of October 2015, and the results reported in deliverable D8.3 *First Prototype repo*rt. It was followed by the Re-design phase. Task T6.1 was overseen by Riksarkivet and executed by the suppliers. It aimed to deliver and demonstrate the first three prototypes, which included:

- the four modules ("Implementation Checker", "Policy checker", "Reporter", and Metadata fixer") that every developer provided for their selected media type (documents, images or audio-video);
- the web application to demonstrate the modules;
- the documentation of the open source software.

Task T6.2 followed the Re-design phase. The aim was to deliver and demonstrate improved versions of the first prototypes developed during the 1st part of the Prototyping phase. These improved versions were to be ready to enter the Testing and validation phase that will be conducted by WP7.

The objective of this deliverable (D8.5) is to provide a report on the improved versions of the three prototypes with information on how the suppliers have:

- provided required functionality;
- established a process for feedback to standardisation organisations;
- adhered to utilising best practices from open source development. In this regard, the present deliverable is supplemented by an updated version of deliverable D8.8 *Monitoring of the Open Source Project Implementation*.

The aim for this deliverable is also to serve as a basis for the start of the final Testing and validation phase.

1.3 STEERING DOCUMENTS FOR THE 2ND PART OF THE PROTOTYPING PHASE

Aside from the Call Documents (Invitation to Tender, Challenge Brief, and Framework Agreement), several other documents have been particularly important for the work preformed during both the 1st part and the 2nd part of the Prototyping phase.

Firstly, the deliverables PREFORMA has submitted, among which the following have provided direction for the work carried out:

- D2.1 *Overall Roadmap*, which provides an overall roadmap for the preparation of the request for tender and the selection of the technology suppliers invited to take part in the project. It offers an overview of the legal and operational procedures and describes the process for gathering, analysing and defining the functional and technical specifications to be used in the Invitation to Tender. During the entire Prototyping phase D2.1 has been an important document when measuring software quality assurance and accuracy.

- D2.2 *Tender Specifications,* which presents the tender requirements and assessment procedures included in the Call for Tender of the PREFORMA Pre-Commercial Procurement (PCP). By compiling the content from the tender documents in one deliverable its aim is to provide:

  o a clear description of the research and development component of the PREFORMA PCP and the relationship between the PREFORMA Challenge and the PREFORMA Tender;

  o the scope and a detailed description of the challenge that the PREFORMA PCP addresses.

  D2.2 has been used as check list to evaluate whether the requirements communicated by PREFORMA have been fulfilled in the software releases submitted by the suppliers.

- D8.1 *Competitive Evaluation Strategy*, where the evaluation framework is defined based on contributions from technical partners and memory institutions. The strategy described was used to evaluate the results at the end of the Design phase in order to select the suppliers who completed the tender. This strategy has also been used when reviewing the suppliers' results during the Prototyping phase.

- D8.2 *Design – First Report*, which builds on the results of the work with the suppliers as well as the work of the suppliers themselves during the Design phase. An important element for the continuity into both the 1st and the 2nd part of the Prototyping phase was that D8.2 also describes the way (methods, measures, principles) the PREFORMA Consortium carried out the formal review, the evaluation, and the preparatory work to decide which suppliers should be invited to submit bids for the Prototyping phase.

- D8.3 *First Prototype Report*, which presents the results of the 1st part of the Prototyping phase. This deliverable summarise the discussions with the suppliers and internally in the PREFORMA Consortium and the results achieved by the suppliers. It analyses the progress made, and resume the outcome of the 1st part of the Prototyping phase, which indicates that the project is moving in the right direction.

- D8.4 *Design – Final Report*, which points out the design changes made by the suppliers to be in line with the feedback provided by WP6 on their first prototypes. Moreover, it also shows how satisfactorily they have addressed this feedback. The evaluation strategy negotiated and established in task T8.2, and described in deliverable D8.2, was

used as an input for informally evaluating the suppliers' results at the beginning of the Re-design phase. The purpose was to evaluate whether the suppliers have fulfilled what they were expected to do.

- First release of D8.8 *Monitoring of the Open Source Project Implementation*, which provides more in-depth feedback on whether the suppliers, during the 1st part of the Prototyping phase, had adhered to the requirements in deliverable D4.3 *Functions of the Open Source Portal*. D8.8 also further evaluates how each open source project has progressed in implementing open work practices for software development.

Secondly, during the Prototyping phase, PREFORMA started to substantiate, in internal documents, important issues not covered by formal documents such as the DoW or submitted deliverables. These internal documents discuss approaches and positions expressed internally within the projects. They can be seen as "work in progress". The most important internal documents are:

- *Data Management Plan for training, testing and demonstration files in the PREFORMA project* (DMP) developed in cooperation with the suppliers. It describes the framework that governs the provision and management of the files to be used to test the prototypes. It also:

  - outlines the distinction between synthetic and organic files;

  - defines the different usage types (training, examination, evaluation, dissemination);

  - describes the workflow for the data provision, storage and orchestration (including responsibilities) of the different actors;

  - analyzes possible legal and copyright implications.

.   An earlier version of the DMP was attached as an annex to deliverable D3.3 *Networking Report Year 2.*

- *Legal opinion*. The purpose of this document is to consolidate the legal sources of the PREFORMA project, to provide analysis and comments of the legal sources, and to conceptualise the PREFORMA project in order to identify possible problems and clarify possible concerns. Like the DMP it is "a work in progress", and is updated as the project advances.

2   FORMAL PROCEDURES

The purpose of this chapter is, firstly, to clarify some basic conditions for the Prototyping phase and, secondly, to present the overall management with special focus on the 2nd part of it, by summarising the discussions held at meetings during this period - with the suppliers as well as within the PREFORMA Consortium.

2.1   STARTING POSITION FOR THE PROTOTYPING PHASE

2.1.1   PREFORMA project phases

The PREFORMA project is carried out in three major phases: A Design phase, followed by a Prototyping phase, and ending with a Testing and validation phase.

The purpose of the first major phase, Design, was to demonstrate the feasibility of the proposed concepts for the solutions posed by the suppliers in their original bids. It took place between November 2014 and February 2015. The six suppliers that were chosen participated in the Design phase under competition like forms. At the end of the phase, the PREFORMA Consortium, based on a formal evaluation, selected three suppliers out of the six participating candidates. The intention of the second major phase, Prototyping, was to develop prototypes based on the functional and technical specifications provided by the suppliers during the Design phase. The Prototyping phase was subdivided into three distinct stages:

- Prototyping phase, part 1, which took place between April and October 2015;
- Re-design phase, which took place between November 2015 and February 2016;
- Prototyping phase, part 2, which began on the 1st of March 2016 and lasted until the end of January 2017.[1]

During this second major phase, the three chosen suppliers had to prove that their respective applications could meet the PREFORMA Challenge set forth in the Challenge Brief.

Finally, in the third major phase, Testing and validation, that will take place for six months starting in the beginning of 2017, these applications will be tested by the memory institutions of the PREFORMA Consortium. A successful competition of the second major phase is required for suppliers to be rewarded contracts for the Testing and validation phase. Therefore, the Consortium has formally evaluated the work of the suppliers at the end of the Prototyping phase.

2.1.2   Contracts and payments

Riksarkivet sent to each of the three selected suppliers from the first major phase (Design) a signed contract for the Prototyping phase (including the Re-design phase), which supplements the Framework Agreement. All the invited suppliers returned signed copies.

---

[1] This part of the Prototyping phase was extended with one month compared with original plans (see section 2.3.2)

A payment plan was set for the whole Prototyping phase: 30% in advance, 30% intermediate, and 40% as final payment for this phase. The first 30% was paid out during April and May 2015. Invoices for the intermediate payment were requested by Riksarkivet before Christmas 2015 and were paid before end of January 2016. Invoices for the final payment were requested in December 2016 and paid as soon as the supplier's final prototypes got the "green light" by the PREFORMA Evaluation Committee.

2.2 INTERACTION WITH THE SUPPLIERS

2.2.1 Intentions

The Prototyping phase started with a formal (virtual) kick-off meeting in April 2015. This was the first in a series of Prototyping phase meetings with the suppliers, organised by WP6. Issues discussed and outcomes from these meetings during the 1st part of the Prototyping phase are reported in deliverable D8.3 *First Prototype Report*.

The first monthly meeting in the 2nd part of the Prototyping phase took place on the 19th of April 2016. In this meeting, a work plan for the rest of the Prototyping phase was agreed with the suppliers (see Annex 1).

Another measure used by the PREFORMA Consortium, installed during the 1st part of the Prototyping phase, was the organisation of a review process with special focus on the software releases. This review process continued throughout the whole Prototyping phase. Besides providing the Consortium with formal releases of software (intermediate and final), the suppliers were required to submit reports describing these releases, the progress compared to previous releases, and several other issues of interest when measuring progress. All feedback on the formal releases and the accompanying reports was compiled into separate feedback reports, which were submitted to the suppliers to react on.

2.2.2 Issues discussed during meetings with suppliers

At the monthly on-line meetings, the main areas for discussion were:

- Report on progress;
- Check of the overall work plan/roadmap;
- Plan for testing and for releases;
- Plan for upcoming events where representatives for the PREFORMA project will attend and present the project and its results;
- Prepare the Testing and validation phase;
- Organise the Experience Workshop in Berlin;
- Evaluate the work of the Technical Working Group;
- Evaluate the results of the second major phase (Prototyping) and prepare for the third major phase (Testing and validation).

Several other topics also came up for discussion and decision during these meetings. An example of such an issue, that engaged both PREFORMA partners and suppliers, was a course on preservation formats at the Archiving Conference in 2017, which PREFORMA was invited to nominate teachers for. This annual conference on digital preservation and imaging, counts both European and American professionals among its audience. In 2017 the Archiving Conference

will be held in May in Riga, and the idea is to organise a general session on preservation formats and assessment of file formats, which could be followed by more focused sessions on PDF/A, on TI/A, and on Matroska/FFv1. PREFORMA accepted the invitation and planning of this activity took place over several of these monthly meetings.

2.3 PREFORMA CONSORTIUM MEETINGS

2.3.1 Project Management Team meetings

In addition to the supplier meetings, the Project Management Team (PMT) met on monthly basis for internal discussions during the 2nd part of the Prototyping phase. Several strategic topics were on the agenda, among them:

- The overall time plan for the 2nd part of the Prototyping phase;

- Evaluatation of the Prototyping phase;

- Preparing of the Test and validation phase;

- Tender release for the Testing and validation phase;

- Outcome of the PREFORMA delegation visits to suppliers in the beginning of 2016;

- Governance issues;

- Interoperability of the conformance checkers (internal and external);

- Policy requirements built into the conformance checkers;

- Progress of testing and test cases;

The outcomes of these discussions are addressed further in this deliverable.

2.3.2 PREFORMA Plenary meetings during 2016

The Plenary meeting in Stockholm, April 2016

On the 6th of April, a plenary meeting was held in Stockholm on the premises of the Swedish National Library. In addition to governance issues, and preparing for the upcoming project review, the following main decision points and action items were on the agenda:

1. Planing for the 2nd part of the Prototyping phase

It was decided to follow the same main procedures in the 2nd part of the Prototyping phase as in the 1st part, with monthly virtual meetings, monthly releases and two main checkpoints in July and October 2016.

2. Collecting test files

The dispatchers chosen for handling test files were to follow up the situation at memory institutions that (thus far) hadn't sent in test files.[2] The ambition was to cover as many subclasses and errors as possible (see section 2.4.2).

3. Reporting issues

Memory institutions were recommended to report issues concerning the three open source projects directly on the GitHub, and the suppliers were asked to help in summarising the activity that will take place there. Selected representatives from the Consortium (Bengt Neiss and Erwin Verbruggen) were appointed to support memory institutions in using GitHub (the GitHub guidance document is enclosed as Annex 2).

4. Experience Workshop

The PREFORMA Experience Workshop was to be organised jointly with EuropeanaSpace Final Conference in Berlin (22-24 November 2016) and hosted by the PREFORMA partner SPK. A first draft for a programme was set.

5. Methodology to move from the second major phase 2 (Prototyping) to the third major phase (Testing and validation)

A first discussion took place on how to move from the Prototyping phase to the Testing and validation phase. Action items and decisions:

- Firstly, the overall procedure was set: (1) Evaluation of the prototypes (straightforward, check whether they deliver everything on time); (2) invitation to tender among the suppliers that has successfully completed the Prototyping phase; (3) evaluation of the tenders by the Evaluation Committee, who will evaluate both the results of the Prototyping phase (Final release and accompanied report) and the bids for the Testing and validation phase.

- Secondly, evaluation criteria will be established in a new version of deliverable D8.1 (D8.1. R2) by the end of October 2016.

- Thirdly, a tender form will be prepared, based on the results of the Prototyping phase. A work plan for the Testing and validation phase will also prepared, including possible visits to procurers, training events and with a clear statement that the development of the prototypes should continue throughout the whole project (taking into account the feedback from the community and the results of the testing phase, and improving the tools accordingly to that).

- Fourthly, contracts for the third phase will be constructed.

6. Testing and validation

Task forces will be set up to define classes and the ground truth for each media type as described in deliverable D8.1. R2. The WP7 leader (UNIPD) will take the lead and circulate a

---

[2] For the function and role of dispatchers, see deliverable D8.3

request to memory institutions in the PREFORMA Consortium and to the suppliers to identify experts for the task forces. The same request was also to be sent to the Associate partners and to the members of the Advisory Board.

UNIPD will prepare a simple spreadsheet to help in collecting classes and defining the ground truth and circulate it to memory institutions together with instructions how to fill it in. Virtual meetings will be organised to explain to the memory institutions what they are expected to do and to monitor this activity.

6. Sustainability and exploitation

Sustainability and exploitation should be more in focus during 2016 and especially during 2017, and plans for that need to be identified. Possibilities for new proposals were discussed (PCP to develop new services and PPI to deploy services procured in PREFORMA).

The Plenary meeting in Berlin, November 2016

On the 24th of November 2016, a plenary meeting was held on the premises of Institut für Museumsforschung in Berlin. Besides governance issues like annual reporting and use of resources, recommendations from the last review, next deliverables etc., the meeting focused on the following items:

1. Changes in the Consortium during 2017

The Project Management Team (PMT) proposed to assign a new partner with focus on sustainability to highlight this area. This will not result in any changes in the daily management of the project or in the relationships between the partners, with the suppliers and towards the EC and external associate partners. The proposal was accepted, and a formal Amendment process will be initiated with the European Commission in the beginning of next year.

2. Results of the Prototyping phase.

A first draft of report from the Evaluation Committee was on hand at the plenary meeting. This report indicated that there were several issues still open and that at least two of the suppliers had a "last mile" to go before getting "green light" for final payment. A feedback report was to be submitted to each of the three suppliers by the end of November/beginning of December, highlighting the open issues to be fixed in the December release before it could be considered approved as the final release of the Prototyping phase.

3. Open Source and licensing related issues

It was discussed which part of the software that must be provided under MPL3+ and GPL2+ licenses (i.e. the Implementation Checker and the Policy Checker) and which part can include third party libraries which are released under a generally recognised open source license compatible with MPL3+ and GPL2+ (i.e. Shell, Reporter and Metadata fixer). This issue was raised already during the Design phase by the - at that time - 6 suppliers, and a clarifying letter was then sent to the suppliers by the PREFORMA Consortium. It was now decided, in Berlin, that this letter was still applicable, i.e. that the core part of the software conforms to the PREFORMA licensing requirements. Furthermore, it was decided that all three suppliers should submit executables, source code, and build environment containing only the core part.

A basic demand was also that it should be possible to install the software also without internet connection.

4. Roadmap for the evaluation and plans for the Testing and validation phase

The roadmap for the evaluation of the results of the Prototyping phase and of the new bids for Phase 3 was updated, based on the first draft report from the Evaluation Committee and the decisions taken upon that report. Since there are still issues to be fixed in the Conformance checkers, it was decided to delay the start of the major phase 3 Testing and validation (WP7) by one month. Consequently, the deadline for preparing test files was set to mid-February 2017. [3]

The overall plan set for the Testing and validation phase and the work of WP7 was that the suppliers will have to download the test files, run the tests and labeling the results according to what has been detailed in D8.1, and after that submit the results to the WP leader UNIPD for processing.

5. Sustainability and exploitation

It was discussed how to get memory institutions more involved. One success factor seems to be to take a step back and start discussing - and also train people in - what a conformance check really is and its importance in long-term digital preservation. Some hands-on activities (training, workshops, test-beds, tutorials, and webinars) had started during the Prototyping phase but need to be increased in 2017. Tutorials, learning material, a booklet and articles on technical magazines about the importance to produce good quality files are examples of ideas that were ventilated.

Ideas for how to get better impact on vendors were also on the table. To ask vendors to bundle PREFORMA with file producers (e.g. scanner vendors) and to give certificate to vendors when files they produce are conformant to standards, were two ideas coming out of this discussion.

6. Planning for a workshop in Padua

A first plan for a public workshop in Padua on the 7th of March was decided. Other planned activities in conjunction with the workshop were (1) a networking session for EC projects in the cultural heritage field, (2) a review meeting, (3) a plenary meeting, and (4) a local hands-on session.

2.4   PREPARING FOR THE TESTING AND VALIDATION PHASE

2.4.1   Initiating activities

Although the official evaluation of the Prototyping phase and the call for tenders for the next major phase, Testing and validation, were planned to take place at the end of the 2nd part of Prototyping phase, the project started to plan for these activities already during the 1st part of this phase. Work needed to start early, both on the testing methodology and on the collection of test files, to ensure that the final versions of the Conformance checkers would meet all the

---

[3] The use of test files is described in depth in deliverable D8.3

requirements that have been identified in the PREFORMA Challenge Brief. Two extra check points were also implemented for the Prototyping phase:

- The PREFORMA Evaluation Committee reviewed the results of the 1st part (final releases and reports) and compared them to the outcomes of the Design phase and the evaluation made at that time;

- A PREFORMA Delegation was set up with the task to visit the suppliers in the beginning of 2016 to get a better understanding on their working conditions.

The outcome of these extra check points has been reported in deliverable D8.4 *Design – Final report*.

2.4.2   Activities to prepare for the Testing and validation phase

During the 1st part of the Prototyping phase, other activities were launched to prepare for the succeeding Testing and validation phase. A call for participation was launched, aiming to involve memory institutions and other content providers outside the PREFORMA Consortium in sending test files and participate in testing. So far (December 2016) 21 institutions and organisations have expressed their interest to contribute. A mailing list[4] and a dedicated webpage on the PREFORMA website[5] for external stakeholders were created.

Furthermore, activities started to identify all the information that needs to be attached to the test files for a proper analysis to be made. These activities included:

- identification of all the possible use cases or classes to be used during the evaluation phase;

- establishment of the so-called ground truth.

As agreed at the plenary meeting in Stockholm, the leader of WP7 started to work on the organisation of the Testing and evaluation phase in April 2016. The first step consisted of the identification of "classes" for each media type targeted by PREFORMA, i.e. of the possible problems/issues to be checked by the conformance checkers. Each class represents/models an issue and each document would have to be labeled as belonging to one or more classes according to its characteristics, i.e. to the issues it suffers. The objective was to identify a list of all possible issues that might affect a file, either because these issues do not respect a rule in the standard specification or because they do not comply with a policy established by a memory institution.

A first virtual meeting between the PREFORMA partners took place in the second week of June to kick-off this activity, agree on the composition of the task forces (one per media type) and plan the next steps and deadlines. The outcome of this activity was continuously reported and discussed on the monthly supplier meetings.

---

[4] *preforma-testing@promoter.it*

[5] *http://www.preforma-project.eu/external-partners.html*

Some basic materials were already at hand when the task forces started to work:

- A document drafted by Riksarkivet as a starting point to identify the possible characteristics/features to be checked in a PDF file;

- veraPDF integration tests that publish results online: http://tests.verapdf.org/. It was not quite a framework at that time, but new test corpora were easy to add;

- veraPDF validation profiles: https://github.com/veraPDF/veraPDF-validation-profiles/wiki.

- EasyInnova XML files that specify the implementation checker rules: https://github.com/EasyinnovaSL/DPFManager/tree/develop/src/main/resources/implementationcheckers.

The task forces produced a first list of classes to be used as an input for the revised release of deliverable D8.1 *Competitive Evaluation Strategy* (October 2016).

2.4.3   Test cases

Throughout the entire Prototyping phase, members of the PREFORMA Consortium and other interested end users have been invited to make usability tests for themselves. A new element during the 2nd part was to initiate organised test cases at memory institutions and service providers in digitisation and archiving, where conformance checking of files is considered as an important step during ingest workflows and ingest reporting. These test cases, or "reality checks", have contributed with important knowledge to the overall evaluation of software produced by the suppliers, and will do so to the process of testing in the upcoming Testing and validation phase. By the end of the Prototyping phase, in total six organised test cases were at hand:

- EuropeanaSpace, focusing on images and integration in their technical Space; PREFORMA PMT was connected to EuropeanSpace for following up of the outcome;

- Media Converting Centre (MKC) in Sweden, focusing on images and large scale production; this centre is a department at the partner Riksarkivet;

- Vlaams Instituut voor Archivering (VIAA), focusing on AV; the outcome was followed up by the partner PACKED and the supplier MediaArea;

- Standards Forum NL (/),[6] focusing on PDF/A; the outcome was followed up by the PMT, the partner PACKED, and the supplier veraPDF;

- MediaArea in collaboration with Archivematica to embed the AV Conformance checker;[7]

- Catalan Open Government Consortium (AOC Consortium),[8] who want to embed veraPDF conformance checker in their product iArxiu; the outcome was followed up by partner AJGI;

---

[6] https://www.forumstandaardisatie.nl

[7] More information is available at https://wiki.archivematica.org/Requirements/MediaConch_integration:

[8] http://web.aoc.cat/

Discussions were also had with the EU financed project E-ARK, where the technical partners showed interest in testing the PREFORMA Conformance checkers. However, the time plan for E-ARK was too tight to allow any tests.

At the Experience workshop in Berlin, some of theses test cases (Vlaams Instituut voor Archivering, Riksarkivet Media Converting Centre) were on the agenda to present initial results and give feedback on how the software could be improved.
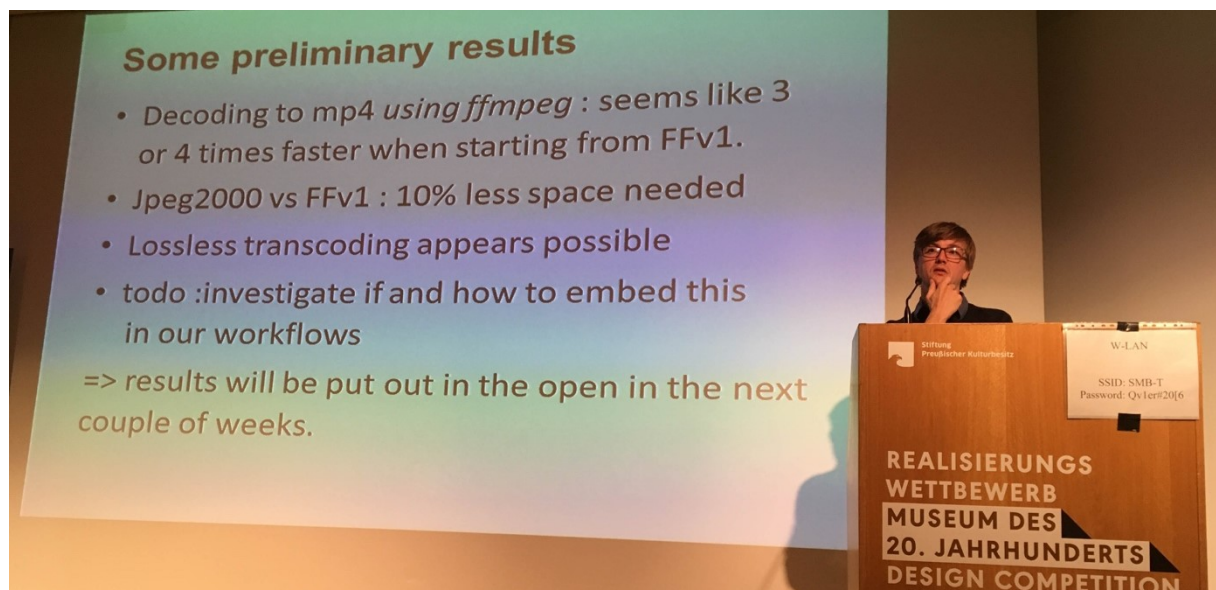
VIAA.s Matthias Priem during the Experience Workshop in Berlin

3   SPECIFIC ISSUES

This chapter summarises some specific issues covered by the objectives of the project and raised during the Prototyping phase.

3.1   SOFTWARE DEMONSTRATIONS

According to the DoW, two demonstrations will be organised by the suppliers and reported in deliverable D8.3 *First Prototype Report* and in the deliverable D8.4 *Design – Final Report*. During the 1st part of the Prototyping phase, demonstrations were not carried out by the suppliers in any pronounced way. On the other hand, during the 2nd part of this phase, the suppliers made several demonstrations at occasions organised by the PREFORMA Consortium or by others. Examples of such demonstrations were two workshops delivered by Dave Rice and Ashley Blewer during late July 2016 at Tate Britain in London on the use of Mediainfo, MediaConch and FFmpeg in the preservation of digital video.[9] At the symposium, called "*No Time to Wait. Standardizing FFV1 & Mastroska for Preservation*", on the 18th – 20th of July 2016 in Berlin, the MediaArea team made a set of presentations that focused on testing FFV1. [10]

The two main public demonstrations organised by the PREFORMA Consortium was at the Open Source Work Shop in Stockholm in April 2016, when the first stable versions of the prototypes were presented, and at the Experience Work shop in Berlin in December 2016. A third occasion, also organised by PREFORMA, was the Image and Research Conference in Girona in November 2016, which included a demonstration of the PDF Manager.[11]

The PREFORMA software has also been demonstrated by project partners at conferences and work shops, both during the 1st and the 2nd parts of the Prototyping phase.[12]

New channels for software demonstrations were introduced during the 2nd part of the Prototyping phase, the one being Webinars.

With the help of the supplier veraPDF, a series of webinars were organised. The first one on the 14th of June 2016 and focused on PCP and PREFORMA in general. Three more webinars were then organised in September 2016 on the use of the conformance checkers, one for each supplier. All three started with an introduction by the PREFORMA Consortium - to explain what

---

[9] See the PREFORMA blog: http://www.digitalmeetsculture.net/article/mediainfo-mediaconch-and-ffmpeg-in-the-preservation-of-digital-video/

[10] See the PREFORMA blog: http://www.digitalmeetsculture.net/article/no-time-to-wait-standardizing-ffv1-mastroska-for-preservation/

[11] See the PREFORMA blog: http://www.digitalmeetsculture.net/article/image-and-research-conference-2016-report/

[12] A comprehensive resume of PREFORMA disseminationactivities can be found in deliverables D4.5 *Dissemination Report Year 2* and D4.6 *Dissemination Report Year* 3.

a conformance checker is - and was followed by a software demonstration presented by rotation of the suppliers.[13]

3.2 TRAINING SESSIONS

During the Plenary meeting in Berlin on the 24th of November, one of the items discussed was how to get memory institutions more involved in the work of PREFORMA. One success factor seems to be to take a step back and start discussing what a conformance check really is and its importance in long-term digital preservation. A first indication of the need for this kind of "re-thinking" came as feedback from participants in a work shop organised by PREFORMA at the Ipres Conference in October 2016 in Bern. At a workshop in November 2016, organised by PREFORMA at the EVA/MINERVA Conference in Jerusalem, this new concept was then tested with very good results. As one of its participants said: "It is hard to sell a car to someone who does not have a driving license".

Some hands-on activities (workshops, test cases, and webinars) had started during the Prototyping phase, but the PMT decided to increase them in 2017 by organising training sessions (recommended also by the reviewers). These training sessions will focus on specific media types and be very practical and hands-on. The following sessions are planned:

- For AV – Partners NISV and PACKED will organise a session in May 2017, probably in Amsterdam in connection to the Reel Thing XL Symposium on film restoration;

- For PDF – Partners Riksarkivet and EVK will organise a session in May or June 2017 in Stockholm or in Tallinn;

- For TIFF – Partners AJGI and KIK-IRPA will also organise a session in May or June 2017, potentially in Gerona.

In the call for tender for the Testing and validation phase, suppliers have been asked to take part in at least one of these training sessions and cover their costs on their own budget.

Hands-on training will also be provided in connection to the planned public work shop in Padua in March and the Final Conference in October 2017.

3.3 DEPLOYMENT OF THE CONFORMANCE CHECKERS

It is important for the PREFORMA Consortium and for the suppliers that the PREFORMA software will be distributed through several channels. The Consortium has mandated the suppliers to use as many distribution methods as they want without requiring a particular one. MediaArea have been using "Homebrew" and "Linuxbrew" as packages but also a few Linux repositories; EasyInnova are using the Maven repository.

During the 2nd part of the Prototyping phase it became more and more urgent to reach an agreement for the suppliers to use the same distribution channels when promoting the conformance checkers. Following a suggestion from the Project Officer, AppHub (www.apphub.eu.com) was contacted to make clear whether it would be feasible to deploy the

---

[13] Information about PREFORMA events are found at http://www.preforma-project.eu/events.html

PREFORMA software in the AppHub Store. Other possible channels and package managers were also considered.

In the end, the suppliers were not convinced on investigating other possible distribution channel at this point. They preferred to keep the PREFORMA Open Source Portal and their own repositories as main distribution channels - at least during the Prototyping phase. Efforts were instead allocated to try to integrate the software prototypes in other systems (e.g. Archivematica) to test the functionality and improve the tools. The choice of distribution channels will be an issue to look further into during the last year of the project when the prototypes have become more mature.

3.4 INTEROPERABILITY OF DIFFERENT CONFORMANCE CHECKERS

At the kick-off meeting for the Prototyping phase in April 2015, the PREFORMA Consortium addressed the issue of developing a common API for the harmonisation and interoperability of different conformance checkers. Interoperability can be:

- **external**, allowing an external system to call and control the conformance checker through a common gateway (the Conformance checker shell); or

- **internal**, enabling the integration of modules and extending the conformance checker with the features of another conformance checker (e.g. allowing the PDF/A Conformance checker to use the TIFF Conformance checker when analysing TIFF images inside a PDF document).

During the 1st part of the Prototyping phase, the PREFORMA Consortium and the suppliers agreed that the **internal level** is in the first place an issue for the suppliers to handle. A specific working group on interoperability was organised and starting its activity based on an internal use case document for a common shell "*Use Case for Interoperability*" prepared by the Consortium. A timeline was set, indicating how and in which order it would be sensible to tackle the issue of interoperability. The aim was that the suppliers should do the developing work on interoperability in the beginning of the Re-design phase, when the PREFORMA Consortium was evaluating the results of the 1st part of the Prototyping phase.

The suppliers continued, as planned, to develop the interoperability layer and to integrate it in their software. All the conformance checkers now support the integration and configuration of external checkers and allow a user to check multiple file formats at the same time. The software checks the file format, invokes the appropriate checker to perform the validation, and reports the result back to the user in a standard homogeneous format.

In the 2nd part of the Prototyping phase new Technical Working Group took over issues concerning internal interoperability.

Interoperability on an **external level** has during the 2nd part of this phase been integrated in some of the test cases organised by PREFORMA, where embedding PREFORMA software into a specific environment could be one part. The Integration of the MediaCronch in Archivematica is one of these test cases (see section 2.4.3).

3.5 POLICY REQUIREMENTS

One of the PREFORMA challenges is to develop an open source conformance checker that not only checks if a file complies with standard specifications but also checks if a file complies with the acceptance criteria or policies of the memory institution.

During the 1st part of the Prototyping phase, the suppliers underlined the importance to have complex policy requirements in place, if possible by the time of the Re-design phase. Requests for valid policy requirements were sent out, but the feed back was minimal. The suppliers were, therefore, advised to use the SCAPE Catalogue of Policy Elements. [14]

The issue of policy requirements has been handled also during the 2nd part of the Prototyping phase. The main result in terms of policy checking is the creation of a public registry of policies where users can publish their policies and make them available for other users. This registry has been created and promoted by the supplier MediaArea. It includes already several policies collected by some memory institutions, and not only policies related to AV but also to the other file formats. During the Experience Workshop in Berlin, the PREFORMA Consortium recommended the other two suppliers to follow the same approach and contribute to this activity in order to improve the registry. Furthermore, the technical side of implementing policy checking became an issue of importance in the evaluation of the software provided by one of the suppliers (see section 4.2.2).

---

[14] SCAPE was an EU-funded project that developed scalable services for preservation planning and preservation actions on an open source platform. See http://wiki.opf-labs.org/display/SP/Catalogue+of+Preservation+Policy+Elements

4   RESULTS OF THE PROTOTYPING PHASE

In deliverable D8.2 *First Prototype Report*, the approach taken was to undertake a situation assessment of the open source projects in PREFORMA. Six progress points were also used to assess the suppliers results so far in the following areas: Capabilities for software; Evolving functionalities; Usability; Testing for quality assurance and accuracy; Achieving reference Implementation; Awareness of what is still missing.

The conclusion was that goals had been achieved in the 1st part of the Prototyping phase. The suppliers had delivered prototypes and provided information on the following areas:

(1) their releases;

(2) how software testing had been carried out;

(3) dissemination and community building efforts;

(4) the open source approach that had been used;

(5) their standardisation efforts.

This deliverable (D8.5) focuses on how the suppliers in their final prototypes have provided the required functionality, established a process for providing feedback to standardisation organisations and adhered to utilising best practices from open source development. Approved final prototypes could then serve as a basis for the final Testing and validation phase.

This chapter is divided into three sections:

- **4.1 Progress made during the 2nd part of the Prototyping phase**, based on the reporting by the suppliers;

- **4.2 Feedback from the PREFORMA Consortium**, based on the feedback reports from the PREFORMA Consortium and its Evaluation Committee, which includes both the software releases with the associated reports provided by the suppliers and the End of Phase report;

- **4.3 Final results of the Prototyping phase**, which describes the evaluation process and the decisions and choices made by the PREFORMA Consortium during this process.

A summary of the main results achieved by the suppliers at the end of the Prototyping phase is also available in deliverable D3.7.1 "*Initial Version of Sustainability and Exploitation Plan*". The focus in that document relates to the key outcomes on future use and exploitation of the PREFORMA results.

4.1 PROGRESS MADE DURING THE 2ND PART OF THE PROTOTYPING PHASE

In the 2nd part of the Prototyping phase, the suppliers were expected - as in the 1st part - to provide two kinds of software releases:[15]

- Monthly stable releases;

- Two formal releases based on the functional and technical specifications from the previous Design phase: Firstly the *Intermediate release* in July 2016 and secondly *the Final release* at the end of October 2016. This release is the release on which the evaluation of the Prototyping phase will be based.

The two formal releases each had to contain two parts:

- Firstly, a functionally stable release;

- Secondly, a report which described the following:

  1. Increased detail on the respective release;

  2. The timeline along with the current status (on time, delayed, ahead);

  3. How the suppliers had provided the required functionality (to date);

  4. Omissions in relation to the original specifications and the plan to address these omissions;

  5. Basic information to be used by PREFORMA in deliverables to be submitted to the EC on work undertaken during the Prototyping phase.

The templates for the releases are enclosed to this deliverable as annexes 3 and 4.

The purpose of the software releases in the 2nd part of the Prototyping Phase was to verify that the PREFORMA requirements were fulfilled. These requirements are detailed in deliverable D4.3 *Functions of the Open Source Portal*, while additional, basic requirements on software releases are provided in the Call documents (Invitation to Tender, Challenge Brief, and the Framework Agreement) and in the deliverables D2.1 *Overall Roadmap* and D2.2 *Tender Specifications*.

In addition to the two formal releases comprising self-reporting parts, each supplier had to provide a formal End of phase report. This report provides basic data for the PREFORMA Consortium to decide if the results of the Prototyping phase were of the required standard to invite the suppliers to take part in the call for tenders for the Testing and validation phase. The objectives of this End of phase report were:

- to provide an overview of the work done in order to measure the results against the objectives and the requirements included in the PREFORMA Challenge Brief;

- to provide a comprehensive report to be shared with stakeholders to facilitate further commercialisation of the product.

---

[15] In open source projects, continuous releases are expected (like daily/nightly builds); the PREFORMA projects requirement is, therefore, in this sense a minimum.

In the following seven sub-sections, focus is on the progress claimed by the suppliers in seven different areas of activities, outlined in the templates of the reports that the suppliers submitted as a part of the two formal releases. These areas of activities are: (1) Software releases and progress, (2) Software testing, (3) Dissemination and community building, (4) Open source approach, (5) Standardisation efforts, (6) Impact assessment, sustainability use and exploitation, and (7) Gap analysis and next step. Each sub-section summarises how the suppliers addressed the tasks specified by the PREFORMA Consortium.

In a final (eight) sub-section, the main points from the suppliers End of phase reports are summarised.

4.1.1 Software releases and progress compared to the last intermediate release (2015)

The task here was to firstly provide the PREFORMA Consortium with an overview of the progress made since the last intermediate release, and secondly to provide the functionalities that were available at the time of this report. In the Final report, the suppliers also had to highlight:

- progress achieved since July 2016 release (intermediate release of the 2nd prototyping phase);

- how they had addressed earlier comments and recommendations received from the PREFORMA consortium;

- their plans for further progress.

A. veraPDF

In the **Intermediate report,** veraPDF highlighted several features that had been put in place or improved in their version 0.18, released in June 2016, compared with their version 0.6 in 2015. As veraPDF understood it, the PREFORMA Consortium's main concern regarding the first intermediate release had been the lack of a command line interface. This feature was released in October 15th and had undergone development since then. They reported that it was now complete. Furthermore, verPDF reported that they over the final 6 months of the Prototyping phase development they planned to:

- Complete the development of their "greenfield" PDF parser implementation, which will replace the current PDF Box dependency used for development;

- Carry out functional, performance and reliability testing of the "greenfield" PDFparser;

- Remove other Apache licensed code in line with PREFORMA's licensing requirement, e.g. their logging is Apache based/PDF Box compatible;

- Develop plug-ins that wrap open source validation/reporting tools for formats used within PDF/A, e.g. fonts, JPEG 2000, ICC colour profiles;

- Perform real world testing against third party/institutional datasets;

- Validate the test suite files via comparison with commercial products and discussions within the PDF Validation TWG;

- Produce the PREFORMA shell incorporating all 3 Conformance checkers;

- Create platform specific installation packages;

- Finish the development of the veraPDF REST services and web interface,

In the **Final report,** veraPDF reported that since version 0.18 in June 2016, most of their development efforts had been spent on replacing the PDF Box based PDF parser with the "greenfeeld" one and on validation model implementations. A comparison of the list of features in place in both releases was required as part of this report. The plan veraPDF presented for the last two months of development was almost identical to the plan they presented four months earlier in their Intermediate report. Only one new planned new step was presented to develop "greenfield" metadata fixing and feature reporting.

B. EasyInnova

In the **Intermediate report**, EasyInnova highlighted in total eleven progress points since their last intermediate report in June 2015. The main ones were:

- The project architecture had been restructured to follow a modular event-driven approach. Java Spring and JacpFX frameworks were now used for building the Graphical User Interface (GUI), and EasyInnova had adopted the model-view-controller pattern for implementing the individual components. This new modular architecture allows DPF Manager to be able to run checks in parallel, which was a new feature. The overall performance of the DPF manager was improved by 500%.

- A Maven artifact had been created to allow external projects to easily integrate the DPF Manager into their solutions.

- The Implementation checker had been rewritten, and could now, instead of being hardcoded in the source, read from an XML file that contains all the rules to validate the ISO standard.

- The configuration file of the DPF Manager, which contains the ISOs to be checked, the policy rules, metadata fixes and report formats, had also been rewritten in XML format to facilitate maintenance and to prepare it for interoperability purposes between Conformance checkers.

- Client-server architecture had been implemented to allow DPF Manager to be executed in server mode in one computer and client-mode in another computer.

- The console window that appeared behind the GUI had been eliminated, as suggested by the PREFORMA reviewers, and the log messages now appear in a console widget integrated into the GUI.

Planned next steps included an improvement on the XML report. The ambition was to enhance the DPF Manager report to include not only the validation results of a TIFF file, but also to show information on its structure and metadata in a language that archivists could understand. Therefore, EasyInnova had decided to use the METS format to ensure that it would be possible to integrate the DPF Manager with memory Institutions that have applied the Open Archival Information System (OAIS) model in their data object lifecycle. In the METS report they also include PREMIS, NISO and Dublin Core data dictionaries.

Finally, EasyInnova had plans for improving the Policy checker configuration in order to allow users to generate custom rules by making it possible to edit single rules in Schematron format directly, or even edit the whole Schematron policy file.

In the **Final report**, EasyInnova stated that the final prototype release of the DPF Manager includes all the features specified in the functional requirements and technical specifications described in the EasyInnova documentation and tender proposal for the Design phase. Moreover, they had also included some additional functionalities and improvements, most of them proposed during the Re-design phase.

EasyInnova claimed that the DPF manager is using the latest developer technologies. To ensure portability they use java (latest versions) portability. EasyInnova provides the installers and executables for Windows, Mac OSX and Linux (Ubuntu, Fedora, Devian and Suse) and is according to EasyInnova, very easy to build the application for other systems using Maven.

EasyInnova also reported that the DPF manager all-in-one application now includes all the interfaces to be deployed and integrated with different infrastructures in one executable. DPF manager can be initiated via a 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.

The application can communicate with other applications to provide interoperability with other Conformance checkers. The Shell of the DPF Manager distributes the files to the appropriate Conformance checkers.

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

The DPF manager can also be used as a framework, ready to be integrated with other applications or frameworks via API. To facilitate the integration, DPF manager has been included in the Maven package repository.

EasyInnova gave in their report a comprehensive description of the functionalities of the different parts of the DPF Manager´s TIFF Conformance checker (the Implementation checker, the Policy checker, the Metadata fixer and a Report module).

C. MediaArea

MediaArea highlighted in both their formal reports progress on general level as well as several progress points. In the **Intermediate report,** they emphasised the following (selected) parts**:**

Progress on general level

Firstly, MediaArea had continued to develop the community via the IETF (Internet Engineering Task Force) CELLAR (Codec Encoding for LossLess Archiving and Realtime transmission) working group, where work had been done to support the standardisation of Matroska and FFV1. The working group's mailing list was continuously very active with members of the core Matroska community and the FFV1 community working on further developing the latest specification for standardisation along with the MediaArea team. The listserv of the CELLAR working group included at the time for the intermediate report 89 members, including 36 active participants.

Secondly, in addition to making the raw source code material available for building MediaConch from scratch and cross-platform automated builds, MediaArea was now able to provide daily builds for users to allow them to download a ready-to-install copy of the software in all three shells, regardless of their preferred operating system platform. To support ease of installation and development, MediaConch is now available through several package systems: Homebrew,

Linuxbrew, and Debian. Forthcoming development includes integration into Fedora.[16] Integration of MediaConch into Artefactual's Archivematica had also begun.[17]

Thirdly, MediaConch had continued to establish presence in a variety of communities that have interest in the developed software. This had been done by presenting or being represented at several conferences.

Progress points since the October, 2015 release:

- The project had, according to MediaArea, made significant progress towards meeting earlier objectives and more recent goals arising from their user's experiences;

- Improved build environment with detailed instructions on how to download builds for Linux, Windows, and Mac OSX;

- Updated website design so users can find installation links, directions, and daily builds more easily;

- Made GH Issue Tracker links directly from Help/FAQ page in the GUI;

- Responded to feedback from user interviews and have opened tickets for each unresolved issue mentioned in the initial feedback along with others in GitHub.

In their **Final report,** MediaArea emphasised the following points of progress:

Progress on general level

The list server of the CELLAR working group had at the time of the final report expanded to 114 members, including 48 active participants.

The October release of MediaConch included the first Metadata fixer features, such as the ability to fix incorrect sizes within Matroska Elements.

MediaConchOnline, the web interface, was now available without registration. MediaConch converts the reporting format of DPF Manager and VeraPDF, in order to support consistency in reporting and display. MediaConchOnline had also been configured with versions of DPF Manager and VeraPDF. As part of a collaboration with VIAA, MediaArea has started to use MediaConch's plug-in abilities to integrate an FFmpeg plug-in. This will allow the use of MediaConch for converting preservation files to Matroska/FFV1 files while running the implementation checker.

Progress Highlights Since the July, 2016 Release

Since the July, 2016 release, the project had, according to MediaArea, made significant progress towards both earlier objectives and more recent goals arising from MediaArea´s research on user experiences. In the upcoming months, MediaArea plan to increase the count of system policies (default policies provided and supported by MediaArea), improve the FFV1 specifications, add more features to the fixer module, add more features to the public policies

---

[16] Fedora is a linux distribution developed by the US based company Red Hat (see http://fedoraproject.org/wiki/Releases#History)

[17] For information about the company Artefacual Systems, see https://www.artefactual.com/

feature, stabilise the software, and make updates based on new requests from the users on the project issue tracker.

4.1.2 Software testing

The task here was to provide the PREFORMA Consortium with a detailed description of the datasets that have been used to test the releases (own, memory institutions, external, etc.), and the respective purpose of testing.

A. veraPDF

veraPDF reported in the **Intermediate report,** that releases and all integration branch merges had been tested against their own synthetic test corpus for all PDF/A flavours, but also against the Isartor PDF/A1b test suite and the BFO PDF/A2 test suite.

The test results for each build are published on http://tests.verapdf.org/ . These are all purpose produced data sets designed to test PDF/Avalidation functionality. The veraPDF test corpus comprised at the time of over 1,500 PDF files, created by the consortium as a comprehensive PDF/A validation suite.

veraPDF also reported that they performed institutional testing on real world data sets that can't be provided because of IPR issues. Testing with heritage sector organisations had focused on reliability, performance and usability rather than testing the functionality of the validator against the PDF/A specifications. One of these heritage organisations was the British Library, their large-scale testing meant that veraPDF could fix serious performance issues with large files containing high quality images, and also problems with text layer fonts.

In the **Final report**, veraPDF underlined that both the PDF Box and "greenfield" veraPDF releases and all integration branch merges are tested against the test suites and synthetic test corpus mentioned in the intermediate report.

veraPDF had targeted organisations that archive PDF/A files in accordance with institutional practice and those that archive PDF/A and other PDF files at scale. Two organisations were approached and they received feedback from seven institutions, ranging from national archives to university libraries:

- British Library
- Koninklijke Bibliotheek (National Library of the Netherlands)
- Parliamentary Archives
- Leibniz-Informationszentrum Wirtschaft (German National Library of Economics)
- University of Yale
- University of Sheffield
- University of Edinburgh

Several volunteers had contributed issues to the veraPDF GitHub Issue tracker, in particular regarding stability and performance improvements. This has, according to veraPDF, led directly to the optimisations in memory usage and better handling of memory exceptions.

B. EasyInnova

EasyInnova reported in the **Intermediate report,** that they had used several sets of images for different testing purposes. For the validation of the different ISO standards (Baseline, TIFF/EP and TIFF/IT) they had formed sets of valid and invalid images for each standard, some manually created to simulate possible errors in the TIFF structure and data.

EasyInnova had also created new validation rules that, although not explicitly stated in the TIFF specification, would produce incorrect TIFF information, like tags out of place, illogical values, or invalid field definitions. In these cases, warning rules were created instead of errors to inform the user that the data in the TIFF file is not coherent.

Unit tests had been created to run checks in all these files and compare the validation result with the expected output. These tests are run every time a build is made through the Maven test plug-in.

EasyInnova had used the Travis continuous integration tool to run all these unit tests every time a modification was committed to the source code repository. This is to assure that new features, incrementally added to the project, will not break anything that was working previously. The Travis tool had been configured to run the tests so it checks the DPF Manager in Command Line Interface (CLI) mode, GUI mode and client-server mode. When the tests are successful in all three modes they are then considered correct.

The DPF Manager had been downloaded and used by some early adopters from different countries, who had agreed to send feedback from their tests. Moreover, EasyInnova reported that the integration of the DPF Manager in the EuropeanaSpace project had been successful. DPF Manager had been also extensively tested by the PREFORMA partner Packed for the Tapies Foundation in Barcelona. [18]

Regarding plans for the future, EasyInnova reported that they are working on the creation of a more extensive test set of TIFF classes, including some new handcrafted files, and creating a new website for TIFF test cases.

In the **Final report**, EasyInnova reported that during the project they had been applied different tests to their releases to ensure a proper operation of the DPF Manager. First, they had applied unitary tests for testing every relevant function in the code and to ensure the same outcome in every code contribution. They also applied GUI tests to ensure that all functionalities in the application are still working after any source code contribution. Travis, their continuous integration tool, had been properly configured for testing all the modules including interface, implementation checker, policy checker, metadata fixer, and reporter. Although internally EasyInnova still use unitary tests, all the unitary tests had been removed from the source code in the final release while waiting for a clarification from the PREFORMA Consortium concerning licenses.

For the TIFF reader/writer and implementation checker, EasyInnova had generated an image dataset to test the application. However, in order to avoid a situation where the testing process could be self-referenced and influenced by their knowledge on the implementation, EasyInnova

---

[18] For information about the Tapias Foundation, see http://www.fundaciotapies.org/site/?rubrique64

have internally used another dataset to test the application, but this dataset is not public for license reasons.

Finally, EasyInnova´s most intensive test came 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 on which kind of TIFF are used at preserving memory institutions, 3.9 Million existing TIFF files in Swiss archives were checked.

C. MediaArea

MediaArea reported in the **Intermediate report,** that their team had helped to move the official Matroska collection of test files from a static zipfile hosted on sourceforge into a new GitHub repository at https://github.com/MatroskaOrg/matroskatestfiles. In addition, MediaArea reported that they could reprocess a collection of a hundred thousand original Matroska files at archive.org under new more verbose parsing options of MediaInfoLib. This will generate more comprehensive test results on the developing implementation checker.

The original work to analyse the massive corpus of the Internet Archive had also commenced since the original analysis parsed files were too selective to cover many types of Matroska implementation checks. The new analysis was reported to be slower but more comprehensive and will allow MediaArea and the CELLAR Working Group to identify implementation errors and relate them to specific samples, specification language, and the creation software.

The development work in CELLAR had also clarified many obligations, constraints, and expectation for Matroska validity. Therefore, the need for comprehensive test sets had expanded. Testing of the policy checker had also expanded because of collaborations with Artefactual and the Tate Museums.[19] Testing and collaboration with these organisations and with other users contributed to identify the improvements needed for the policy checker.

In the **Final report** MediaArea mentioned that the re-processing of a hundred thousand original Matroska files at archive.org, reported in the intermediate report, was progressing.

When testing the policy checker in collaboration with Artefactual Archivematica and the Tate Museum, MediaArea found that the original policy logic was too simple to support many in-practice policies. A blog post about this is found at https://mediaarea.net/MediaConch/2016/10/04/policy-refactor/

4.1.3 Dissemination and community building

The task here was to provide the PREFORMA Consortium with a list of dissemination activities that had been undertaken to promote the suppliers open source project (web pages, blogs,

---

[19] Tate is a family of four art galleries in London, Liverpool and Cornwall (see www.**tate**.org.uk/)

1.
2.

PREFORMA - Future Memory Standards
PREservation FORMAts for culture information/e-archives
EC Grant agreement no: 619568

newsletters, press releases, papers, presentations, etc.). Any potential long-term collaborations/partnerships entered, should also be reported by listing the organisation/s and the role they have played in the project. Furthermore, the suppliers had to report how they had progressed in setting up an open source community around the developed tools.

A. veraPDF

In the **Intermediate report** veraPDF reported on:

1. Web presence (continuously updates of the veraPDF website);

2. Recent events/conferences (5);

3. Webinars (1 occasion);

4. Press releases (four);

5. Mailing lists (20) and LinkedIn interest groups sent to;

6. Regular updates posted for the industry community, the public, and for PDF Association members only;

7. Twitter account **(**71 followers);

8**.** veraPDF news (144 subscribers);

9**.** The PDF Association: PDF Validation TWG (55 subscribers);

In the **Final report** veraPDF updated the information in the intermediate report as follows:

1. Partnerships: KEEPS and Archivematica (as previously reported);

2. A new mailing list had been set up for users;

4. Presentations/demonstrations had been preformed at 5 events/conferences since the intermediate report;

5. One article had been published in *Digital Library Perspectives Journal: Special issue on digital preservation tools and partnerships;*

6. Three webinars had been organised, one per supplier;

7**.** Seven press releases had been sent out since the intermediate report;

8. The twitter account had now 97 followers;

9. veraPDF news now had 160 subscribers;

10. The PDF Association: PDF Validation TWG now had 57 subscribers.

B. EasyInnova

In the **Intermediate report** EasyInnova explained that the TI/A community had been built up around three online channels (and many offline communications):

- the TI/A website www.ti-a.org, which at the time of publication of the Intermediate report had received 5.500 visits with 3.900 unique visitors in 15 months;

- the TI/A twitter account twitter.com/TI_A_Standard, which had almost 350 followers; EasyInnova had published around 770 tweets related to the TI/A format and digital image preservation in general;

- The TI/A Intranet for the involved experts intranet.ti-a.org, which at the time of publication had 72 experts from 16 different countries registered.

To raise awareness in the scientific community, EasyInnova had published a white paper about the TI/A format, and presented the TI/A and DPF Manager initiative at several scientific conferences. EasyInnova could also rely on a valuable collaboration with the Swiss Coordination Centre for the Long-Term Preservation of Electronic Documents (KOST-CECO - http://www.kost-ceco.ch/) and some of the bigger Swiss Archives.

In the **Final report,** EasyInnova declared that they had been moving forward in their strategy for creating archivist awareness around TI/A. A first draft of the specification for TI/A was published on September 30th. This had started a discussion in the community of experts, which had helped EasyInnova in producing a second draft hat has been sent to the ISO Working Group to be discussed during its meeting in Sydney at the end of November.

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

- The TI/A website has received 6.500 visits with more than 4.500 unique visitors in 18 months, with an increment of approximately 25% after the publication of the 1st draft;

- The twitter account has 410 followers and 850 tweets related to the TI/A format and digital image preservation had been published. The growth of followers had been around 20% during the last month;

- In the TI/A Intranet, 80 experts from 17 different countries were registered at the time for the report;

- In the last 3 months, TI/ Initiative news has been published in external blogs like Digital MeetsCulture, 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, EasyInnova had presented the TI/A and DPF Manager initiatives at international conferences in Europe, Asia, and US.

C. MediaArea

In the **Intermediate report**, Media Area has listed the following dissemination activities:

1. Publishing newsletters;

2. Organisation of the "No Time To Wait!" symposium, held during IETF96 with over 60 attendees;

3. Website set up;

4: Documents Repository available;

5. Twitter Storify (2 reports);

6. Presentations/workshops at conferences/meetings (4);

7. A mailing list had been set up.

In the **Final report**, MediaArea updated the list of activities as follows:

1. MediaArea participated in a workshop on October 5th;

2. CELLAR continues to grow in number. On October 30th, MediaConch team members held an informal meet up with Nick Krabbenhoeft at New York Public Library to introduce interested CELLAR members to this work;

3. Artefactual has continued to integrate MediaConch into Archivematica, as summarised in the intermediate report. In October, MediaArea hosted a webinar together with Artefactual which had its focus on the integration of MediaConch. Artefactual also gave a poster presentation at the Association of Moving Image Archivists conference in early November regarding Archivematica and MediaConch. MediaArea staff will also teach the usage of MediaConch as part of a "Digipres 101" workshop;

4. MediaInfo and FIMS had an EBU booth at IBC (EBU booth), including discussion on MediaConch features, on September 10th;

5. MediaArea have plans to expand software related to community-building to include ability for users to share their policies and a wiki for embellishing failed implementation reporting;

6. MediaConch is now part of the latest official Ubuntu version (MediaConch is directly provided by Ubuntu on their DVD) and part of Debian SID (the "work in progress" version, which will be the next official Debian version), as well as the open source repository Homebrew in order to meet the classic requirement for good, open source community building. MediaArea reported that they are also working with other Linux distribution (e.g. Fedora) to reach different open source communities.

4.1.4   Open source approach

The task here was to provide the PREFORMA Consortium with a description of how the suppliers had addressed the relevant open source topics, best practices, and licensing. Furthermore, the suppliers were required to report on how they had progressed in setting up an open source community around the developed tool.

In the Final report, the suppliers also had to provide a description of how they had addressed the relevant open source topics, best practices, and licensing issues identified by the University of Skovde in the feedback reports on their Intermediate releases.

A. veraPDF

In the **Intermediate report** veraPDF reported that they - for every software release - had uploaded the following zip archives to the PREFORMA open source portal:

- One containing the cross-platform installer and startup shell files for Windows and bash;

- Another containing the full source of veraPDF and its dependencies except for log4j (see below);

- Three archives containing the build tools needed to compile the software, Java JDT and Maven, with example scripts to install the tools and compile the code. The build environment comes in 3 forms, one for Linux, one for MacOS and a final one for Windows.

veraPDF also reported an issue with the source archive concerning building log4j from source. The log4j dependency, inherited from PDF Box, will be removed when they merge the "greenfield" parser development later this year.

The website had been re-designed and its content revised in response to PREFORMA Consortium review feedback. veraPDF had also introduced a dedicated documentation site. External contributions to the documentation will from now on be made via GitHub pull requests, in the same manner as contributions to the source code.

Reported Open Source Best Practices in use:

- Source code, validation profiles, test corpus, and documentation on GitHub;

- TravisCI for first stage of continuous integration;

- Jenkins server for continuous deployment;

- Continuous integration testing against acceptance corpora;

- Continuous deployment of development and release installation packages;

- Maven repository for all development and release source, javadoc and jar packages;

- Signed GitHub tags for all development and release versions;

- Codecov for test coverage (moved from Sonar to allow community administration);

- Codacy for static code analysis (again moved from Sonar for community administration ease).

In the **Final report** veraPDF reported that since the last report, they had developed the "greenfield" PDF parser and validation model that will replace the Apache licensed PDF Box dependencies, used by veraPDF until now. At the same time, they had removed the Log4J code from their library. Thus, the forthcoming release 0.26 will be the first source package that will meet the PREFORMA licensing conditions.

B. EasyInnova

EasyInnova reported in the **Intermediate report** that, following the recommendations of the PREFORMA Consortium in the 1st part of the Prototyping phase, they are now using all functionalities provided by their open source repository GitHub.

They also reported on a selection of areas of progress:

- All the developments of the DPF manager have been organised into milestones with information provided on when a new feature or a bug fix will be released. All the enhancements and bugs will be reported and linked to the corresponding milestone in the issue tracker section. All the DPF Manager releases have been fully documented in the GitHub with all new features and bugs fixed.

- All the bugs reported in the GitHub would now be treated as high priority and will be answered as soon as possible. Following the evaluation of the bug a milestone release is assigned and the bug will be immediately corrected in the master branch as a quick fix. When required, a new test will be created to prevent the same bug appearing again.

- All the code is now self-documented, using the java doc standard.

- All the code submitted is now licensed with the required MPL v2+ and GPL v3+ licenses and all the test images will be distributed using the CC Attribution-ShareAlike 4.0 International license. To guarantee that any contribution in their repository accomplish

with license requirements, EasyInnova are planning to introduce a Maven artifact able to check for any licensing requirements.

In their **Final report**, EasyInnova had attempted to address questions and comments made by the PREFORMA partner University of Skovde on the intermediate report.

Provisioning the source code

The University of Skovde reported that some contributions were made under licences other than the specific licenses required by PREFOMA. EasyInnova reported that they immediately had resolved this issue and updated all the files with the required header license information.

Provision of a roadmap on the development platform

EasyInnova reported that they have updated the GitHub roadmap and milestones plans for the period until 2020 with a release plan which targets external contributors. They expect that the release of the TI/A draft implementation will promote the collaboration of external contributors whose aim will be to integrate the DPF Manager in their software.

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

To fulfil the requirements while waiting for clarification from the University of Skovde, EasyInnova removed their unitary test that use a testing dependency with Junit framework under EPL licence.

The Skovde partner reported some dependencies from libraries under Apache 2.0. In deliverables during the Design phase, EasyInnova included a list of dependencies that would be used during the Prototyping phase. The document named "Intellectual property rights report" included the review and analysis from the perspective of IP law. 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.

EasyInnova declared that they are aware 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 copy lefts licences). On the other hand, software under GPL3+ or MPL2+ can use libraries under Apache 2.0 and be freely and legally distributed. As soon as EasyInnova have received clarification from the PREFORMA Consortium, they 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 to avoid java dependencies. In the final release, the runtime will no longer be included in the executables, in order to fulfil the PREFORMA licenses requirements

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

EasyInnova have explicitly published all the rules used to validate the implementations in their web page documentation. This information taken from the GitHub folder contains the current files used to validate the specifications. These files are in an XML format and could, therefore, be read by any text editor. To facilitate the navigation, EasyInnova included a search box which can search any text in the rules description, title or error code.

Provision of software which can be redistributed in a cascade

Once the previous issues are resolved, the software will be redistributed in a cascade with the PREFORMA licence requirements fulfilled.

C. MediaArea

In the **Intermediate report** MediaArea points out that by rewriting specifications into Markdown and posting on Github, more people will be able to contribute to the specifications, including archivists without development or prior GitHub experience. Work on this integration has grown since the last report and the community is thriving more than before, particularly after recent outreach efforts.

MediaArea is in the position of being able to leverage its existing software packages, MediaInfo and MediaTrace. MediaInfo already has a very large user base among audiovisual archivists. MediaTrace is used by MoMA (NYC).

In the **Final report** MediaArea focus on progress since the intermediate report.

They have continued to work closely with FFV1 and Matroska format designers to clarify the format specifications. This work has been done in the open via the IETF CELLAR listserv and Matroska and FFV1 GitHub pages so that it is transparent and anyone can contribute.

MediaArea continuously encourage community members to get involved through 1-on-1 coaching and instructions, giving speeches at conferences, and by holding workshops about MediaConch, file validation and conformance checking.

In November 2016, MediaArea plan to co-host an event aimed at supporting learning and assisting in the standardisation of Matroska and FFV1 at the New York Public Library. Both MediaConch and Archivematica (who has integrated MediaConch) will lead a workshop for the Association of Moving Image Archivists to demonstrate to archivists how to participate, contribute to, and use these open source projects.

MediaArea has also funded the development of FFV1 and Matroska features in external open source projects in support of the goals of the PREFORMA Challenge. Through these efforts MediaArea has offered financial support to several development tickets.

MediaArea has also supported FFmpeg developers in improving the effectiveness for FFV1 in film preservation and provided recommendations and reviews to the FFmpeg encoder and decoder of FFV1.

The source code made by MediaArea has been made freely available in open source form as per the agreement between the 6 suppliers from the Design phase and the PREFORMA Consortium: as stipulated in the a letter "Clarification from PREFORMA on licensing requirement" by the Consortium "All code (software and libraries) required to compile and/or execute the Conformance Checker in a production environment has to be freely available in open source form under generally recognized free software licenses compatible with the GPLv3++ and MPLv2++ to enable redistribution of the whole package under these two licenses."

MediaArea declared that they had carefully considered patent issues and had acted to significantly reduce patent risk in the code delivered to PREFORMA.

4.1.5 Standardisation efforts

The task here was to provide the PREFORMA Consortium with a description of how the suppliers had actively contributed to the standardisation process in their domains, by means of providing feedback on the existing standards as well as on how to support emerging standards.

A. veraPDF

In the **Intermediate report,** veraPDF informed that since the last report, the ISO committee for PDF/A (ISO TC 171 SC 2 WG 5) had met in Ghent, Belgium in May 2016 with a number of PDF Validation TWG members attending. At this occasion, the PDF Validation TWG submitted another ten points of ambiguity in PDF/A to the ISO WG for consideration and resolution.

veraPDF reported that they are planning to submit additional items for clarification in PDF/A-Next for consideration at the WG 5 meeting in Sydney in November, 2016. veraPDF anticipated that their collaboration with WG 5 and WG 8 will continue throughout the course of 2016 and into 2017, as they continue to resolve remaining matters of interpretation regarding the specification.

In the **Final report**, veraPDF reported about the plans for the meeting of the ISO committee for PDF/A (ISO TC 171 SC 2 WG 5) in Sydney. Their standards-development efforts were also summarised in the following bullet points:

- They have established that existing Parts of PDF/A will not be amended in any way. Any clarifications to existing ambiguities will be addressed in the forthcoming new Part for PDF/A, presently termed "PDF/A-Next";

- They have driven awareness of the need for PDF/A-Next, and led in its development;

- They plan to submit a request for an ISO WG resolution calling on the PDF Association to publish its Resolution of Ambiguities document as a PDF Association Technical Note.

B. EasyInnova

In the **Intermediate report** EasyInnova recapitulated that they have continued working on a definition of a technical specification for a dedicated TIFF format for the use in archives (called TI/A) to be published as an ISO recommendation.

Since last intermediate report they had made progress on several fronts. First, EasyInnova met with Adobe and made clear that they had followed Adobe's advice to work on an ISO recommendation instead of a new file format. Adobe declared, that they understood the task of handling existing assets in museums and archives, and that they were willing to help EasyInnova with their plans and in the discussion with ISO.

A first draft of the recommendation had been produced by EasyInnova, based on the discussions with experts inside the TI/A Intranet and the community of memory institutions. Besides this technical approach, EasyInnova wanted to get a detailed picture of the assets already existing in memory institutions, to avoid that the recommendation implies unnecessary migration. They had, therefore, asked for access to "hot" image data of large institutions to do an analysis of the structure of TIFF files already archived by memory institutions. For this analysis, EasyInnova had organised access to about 2 Million TIFF files from three large archives in Switzerland. EasyInnova also set up an infrastructure for the job in those archives

because they were neither allowed to take the data out of the building nor use the institutions own infrastructure.

In the **Final report** EasyInnova highlighted continuous progress concerning the work on the TI/A recommendation. The outcome of the analysis of TIFF files already archived by memory institutions showed 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 popular feature. Files with a quantisation depth of 16 bit were on the other side very rare, in strong contrast to today's best practice. LZW or JPG compression can only rarely be found, most likely due to lack of acceptance of such compression schemes. Discussions with various experts in digital archiving had also made it clear that the most controversial topic is "to compress or not to compress ". Therefore, the recommendation EasyInnova will hand over to ISO should not go into technical details of image compression, nor will it force any institution to aband the use of image data compression as long as the technology applied can be regarded as technologically stable, well documented and supported by software.

The draft of the recommendation had been brought to a pre-final stage. The final draft is available at the TI/A intranet and submitted to the ISO TC171 6 to start the official process of standardisation. A deadline had been imposed by the Technical Committee so to include the discussion about TI/A in the agenda for the working group meeting in Sydney.

C. MediaArea

MediaArea underlined in their **Intermediate report** that they are collaborating with several communities to make progress in the standardisation efforts for FFV1, MKV, and EBML (the foundational format of MKV). Being able to work closely with the FFV1 and Matroska format designers will ensure standardisation in line with the formats vision. MediaArea had also worked closely with other preservationists to collect user needs and feedback, to be applied to the forthcoming standardised specifications. The standardisation work is done through IETF, specifically the CELLAR working group. The MediaArea team participated in the first meeting of the CELLAR working group at IETF96 in July 2016 in Berlin.

Surrounding the IETF meeting, MediaArea co-hosted a symposium in Berlin, called "No Time to Wait" that focused on the standardisation and use of FFV1 and Matroska in archives. The symposium provided excellent opportunities to learn from and collaborate with related standardisation efforts such as Google's work to standardise a Matroska branch and the Library of Congress's work in defining AS-07.

In the **Final report** MediaArea highlighted their continuous work on standardisation. Since their intermediate report, the draft for EBML has been upgraded to a working group item and received another revision in the CELLAR document tracker.

Within the MediaConch project, their team had developed XML Schemas for each of their XML based reports in order to better define and standardise the XML-based expressions used in the software.

Community work had also continued to prepare for a set or recommendations for the archival use of Matroska and FFV1 in a manner analogous to the Library of Congress's similar standardisation effort in AS-07 which relies upon MXF and JPEG2000.

4.1.6   Impact assessment, sustainability, future use and exploitation

The task here was to provide the PREFORMA Consortium with a description of the supplier's ideas and plans related to the sustainability, future use and exploitation of the results of their projects. The suppliers were required to include some evidence of the impact that their projects had generated so far for the memory institutions and for any other relevant target group.

A. veraPDF

In the **Intermediate report**, veraPDF stressed that the active use of a tool such as a format validator depends strongly on its stability and feature completeness. It was for this reason, they claimed that PREFORMA has not seen an installation of PDF Manager in a working environment yet. However, during their dissemination efforts veraPDF had seen strong interest in their software, both from the digital preservation community and the document industry. Several vendors of archiving systems had also expressed an interest in including veraPDF in their offering. A first approach from outside these communities came from the Dutch office for standardisation, who asked for assistance with the integration of veraPDF into a public website, which should verify the formats of content of Dutch governmental (and potentially third-party) websites. Negotiations about this were reported to be on-going.

Part of the Open Planet Foundation´s (OPF) mission is the sustainability of project results in the digital preservation field. Since the end of the Planets project, OPF had sustained and made available the results of a whole set of projects. They had created the tools and processes needed to keep software and knowledge collected in research projects available for long-term use. These tools were already in use for the PREFORMA and will sustain veraPDF results after the project lifetime. However, OPF also has a model to extend such results, which they call stewardship. Currently, this model is applied to JHOVE but it will be extended to veraPDF after the PREFORMA officially ends.

veraPDF also presented a comparison with the planned delivery schedule for the current milestone, defined at the end of the Design phase.

In the **Final report,** veraPDF focused on four topics:

1. The case for maintaining and sustaining veraPDF

The feedback received from users had provided eveidence that veraPDF is meeting a deeply-felt need on the part of memory institutions, and commercial organisations concerned with the long-term viability of their records. PDF/A, however, represents a small fraction of the files such organisations process; the clear majority are simply PDF files, and do not claim to conform to PDF/A. Memory institutions really need the ability to validate the conformance of PDF itself, but that would be a large scale project. In addition, there are other PDF subset specifications of interest to memory institutions, including PDF/A-Next, PDF/E (engineering), PDF/UA (universal accessibility), PRC and more. veraPDF may extend to cover all aspects of PDF, PDF subset standards, related specifications, and even third-party standards such as XMP or PRC.

2. Sustaining veraPDF

Maintaining and developing the software in a professional fashion and maintaining the attention of the industry, requires financial resources and organisational commitment. The veraPDF consortium plan to develop a revenue-generation system based on the veraPDF software and veraPDF.org to achieve these resources.

3. The veraPDF Project

In order to fund software maintenance and future development, the veraPDF consortium plan to:

- create mechanisms to facilitate the aggregation of anonymous test data from veraPDF users and to generate conformance reports from the universe of files tested;

- provide memory institutions and commercial (industry) organisations with access to the conformance reports based on an annual subscription;

- provide a means of demonstrating support for the veraPDF project via a "sponsors" page, or similar.

4. Grants

In addition to activities intended to generate revenue directly, the veraPDF consortium had approached 3rd parties to request grant monies to continue development of the software.

5. Consulting

The veraPDF consortium had been approached by memory institutions with requests for commercial applications based on veraPDF.

B. EasyInnova

EasyInnova underline in their **Intermediate report** that low technical knowledge about the TIFF format makes memory institutions unaware of when their TIFF assets do not conform to the standard requirements. Therefore, EasyInnova´s first activity had been to create awareness.

Getting as many early adopters as possible had been crucial for the EasyInnova strategy. Some of the early adopters 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. These early adopters have allowed EasyInnova to receive 2.526 reports of analysed 5.603 files, which had led to the discovery of 18 private tags and 16 typical errors in the baseline.

DPF Manager had been used intensively in two cases:

- The PREFORMA partner Packed validated the conformance of around 40.000 TIFF files of scanned paintings for *La Fundació Tàpies* (http://www.fundaciotapies.org/).

- The University of Basel had analysed 2 Million TIFF files from 3 big memory institutions in Switzerland to understand which variants and tags that have been used in the past to create TIFF files.

Collaboration with associations or entities protecting the interests of memory institutions is crucial, veraPDG stated in their report. Examples of such on-going collaborations are:

- KOST-CECO in Switzerland, which is managing the digital preservation issues of their 30 members;

- EuropeanaSpace project, which is using DPF Manager to validate the TIFF files uploaded by memory institutions.

The long-term success of the DPF Manager depends on the establishment of a successful community around the open source project, but also on the development of a set of commercial

services to ensure the project is sustained when the public funding period is over. In the documentation submitted at the end of the Design phase, EasyInnova included a brief business plan, which outlined that their 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. With the purpose to offer services in the near future, veraPDF had already registered the domain www.dpfmanager.com in order to offer services in the future.

In the **Final report**, EasyInnova gave the following complementary information:

- Memory institutions increasingly understand that they must validate their TIFF files conformance to the standards and to the TI/A ISO Recommendation. The list of early adopters using DPF Manager before the final release is getting longer. The institutions using the DPF Manager at the time of the final report were the Swiss Federal Archives and the cantonal archives of Basel City and Saint Gallen. The increase in use by early adopters in the last months isvery significant; they have analysed the same number of files in 2 months than in all the years before. This means that by the submission of the final report, more than 4 Million TIFF files have been analysed using DPF Manager, a crucial test when validating the robustness of the tool, according to EasyInnova.

- When the final DPF Manager release is submitted, and all the functionalities (including the TI/A first specification) are available, EasyInnova will start looking for final users. The plan for the next months is to improve and complete the business plan.

C. MediaArea

In the **Intermediate report,** MediaArea highlighted three topics:

1. Extending into communities via Archivematica

The outcome of MediaArea's partnership with Artefactual will be to integrate MediaConch into Archivematica workflows. Archivematica is a popular framework for OAIS-compliant digital preservation with a robust user community.

2. Extending into other formats

Functionally, MediaConch have already the capacity to expand and support other file formats beyond Matroska, i.e. FFV1, LPCM and, via integration with the other PREFORMA suppliers, PDF and TIFF. MediaArea underlined in the intermediate report that there is a potential for MediaConch to become the conformance checking software for *any* audiovisual format, not just Matroska and FFV1.

3. Importance of symposium feedback

By using the "No Time To Wait! " symposium in July 2016 as an example, MediaArea point at the importance for SMEs to interact with the archival community regarding their needs and workflow specifications. It will help both parties to reach agreements on goals and principles in digital archiving.

In their **Final report** MediaArea provided an update on progress since the intermediate report:

1. Extending into communities via Archivematica

Initial integration of MediaConch was now available in the latest Archivematica release, which has been promoted during Artefactual's Archivematica Camp and a webinar specifically organised on this topic.

2. MediaConch integration at VIAA and Workshops at Tate Museum

As reported earlier, MediaArea has collaborated with VIAA and the PREFORMA partner PACKED to integrate MediaConch into VIAA's archival environment. MediaArea had also collaborated with the Tate Museum in the context of their Pericles project.

3. Future use

MediaArea had started to plan for the commission of supplemental media format diagnosis and support that could be added to MediaConch after the end of PREFORMA. They had also started conversations with cultural heritage institutions to find out the level of interest in supplemental media format sponsorship and software integration opportunities for institutional workflows.

4.1.7   Gap analysis and next steps

The task here was to provide the PREFORMA Consortium with a description of the status of the work compared to what was planned in the functional and technical specifications the suppliers had provided at the end of the Design phase. The suppliers were also supposed to critically highlight what was still outstanding in the current release and their plans to overcome these gaps. An updated version of their work plans and timelines should be included, so the PREFORMA Consortium members can presently and in the future review how requirements of the project have been met as well as the level of compliance to the suppliers own technical and functional descriptions.

A. veraPDF

veraPDF compared in the **Intermediate report** the present situation and the planned delivery schedule for their internal milestones, defined at the end of the Design phase. Some features were still in progress and not finished according to plan:

- Policy profiles Prototype implementation had not been tested against real institutional policy as planned, mainly due to the lack of real requirements from memory institutions;

- Internationalisation had not been worked on, as all the veraPDF documentation and error reporting was tied to the PDF/A standards which are only in English.

- PDF report templates were still in development and behind schedule.

An update of the situation was made by veraPDF in their **Final report**. The list of features not finished as planned was as follows:

- Policy profiles: Some concrete policy cases from the test corpus meetings were in place. New schedule - November 2016;

- Translations: There had been no call for translated versions, and the PDF/A specifications are in English. New schedule - the requirements for translations will be assessed during the Testing and validation phase, possibly in 2017;

- Report templates: PDF reports were still in development and behind schedule. New schedule - November 2016;

- Jpylyzer plug: In development and not completed. New schedule - veraPDF may be able to implement it in 2016, but it might be later (unsure).

- PREFORMA shell: Not completed, integration had taken a back seat until veraPDF had considered their validator functionally complete. New schedule - full prototype for PREFORMA experience workshop in Berlin in November 2016.

B. EasyInnova

In the **Intermediate report,** EasyInnova started with an exposé of development aiming at presenting a "red line". As described in the final report of the Re-design phase, and in the technical specification of the DPF Manager, the first prototype was focused on the TIFF Conformance checker. During the Re-design phase, EasyInnova internally re-designed the application from bottom to top focusing on the development of the shell component.

In the beginning of the 2nd part of the Prototyping phase, EasyInnova developed the re-design of the shell architecture, provided a basic interoperability between Conformance checkers, completed the planned interfaces, and developed the scheduler functionality. Regarding the TIFF Conformance checker, EasyInnova changed the Implementation checker module and designed the new report. The only task planned for this period that hadn't been finished was the Metadata fixer module, but EasyInnova was convinced that all the functionalities would be ready on time by the end of the Prototyping phase (without affecting the global plan).

Tasks in progress were the improvement of the TIFF/IT and TIFF/EP implementation. EasyInnova had also started to define TIFF classes for the Testing and validation phase and were improving the DPF Manager documentation. Next steps in the development of the DPF Manager, to be ready by the end of the Prototyping phase, would be the following ones:

- Complete the interoperability between Conformance checkers;

- Finish the XML configuration file, using and XSD standard, to be properly validated before being used by the program;

- Implement the TI/A standard validation, using the draft that will be submitted to the ISO working group;

- Finish the new report with the remaining formats JSON and PDF;

- Improve the Metadata fixer, making it able to not only detect and correct incoherences inside the metadata but also to discover file transformations not previously reported in order to reconstruct the file provenance.

The final task to finish was the evaluation and testing. EasyInnova explained previously in the end of Re-design phase report, that they wanted to create a reference public repository of test TIFF files to be used as a benchmark of TIFF reader/writer tools. Their aim was to prove that DPF Manager can validate the ISOs implementations more accurately than other tools. For that reason, EasyInnova plan to build an evaluation platform that can be used not only for the DPF Manager but also for other solutions.

In the **Final report** EasyInnova declared that they had finished all tasks described in the Phase 2 proposal and, on top of that, also finished some new functionalities and improvements not initially planned.

There was only one task that had not been developed yet, the integration on the DIRECT system. EasyInnova had also delayed the construction of a website to analyse the performance of the DPF manager compared with other Conformance checker tools. This task was not initially planned, although it became clear during the Re-design phase that it would be useful to have. Therefore, EasyInnova plan to include it in their tender proposal for the Testing and validation phase.

Planned next steps were focussed on the proposal for the Testing and validation phase, promotion of the collaboration with external contributions to ensure the long-term sustainable of open source communities as well as promoting the integration of the DPF manager into Digital Asset Management software.

Some of the new features that EasyInnova would like to promote to the external contributors were the following ones:

- DNG implementation checker; DNG is an open format from Adobe for storing RAW information. DNG uses 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, which can be easy integrated in our implementation checker;

- Implementation checker rules based on Xpath and XQuery selectors. The implementation checker in place uses custom rules, but EasyInnova would like to provide more flexibility by including different kind of selectors.

- Integration of the DPF Manager into the Archivematica open-source digital preservation system.

Regarding the TI/A Initiative, EasyInnova´s aim was still to convert their current proposal to an ISO Recommendation, as described in earlier reports.

Finally, from the community point of view, all communication will be oriented towards attracting new users of DPF Manager, memory institutions as well as software developers.

C. MediaArea

In the **Intermediate report**, MediaArea declared that verbosity settings are present in MediaConch but should be expanded to be more complex. There were plans to extend verbosity from today's two levels to five.

Integration with other conformance checkers was now an existing feature. MediaArea had merged the MediaConch GUI and the Web UI into a common UI for consistency and for ease of updating.

The IETF CELLAR working group had its first meeting as an official working group on 19 July 2016. In Media Area's timeline for July is "Present proposed Standard for FFV1 and Matroska atIETF 96." This was successfully done and approved by the group. The first RFCs for Matroska, FFV1, and EBML were submitted and comments were discussed and modified on the CELLAR listserv and on GitHub.

According to the MediaArea work plan and timeline, the fixer component should have been integrated into the software, but they were behind schedule with the Metadata fixer. Instead

MediaArea had prioritised the "performance optimisation" of the application, increased implementation checker work, and reviewed the policy checking elements of MediaConch.

In the **Final report,** MediaArea declared that the Metadata fixer, which was behind schedule at the time of the intermediate report, will be implemented in their coming October release.

The development of the Implementation checker followed the progress of the Matroska and FFV1 specifications within the CELLAR working group. However, more changes, clarifications and refinements are expected from CELLAR and MediaArea. Therefore, plans are already in place to continue to develop MediaConch in order to follow and contribute to CELLAR. As regions of the standard are clarified, implementation checks will be added to MediaConch.

The WebUI mentioned in the intermediate report will be implemented in November as a collaborative web framework for sharing institutional policies.

MediaArea had found that there are more opportunities to optimise the Implementation checker. The checking process often covers tens or hundreds of thousands of Matroska elements and FFV1 frames per file, so optimisation is essential both to allow expanding use and for the community of users.

For further context of current tasks and next steps, MediaArea made a reference to their issue trackers at GitHub.

4.1.8   End of phase reports

The aim of the administrative (and partly technical) report to be provided at the end of the Prototyping phase, was to give the PREFORMA consortium basic data in order to review the three suppliers improved prototypes.

Apart from summarising the main results achieved, the End of phase reports should focus on:

- identifying possible deviations from what was in the original tender;
- describing the innovative aspects of the work;
- describing the activities that have been performed;
- describing any potential long-term collaboration / partnership entered;
- describing what the supplier has gained from this project and what have been the main benefits;
- describing the potential for exploiting the work;
- providing a description of how the money have been used and allocated.

The template for the End of phase report contains twelve specific questions and a request for a financial report, which combined clearly outline what the PREFORMA Consortium required in terms of answers. The template is enclosed as Annex 5.

The End of phase reports are available in the PREFORMA repository, and their content is shortly summaried in the following text.

A. veraPDF

The veraPDF consortium's main objectives were to:

- develop an open source PDF/A validation library;

- produce a comprehensive and authoritative test corpus providing pass and fail test files for all testable clauses of the PDF/A specifications;

- develop a cross platform, open source Conformance Checker application that meet the PREFORMA challenge for text documents;

- follow open source project best practices to establish an active community.

The open source components had to be dual GPL3+/MPL2+ licensed to comply with PREFORMA's software licensing requirements.

It was anticipated, that the process of creating the test corpus would identify ambiguities in the PDF/A specifications. The PDF Association established its PDF Validation TWG to discuss any issues uncovered and to connect the project to the ISO community responsible for PDF/A in order to ensure these issues were addressed in current and future parts of PDF/A. The PDF Validation TWG provided a forum for review, analysis and resolution of these ambiguities in a transparent manner. The PDF Validation TWG's relationship with the ISO community responsible for PDF/A, allowed the veraPDF consortium's work to be applied directly to enhancing development of next-generation PDF/A specifications.

veraPDF have developed several software components, all in Java. These components came together to meet the PREFORMA Challenge Brief concerning the Implementation Checker and Metadata Fixer, the Policy Checker, the Reporter, and the PREFORMA shell.

The veraPDF test corpus, available online, comprises over 1,500 PDF files created by the veraPDF consortium as a comprehensive PDF/A validation test suite. All veraPDF development and release software is tested against our validation test corpus. Test have also been made against existing PDF/A test corpora produced by the PDF Association.

The original technical specification suggested the use of Apache FOP for generating PDF reports. Given this isn't compatible with PREFORMA's licensing requirements, veraPDF have used an appropriately licensed alternative.

veraPDF reported that the most significant outcomes of their work include:

- Industry acceptance of the project's purpose, technical, and functional design;

- Successful development of a test-suite and software under the supervision of a large and diverse group of implementers from across the industry, including all industry-leaders in PDF/A development

- Successful engagement with the ISO community responsible for development of both PDF/A and PDF (ISO 32000) specifications, resulting in significant influence on the future development of both;

- Successful establishment in principle of an industry-supported validation project, thus opening the possibility of similar projects in the future (for example, validation of the next part of PDF/A, or other archival-related specifications such as PDF/E and PDF/UA);

- Successful implementation of a PDF/A conformance checker, a feature extraction engine for further policy checks and a metadata fixer as defined by the tender functional and technical specifications;

- Successful implementation of a "greenfield" PDF parser, currently the only open-source PDF parser under MPLv2+/GLPv3+ license;

The veraPDF project has established a dialog between the PDF industry and the cultural heritage/digital preservation communities. The heritage community cannot be expected to provide expertise across all formats they are responsible for preserving. Active dialog and co-operative projects with the industry sectors where specialist expertise exists is essential to the preservation of digital heritage.

veraPDF believe that their software will meet a deeply-felt requirement on the part of memory institutions, and commercial organisations concerned with the long-term viability of their records. PDF/A, however, represents a small fraction of the files such organisations process; the clear majority are simply PDF files, and do not claim to conform to PDF/A. As stated in earlier reports: what memory institutions really need is the ability to validate the conformance of PDF itself and eventually also other PDF specifications. veraPDF have in mind to cover all aspects of PDF, PDF subset standards, related specifications, and even third-party standards such as XMP or PRC.

B. EasyInnova

At the outset of this piece of work, the aim was to create an innovative tool and framework, DPF Manager, for gaining full control over the technical properties, metadata information and structure of digital content data objects intended for long-term preservation. It has been designed to address present as well as future challenges in digital preservation and in the standards compliance area. Another aim was to raise awareness of the fact that when storing a TIFF file, it is not enough to ensure that it will be technically preserved; there is a need to evaluate other parameters and features as well. Therefore, EasyInnova had as an objective to start a standardisation process in the ISO organisation to create a new standard for TIFF file for archival purposes. A third objective was to develop an application that not only fulfills all the functional requirements described in the PREFORMA Challenge Brief, but also satisfies the software quality model described in the ISO/IEC 25010

To ensure the viability of the open source community, EasyInnova created two different communities: one for the archival standardisation process of TIFF and one around the DPF Manager.

Finally, the aim was to provide the features needed to easily integrate DPF Manager in memory institutions using the OAIS model.

At the end of the 1st part of the Prototyping phase, the Implementation checker was able to check the TIFF baseline 6.0. During the 2nd part EasyInnova saw a need to split the TIFF Baseline 6.0 in two different implementations, following this the TIFF Baseline 6.0 document is organised in two parts: Part 1, the Baseline section, describes the features that all general-purpose TIFF readers should support; Part 2, the Extensions section, describes a number of features that can be used by special or advanced applications and not recommended for general interchange.

The standardisation process also deviated from the original aim, which was to create a new file format called TIFF/A. However, this approach conflicted with Adobe plans for TIFF. Adobe does not want to support a development of a new TIFF format, and they did not permit EasyInnova to use the TIFF trademark. The name was, therefore, changed to TI/A (Tagged Image for Archival), fully supported by Adobe. Thus, TI/A is now a group of recommendations for the TIFF files for archival purposes, specially focussed on monitoring the existing files in archives but also on giving advice about the creation of new ones.

Another significant outcome of the project is the TI/A draft proposal, which has been brought to a pre-final stage. The final draft was submitted to the ISO TC171 by October 28th 2016, to start the official process of standardization. During the TI/A standardisation process research had been done to take into account in the TI/A specification the current status of the TIFF files already preserved by memory institutions.

C. MediaArea

In the beginning of the 2nd part of the Prototyping phase, the overall goal of MediaArea was to design and develop software that consists of an Implementation checker, Policy checker, Reporter, and Metadata fixer, aimed at a selected list of formats, and to collaborate with partners who would integrate the software into their workflows.

MediaArea are in partnership with Artefactual Systems, a consulting and development company that develops an open source and OAIS-focused repository solution called Archivematica, which is used by several memory institutions throughout the European Union. During this partnership, selections of the PREFORMA developments have been integrated into Archivematica, including the Conformance checker.

The standardisation process started with standard drafts for both Matroska and FFV1. This standardisation process will last longer than the PREFORMA project, but MediaArea plan to keep working on it during the Testing and validation phase and after the end of the PREFORMA project. MediaArea are ahead in the standardisation and development of EBML/Matroska but behind with FFV1.

MediaArea has changed some of the priorities compared with initial plan:

- Fixer: Has lowered the priority of the fixer in order to prioritize making the policy checker more extensible;

- Policy checker: Has focused on addressing the initial user feedback, which included a need for more complex policies in the policy checker and support for more container formats;

- Public policies: This feature was not included in our initial goal but has often been requested by users, so it was prioritised and is now offered by MediaArea;

- Reporter: Did not consider the display concept early on, but had extended it to make it possible to customise reports.

The formation of CELLAR within IETF has brought together a strong community of open source developers and preservationists interested in the development of these standards.

Moving forward, there are several directions MediaArea hope to continue exploring in order to continue to develop, promote, and apply MediaConch and related work in archival, broadcast, and other media applications.

4.2   FEEDBACK FROM THE PREFORMA CONSORTIUM

As mentioned in the beginning of chapter 4.1, the PREFORMA Consortium provided feedback on the two formal releases submitted during the 2nd part of the Prototyping phase, the intermediate release and the final release. This was undertaken on both occasions through the Evaluation Committee (see section 4.3.1).

The evaluation carried out on the intermediate release was an informal one, but the final release was simultaneously completing the Prototyping phase. Therefore, the final release report, together with the End of phase report, was part of the material that the PREFORMA Evaluation Committee had to formally evaluate in relation to the PREFORMA Challenge Brief and to the suppliers functional and technical specifications from the previous Design phase.

All comments on each release were compiled into formal feedback reports and sent to the suppliers respectively.

Feedback and questions on the monthly releases was also continuously raised in the GitHub or by mail directly to the suppliers by external testers of the software. In deliverable D3.7.1 Chapter 3, information is available on contributions of non-consortium members to the development of the suppliers Conformance checkers.

The following sub-sections summarise the feedback from the PREFORMA Consortium on the formal releases, although in more detail on the Final releases, and on the End of phase reports.

4.2.1   Feedback on the Intermediate releases

In the feedback document on the intermediate releases, the PREFORMA Consortium compiled the comments received from the members of the PREFORMA Evaluation Committee (end of July 2016), including those on the text of the accompanying reports. These comments related both to specific and more general issues. Sometimes they reflected different opinions, but were overall meant to provide useful input for the last period of the Prototyping phase, and serve as a base for further discussions. Most of the comments on the intermediate reports were related to the open source approach.

A. veraPDF

Results from examination

The Evaluation Committee noted that archivists who participated in demonstrations of the tool had reported that they will be able to test:

- more than one file at the time without having to use batch commands that are usually not suitable for a normal use;

- an unknown PDF/A-file and get the response on which PDF/A type it is. (veraPDF 0.20.3 seems to have an "Auto-detection" feature.)

It was also noted by the Evaluation Committee that http://demo.verapdf.org/ is up and running, although the report does not provide detailed information on reasons for errors and on links to further explanation, possible solution, etc. A question was also raised about the progress on user friendly error messages.

The release had been tested using veraPDF corpus, files from Isartor test suite, and local PDF and PDF/A files. Validation of PDF/A compliant and not compliant files against selected profiles had been successful. Reports can also be generated and saved.

The software validated as it should, and it could also produce view reports in XML and HTML and save reports. However, these reports were still not user friendly enough for non-technical end-users.

The Conformance Checker

The Evaluation Committee found installation, setup and use easy to handle but it would be useful to have batch processing functionality. It was noted that manual selecting of every single file for checking large number of files could affect the acceptance of the tools. The information available via menu "About" did not say much about the Conformance Checker.

The Intermediate report

veraPDF had, according to the Evaluation Committee, provided a good overview of changes made from version 0.6 to version 0.18. The supplier had also provided a detailed description of completed features.

Concerning testing, the Evaluation Committee noted that real world data sets were not shared due to IPR, but it would have been good if the results from testing had been shared. The Committee also noted that volumes and range that had been tested appears adequate. The large-scale testing of British Library files was very useful.

Regarding dissemination and community building, the Evaluation Committee noted that veraPDF had not reported significantly on community building. They had, however, been very active at conferences and published a many press releases, but the number of twitter followers, veraPDF news subscribers, and PDF Validation TWA subscribers was relatively low.

The Evaluation Committee underlined that the activities concerning standardisation were impressive. It was also noted that veraPDF had provided a detailed description of the status of work done and of planned next steps.

Open source approach

Concerning the open source approach, the PREFORMA expert partner on open source, University of Skovde, had reviewed this task. Overall, they found that the supplier had made progress, but they made several observations summarised in the following nine issues:

Issue #1 – Provision of source code

The complete source code had been provided as single zip-files on the OSP (Open Source Portal) but not always on monthly basis as required by PREFORMA. It was noted that the software provided contains code licensed under Apache 2.0. However, the supplier has planned to remove Apache licensed code to be in line with PREFORMA's licensing requirement.

Issue #2 – Provision of a roadmap on the development platform

It was noted, that the report from veraPDF lacked details concerning the requirement for an up-to-date roadmap, focused on external contributions and development beyond the life-time of PREFORMA. It was also unclear to what extent the supplier plan to address this requirement. Furthermore, it was noted that the content of the roadmap currently provided on the veraPDF´s website primarily targets the PREFORMA Consortium instead of external potential contributors.

Issue #3 – Identical software under both "GPLv3 or later" and "MPLv2 or later"

It was noted, that source code had been provided on the OSP, but not under the two specific licences required by PREFORMA. Many files containing source code, in the zip-file provided on the OSP, lacked licence information in the header.

Issue #4 – Provision of executable of the software on the open source portal

It was noted, that the supplier must provide an executable of the source code for each deployment platform on the OSP in order to fulfil PREFORMA licensing requirements.

Issue #5 – Provision of executable of the software for use via web browsers

PREFORMA requires that the software can be used via standard web browsers. It was noted, that the report from veraPDF lacked details concerning this requirement, and that it was unclear when it will be fulfilled.

Issue #6 – Provision of detailed documentation concerning interpretation of the technical specification of each file format

PREFORMA requires that the supplier provide detailed documentation concerning interpretation of the technical specification of each file format used. It was noted, that the report from veraPDF lacked details regarding the requirement on documentation, and that it was unclear when it will be fulfilled.

Issue #7 – Provision of software which can be redistributed in a cascade

PREFORMA requires that the supplier provide all code (i.e. all source code; tool chain for building executables; and executables etc.) under open source licences on the OSP, and that all code can be distributed and redistributed by any individual. It was noted, that the report from veraPDF lacked details concerning this requirement, but source code had been provided on the OSP, but not been under the two specific PREFORMA licences, which inhibits provision of software that can be redistributed in a cascade.

Issue #8 – Clarifying that the open source project has obtained all necessary rights to promote external contributions

It was noted, that the software provided implements (or there is a plan to implement) files in formats which was not included in the tender. One strategy which would contribute to clear the communication concerning IPR would be to clarify and improve the content in the roadmap to convince any potential contributor that the supplier has obtained all necessary rights.

Issue #9 – Clarity in licensing of provided source code to promote opportunities for re-distribution in a cascade

It was noted, that source code was provided under unclear conditions and did not fulfil PREFORMA licensing requirements. To clarifying licensing conditions for the PREFORMA software, all files containing source code must contain explicit information which clarify that these files are provided under the two specific open source licences required by PREFORMA.

B. EasyInnova

Results from examination

Members of the Evaluation Committee had found it easy to install the GUI, just one click. When checking the online-validator, there were problems and an examination of it was not possible. However, the overall impression of the software compared to earlier version was that it had evolved in a positive way. Several stability and performance improvements had been made, and the reporting tools were also working well. In the runtime check, four issues were identified: two of them were bugs and two were feature wishes.

The Conformance Checker

The Evaluation Committee noted that it was easy to set up the conformance checker.

The Intermediate report

The Evaluation Committee found it very impressive that EasyInnova had improved the performance of their software by 500%. The use of Maven was also somtthing very interesting,

as it allows DPF manager to connect with any external project, as well as the added multilanguage feature.

It was noted by the Evaluation Committe that EasyInnova had manually created test cases, and that they had identified new private tags because of contacts with early adopters of their software. Futhermore that EasyInnova want to create a new website on TIFF test cases to evaluate DPF Manager and compare it with other existing tools.

EasyInnova had also done a good job in building an ISO recommendation about TI/A.

Open source approach

The open source approach was also for this supplier reviewed by the PROFORMA expert partner on open source, University of Skovde. Overall, they found that the supplier had made progress, but they made several observations:

Issue #1 – Provision of source code

It was noted, that the complete source code (provided as single zip-files on the OSP) had been provided monthly for almost all months. Furthermore, it was noted that the software provided contains code licensed under Apache 2.0, and consequently the complete source code had not been provided under the two specific PREFORMA licences.

Issue #2 – Provision of 'roadmap' on the development platform

It was noted, that despite comments concerning provision of a roadmap in earlier review comments, the report from EasyInnova lacked an up-to-date roadmap focused on external contributions and development beyond the PREFORMA project. It was also unclear to what extent the supplier plan to address this requirement.

Issue #3 – Identical software under both "GPLv3 or later" and "MPLv2 or later"

It was noted, that source code had been provided on the OSP, but not under the two specific PREFORMA licences required. Several files with source code provided on the OSP had licence information in the header, which implies that the complete source code had not been provided according to PREFORMA licensing requirements. There were also references to source code which was not included in the zip-file provided on the OSP. This implies that the complete source code had not been included in the zip-file. Furthermore, some import statements in the source code provided in the zip- file, refered to source code provided under licences that do not fulfil PREFORMA licensing requirements.

Issue #4 – Provision of executable of the software on the open source portal

Since the complete source code had not beeen provided under the two specific PREFORMA open source licences, it follows that the executables did not fulfil PREFORMA requirements. It was stated that the supplier need to provide an executable of the source code for each deployment platform on the OSP to fulfil PREFORMA licensing requirements.

Issue #5 – Provision of executable of the software for use via web browsers

PREFORMA requires that the software can be used via standard web browsers. It was noted, that the report from EasyInnova stated that there is "an online validator, which can be accessed publically in the DPF Manager website". However, when University of Skovde tried to use it to validate files, the application returns an error message for all files.

Issue #6 – Provision of detailed documentation concerning interpretation of the technical specification of each file format

It was noted, that the report from EasyInnova – like the report from veraPDF - lacked details concerning this requirement, and that it was unclear when it will be fulfilled.

Issue #7 – Provision of software which can be redistributed in a cascade

PREFORMA requires that the supplier provide all code (i.e. all source code; tool chain for building executables; and executables etc.) under open source licences on OSP, and that all code can be distributed and redistributed by any individual. It was noted, that the report from EasyInnova lacked details concerning this requirement. It was also noted, that source code had been provided on the OSP, but not under the two specific PREFORMA licences, something which inhibits provision of software that can be redistributed in a cascade. Furthermore, it was noted that a tool chain for building executables had not been provided in zip-files on the OSP, but was instead provided under a proprietary licence. The PREFORMA requirements must be addressed by the supplier.

Issue #8 – Clarifying that the open source project has obtained all necessary rights to promote external contributions

It was noted from the EasyInnova report that the software provided on the OSP implements (or will implement) a file format that was not included in the tender (TI/A). It is fine for a supplier to provide software beyond PREFORMA requirements, but the supplier then needed to clarify issues concerning IPR and improve the content in the roadmap so that any potential contributor is convinced that the supplier has obtained all necessary rights for the implementation of the software in mind.

Issue #9 – Clarity in licensing of provided source code to promote opportunities for re-distribution in a cascade

It was noted, that source code was provided under unclear conditions and did not fulfil PREFORMA licensing requirements. The situation was more or less the same as for supplier veraPDF.

Issue # 10: Modular software architecture

It is important that the architecture for PREFORMA software is modular, to enable easy use of a specific part of the software which covers only one of the file formats tendered. It was noted, that the report lacked information concerning transparency and modularity for provided software. The supplier needed to clarify how any potential user can use a specific subset of the software that only deals with one version of TIFF.

C. MediaArea

Results from examination

The Evaluation Committe noted, that tests done with AVI/FFV1 files produced at the Austrian Mediathek worked well. The GUI for Windows was used for the tests. A comment from testers was that the offered profiles generally didn't fit but it was quite easy to define an individual profile. However, to define an individual profile that covers all the important fields required was more sophisticated. When checking local files and folders with own datasets, only smaller inconsistencies were noticed when using MediaConch (like incorrect text on a button).

The Conformance Checker

The Evaluation Committee noted a large-scale improvement when compared to the last evaluation. The checker has become quite complex, but is in most cases self-explaining. For

those that require more details, the existing help functions are useful. The Metadata fixer was missing, but a satisfactory explanation was provided in this regard,

The Intermediate report

The description of the changes made since the intermediate release is satisfactory and in line with the experience of the Evaluation Committee during the practical tests. The Committee also found that the publcation and explanation of each testfile of the Matroska test collection set in GitHub was very helpful.

Open source approach

Like the intermediate reports from the other two suppliers, the open source approach in this report has been reviewed by the PROFORMA expert partner on open source, University of Skovde. Overall, they found that the supplier had made progress, but they made several observations:

Issue #1 – Provision of source code

It was noted that the complete source code (provided as single zip-files on the OSP) had been provided monthly as required by PREFORMA. It was also noted from an inspection of the content on the OSP, that source code for Windows was provided as a single zip-file on the OSP. In norder to fulfil the requirements from PREFORMA, the complete source code in zip-files a need to be provide under the two specific PREFORMA licences on the OSP. The software provided also contained code licensed under other conditions. Furthermore, software included in the zip-file provided on the OSP had dependencies to other software provided under other licences which was not included in the zip-file.

Issue #2 – Provision of 'roadmap' on the development platform

Despite comments concerning provision of a roadmap in review comments provided earlier, the report from MediaConch lacked details to meet the requirement for an up-to-date roadmap focused on external contributions and development beyond PREFORMA. It was also unclear to what extent the supplier plan to address this requirement. It was noted, that the supplier had provided a roadmap on its own website, but this roadmap was PREFORMA specific and not available on the open development platform. Further, its content did not reach beyond the PREFORMA time frame.

Issue #3 – Identical software under both "GPLv3 or later" and "MPLv2 or later"

It was noted, that source code has been provided on the OSP, but not under the two specific PREFORMA licences. Several files containing source code in the zip-file provided on the OSP, contained licence information in the header, which implied that the complete source code had not been provided according to PREFORMA licensing requirements. It was also noted that there are references to source code which was not included in the zip-file provided on the OSP. Furthermore, some of the included statements in the source code provided in the zip- file refer to source code provided under other licences that do not fulfil PREFORMA licensing requirements.

Issue #4 – Provision of executable of the software on the open source portal

PREFORMA requires that an executable shall be provided for each platform. The supplier, therefore, needs to provide an executable of the source code for each deployment platform on the OSP to fulfil PREFORMA licensing requirements as detailed in deliverable D4.3.

Issue #5 – Provision of executable of the software for use via web browsers

PREFORMA requires that the software can be used via standard web browsers. It was noted, that despite previously identified problems related to the software release in November 2015, and related to use of software via web browser without a requirement for a login, the report from MediaConch lacked details concerning this requirement. It was also unclear when the supplier will fulfil this requirement.

Issue #6 – Provision of detailed documentation concerning interpretation of the technical specification of each file format

PREFORMA requires that the supplier provide detailed documentation about the interpretation of the technical specification of each file format used. It was noted, that the report from MediaArea lacked details concerning this requirement, and it was unclear if this requirement will be fulfilled.

Issue #7 – Provision of software which can be redistributed in a cascade

PREFORMA requires, as was previously mentioned, that the supplier provides all code (i.e. all source code; tool chain for building executables; and executables etc.) under open source licenses on the OSP, and that all code can be distributed and redistributed by any individual. It was noted, that the report from MediaArea lacked details concerning this requirement. However, the source code had been provided on the OSP, but not under the two specific PREFORMA licences, which inhibits provision of software that can be redistributed in a cascade as required by PREFORMA.

Issue #8 – Clarifying that the open source project has obtained all necessary rights to promote external contributions

To increase interest and maximise the opportunities for attracting external contributions, it is important to clarify to all potential contributors that the open source project has obtained all necessary rights for implementing all file formats in software. It was noted, that it looks like software provided on the OSP implements file formats was not included in the tender. It is of course beneficial if a supplier can provide software beyond PREFORMA requirements, but it is also important that the supplier in such cases clearly communicate and convince any potential external contributor that all necessary rights for implementing and distributing software under the PREFORMA licences has been obtained.

4.2.2   Feedback on the Final release and the End of phase report

The process for providing feedback on the Final release and the End of phase report varied from the methodology used on the intermediate release.

Firstly, the feedback was now part of the formal evaluation of the Prototyping phase, which also included the bids for the Testing and validation phase (see section 4.3.2).

Secondly, it was obvious from the appandent reports that the suppliers were still progressing but some of them had not managed to fully finalise the development in such a way that the Evaluation Committee could accept their last releases as final ones. Therefore, the PREFORMA Consortium decided at the plenary meeting in Berlin in November 2016 to extend the Prototyping phase for one month. A special task force was set up with members from the PREFORMA partners to evaluate the DPF Manager from EasyInnova, which appeared to have significant problems.

Thirdly, the open source experts, i.e. the partner University of Skovde, had in their review of the intermediate releases highlighted several issues that they judged as still pending. One issue in

particular was discussed both by the PREFORMA Consortium and by the suppliers. Could subsidiary components in the Conformance checker be provided in other open source formats than in GPLv3 or later and MPLv2 or later?

In Berlin, the PREFORMA Consortium confirmed, as mentioned earlier (see section 2.3.2), that the letter sent to the suppliers during the Design phase, saying that the core part of the Conformance checker should conform to the PREFORMA licensing requirements, was still applicable. It was also decided in Berlin that all three suppliers must isolate their core components from the subsidiary components (the latter provided as open source under generally recognised free software licenses compatible with the GPLv3 or later and MPLv2 or later). In the feedback reports to the suppliers, it was required that they should provide a "Core Distribution Package", defined by the PREFORMA Consortium.

Other, specific requirements were also set up, which in fact told the suppliers what the Evaluation Committee expected from them to approve their final releases. This was not the same as saying that all issues, noted as still pending in the feedback report on the intermediate release, were solved. Instead, it indicated an acceptable level of compliance with the PREFORMA requirements, which the Evaluation Committee later complemented with recommendations for further actions during the Testing and evaluation phase (see section 4.3.3)

A. veraPDF

- Update the documentation;

- Add an option to check multiple files from the GUI, including the possibility to choose specific files within a directory and making sure that the HTML report fuctionality is available also when checking multiple files;

- Provide a full release which includes the Policy Checker, and which is compliant with the PREFORMA licensing specifications (i.e. based on the" greenfield" solution).[20]

B. EasyInnova

1. Add documentation about the Metadata Fixer and the Policy Checker

The Evaluation Committee had checked the installation procedure, the user manual and reference documentation, and based on that, the following actions were required:

- Update the installation guidelines;

- Complete the documentation on the commands and options for the command line interface;

- Add documentation on the Policy checker in the reference documentation;

---

[20] The "greenfield" implementations of all core functionality and the PDF Box dependencies had caused a lot of discussion and obviously also delayed the final release.

- Add documentation on the Meta fixer in the reference documentation.

2. Align functionality between GUI and CLI

Required actions:

- Add functionality to the CLI for creating and editing configuration files;
- Add functionality to the CLI for fixing files.

3. Relaxing rules only as part of the policy checker

It is important that the Implementation checker is considered as an authority source, which can not be customizable in any way by a standard user. Instead, a standard user shall modify the behaviour of the Implementation checker through the policy checker, where it should be possible for him or her to override the default behaviour of the Implementation checker by setting policies, aimed either at restricting the Implementation checker rules or at extending/relaxing them. Therefore, the Evaluation Committee agreed that implementing rules within the Implementation checker module would be not compliant with the R&D objectives defined in the PREFORMA Challenge Brief. However, the Committee agreed that adding functionality to the Policy checker that identifies intended/desirable deviations from the specifications would be an alternative approach in line with the Challenge Brief. Actions required:

- Add functionality to identify intended/desirable deviations from the specification as part of the Policy checker and clearly distinguishable from the Implementation checker;
- Ensure that the report clearly states that conformance with the specification is unambiguous and that a report on intended deviation in invalid files is provided.

4. Reporting for large batches of files

The reviewers checked batches of 10, 100, 1000 and more files with DPF Manager, and required EasyInnova to increase performance for DPF Manager for batches up to 10 000 files for a single instance.

5. Explain how you managed checking 10/20 files/sec > we reached <2 files/sec

The Evaluation Committe had checked TIFF files with sizes generally between 20Mb and 60Mb at different machines. In these tests, the Committee could process files at 0,2 to 1 files per second. Required actions was to improve process speed up to 5-10 files/second and explain if and how it is possible to obtain the speed claimed by EasyInnova (10-20 files/second).

6. Clarification about TI/A

Statements about TI/A imply that TI/A is a file format and/or recommendation and/or guidelines. Therefore, it was required to clarify and remove any ambiguities about TI/A.

7. Clarification about Junit/Maven

The supplier had not been clear on the use of Junit and Maven, and are therefore required o clarify if Junit and Maven are necessary for building and/or running the Conformance Checker. If necessary, they must be provided with the distribution package.

8. Clarification about Java bundle

The Final report stated that the executables distributed up to now included the java runtime to avoid java dependencies. In the final release, the runtime will no longer be included in the

PREFORMA - Future Memory Standards
PREservation FORMATs for culture information/e-archives
EC Grant agreement no: 619568

executables, in order to fulfil the PREFORMA licenses requirements. In the End of Phase 2 Report it says, that EasyInnova will "provide two kind of installers", one with and one without a Java runtime environment. The Evaluation Committee, therefore, required the following action: Assuming that Java is required to build the source code, and knowing that all software necessary to build and run the Conformance Checker must be included in the distribution package, clarify that a "full package" (Java bundled) will be provided.

9. JSON license

The JSON license is according to GNU not compatible with GPL. Therefore, EasyInnova must provide instructions on how to disable this feature and completely remove the feature from future releases.

C. MediaArea

The Evaluation Committee required two actions regarding the open sorce approach in addition to the provision of a Core Distribution Package:

- To provide a build environment for each deployment platform on the OSP, which is provided under a license approved by Open Source Initiative (OSI), i.e. without the need for installing proprietary software (e.g. Microsoft Visual Studio 2013).

- To ensure that there is no reference to a file format that has not been included in the tender (e.g. JPEG 2000, MPEG-4, and MXF) in the PREFORMA Core Distribution Package.

4.3   FINAL RESULTS OF THE PROTOTYPING PHASE

4.3.1   The end of phase evaluation

At the end of the Prototyping phase, the results of the three open source projects were evaluated, following the same approach as the Design phase (see deliverable D8.2 *Design - First Report*).[21] Similar categories of reviewers were also involved but allocated differently.

Each memory institution and external expert had to review and evaluate one proposal, but was of course free to evaluate more proposals if they had capacity. The technical experts were asked to contribute by giving their opinion on two or even all the three proposals. Finally, the open source experts were asked to specifically look at open source and licensing related aspects.

Media format	Supplier	Technical experts	Domain experts	External experts	Open source experts

---

[21] The strategy used to evaluate the results at the end of the Design phase to select the suppliers who completed the tender was defined in deliverable D8.1 *Competitive Evaluation Strategy*. This strategy was also used when reviewing the suppliers' results during the Prototyping phase

text	VeraPDFa Consortium	FRAUNHOFER PACKED RA (Benjamin)	RA (Magnus) EVKM LGMA	Jozo Ivanović (National Archives of Croatia)	University of Skovde
image	Easy Innova - IMAGE	FRAUNHOFER PACKED RA (Benjamin)	SPK KIK-IRPA AJGI	Jan Dalsten Sørensen (National Archives of Denmark)	University of Skovde
av	MediaArea.net	FRAUNHOFER PACKED RA (Benjamin)	S&V KB GFC	Peter Bubestinger, Hermann Lewetz (Österreichische Mediathek)	University of Skovde

Table 1. The allocation of reviewers in the Prototyping phase

4.3.2 The decision-making process

The evaluation was made in two steps:

In a first step, all the reviewers had to assign their scores (and briefly justify them) to specified items under the categories "Impact on the Challenge" and "Technical Approach" in the template used for the evaluation (see Annex 6).[22] Evaluated in this step were (1) the Final releases including reports and (2) the End of phase reports, submitted by the suppliers between the end of October and the beginning of November 2016.

In a second step, the remaining categories ("Proposed Approach for Phase 3", "Quality of the Tender", and "Costs") were evaluated in the same way after the suppliers had provided their bids for the major Phase 3. The call for tender for Phase 3 was distributed in mid-November, and all three suppliers participating in the Prototyping phase were invited to submit bids. The PREFORMA Consortium decided that the results of the Prototyping phase, to date were of the necessary standard to permit the suppliers to proceed into the next phase.

When the tender form was distributed, the suppliers were informed that the indicative amount planned for the Testing and validation phase was revised. Originally, the plan was to allocate for an indicative amount of 105.000 EUR for each proposal. But since two of the memory institutions participating as procurers in the PREFORMA PCP have had problems affording the financial commitment planned at the beginning of the project, the PREFORMA Consortium revised the indicative amount to 90.000 EUR. It was at the same time clearly stated that this amount, being indicative, should not prevent the suppliers from offering a different price.

The electronic version of the tender form had to be submitted by email strictly by the 4th of December 2016. The paper version could arrive later, by the 12th of December 2016. Bids were submitted in time by all suppliers in the Prototyping phase.

A first version of the report from the Evaluation Committee and the assessment of open source implementation (updated version of deliverable D8.8) was available for discussion at the plenary

---

[22] The descriptions of all the items are available in deliverable D8.1 *Competitive Evaluation Strategy*

meeting in Berlin at the end of November 2016. These reports pointed at several issues still pending, which should be fixed by the suppliers in their coming December releases. Otherwise these releases would not be approved as the final ones of the Prototyping phase.

4.3.3   Decisions made by the PREFORMA Consortium

At the time of the plenary meeting in Berlin in November 2016, there were issues still to be fixed in the Conformance checkers, therefore the PREFORMA Consortium decided to:

- Send feedback reports to the suppliers asking them to fix all "open issues" and incorporate the results in their December releases. These were issues that the PREFORMA Consortium and Evaluation Committee considered mandatory and these issues needed to be addressed before the PREFORMA Consortium could decide if the suppliers had fully completed the Prototyping phase. The feedback reports on the final releases and the End of phase reports are discussed more in depth in section 4.2.2.

- Delay the start of the Testing and validation phase by one month (start at the 1st of February);

- Update the roadmap for the evaluation of the results of the Prototyping phase and of the new bids for the third Phase in accordance with the delay; the roadmap is attached as Annex 7.

On the 20th of January, a consensus meeting was held with the PMT and the Evaluation Committee in order to decide upon the evaluation of the results of the Protoyping phase and on the bids for the Testing and validation phase. For the suppliers EasyInnova and MediaArea it was decided to accept their end of December releases as final and to accept their bids. The acceptance of the releases was connected to several recommendations, pointing at future actions which the Evaluation Committee expect to be addressed during the next phase (see the Final evaluation report, enclosed as Annex 8). The process of final payment for these two suppliers started immediately. The third supplier veraPDF had, according to the Evaluation Committee issues still pending, and the Committee postponed this part of the evaluation. The pending issues were:

- verPDF had not yet uploaded on the PREFORMA server the so called "Core Distribution Package" as defined in the feedback report sent in December 2016.

- It was not possible to fully test the functionality of the Policy Checker, because the reviewers could not find the documentation on how to create and apply a policy file, and it looked like the GUI did not allow a user to create a policy file.

A deadline for veraPDF to provide missing software and documentation was also set.

At the 3rd of February, the Evaluation Committee decided to accept the end of January release from veraPDF as final, and the decision was supplemented with a list of recommendations on expected future actions. Their bid did not convince the Committee, being very far from the indicative amount of 90.000 EUR, but the veraPDF team reviewed it and reduced the price to a level in line with what the Evaluation Committee could accept. Therefore, in the end all three bids were invited to sign a contract for the Testing and validation phase. A formal decision on the acceptance of the bids was signed on the 10th of February by the coordinator on behalf of the Consortium, and the contracts were at the same day signed by the Swedish National Archivists.

5   POINTS OF PROGRESS

This chapter concludes, and to some extent also explains, the main points of progress made during the Prototyping phase. It is organised following two important aspects of progress: governance and compliance with requirements.

Good governance is crucial for all kind of projects, but for PCP based one's governance issues are particularly important because of the soluble character of these projects with the outcome based on development instead of purchased services and products. Section *5.1 Governance issues* summarises how governance issues, during the Prototyping phase, have progressed in parallel with the software development – on strategic level as well as on tactical and operational level.

Compliance with the PREFORMA requirements is the first issue of importance when summarising progress. Section *5.2 Complience with PREFORMA requirements* descibes the way the PREFORMA Consortium and its Evaluation Committee have tackled the issue of compliance but also the outcome of this process.

5.1   GOVERNANCE ISSUES

On a **strategic level**, the overall aim of the PREFORMA Consortium has been to secure the long-term perspective of the Prototyping phase from its start. This has been achieved through a progressing and iterative software development ending in an approval of the final releases of the supplier's Conformance checkers and in the provision of detailed recommendations for the future work. In a short-term, the task was to pave the way for the succeeding third major phase of PREFORMA, the Testing and validation phase. In the longer term to start planning for the after-life of PREFORMA and the sustainability of its outcome.

The PREFORMA Consortium has chosen two main approaches for handling the strategic issues:

- firstly, establish a forward-looking relationship with the suppliers;

- secondly, set up clear procedures for the follow-up (agreed with the suppliers) and for the management of a consolidated outcome, both activities documented in work plans and roadmaps (for the 2nd part of the Prototyping phase, see Annex 1 and Annex 7.)

These main approaches were converted into practical arrangements, some already established before the Prototyping phase began. The most apparent ones are:

Firstly, directional guidance in the form of expectations was placed on suppliers. Input was partly provided in the Call documents, partly outlined in the Work Package plan, but further developed in deliverable D4.3 *Functions of the Open Source Portal* and in clarifications and explainations by the PREFORMA Consortium, given so to speak "on the fly". These developed expectations covered how the supplier's software should be developed, distributed and released throughout the Prototyping phase, and were mostly addressed in the feedback reports on software releases in addition to supplier meetings and to clarification letters and e-mails. Specific working groups were also set up for specific tasks, such as the working group on interoperability of the different conformance checkers and the working group on special technical matters.

Secondly, a structure of software releases was set in place in the Prototyping phase, with immediate and final releases based on the functional and technical specifications presented by the suppliers during the Design phase. This meant that suppliers had to show continuity with the concepts and plans that they had previously presented, and to show that their work had reached a level of at least reasonable satisfaction for those who tested the software, i.e. the PREFORMA Evaluation Committee.

Thirdly, after the 1st part of the Prototyping phase, the PREFORMA Consortium intended to find visible and noticeable strong parts in the supplier's development work by measuring six points of progress: (1) Capabilities for software, (2) Evolving functionalities, (3) Usability, (4) Testing for quality assurance and accuracy, (5) Achieving reference Implementation, and (6) Awareness of what is still missing. Some extra check points were also implemented:

- The PREFORMA Evaluation Committee reviewed the results of the 1st part (final releases and reports) and compared them to the outcomes of the Design phase and the evaluation made at that time; in the 2nd part the Committee also undertook an extra informal review of the intermediate releases.

- A PREFORMA Delegation visited the suppliers in the beginning of 2016 to get a further understanding of their working conditions.

The overall outcome of these checks was that the project at the end of the 1st part of the Prototyping phase followed its compass course pretty well (for further information, see deliverable D8.3 *First Prototype Report*).

Fourthly, a discussion how to better highlight sustainablility started during the 2nd part of the Prototyping phase, pointing at two kind of actions:

- assigning a new partner in the project with focus on sustainability (a formal Amendment process will be initiated with the European Commission in the beginning of 2017);

- increasing the number of hands-on activities (workshops, test cases, and webinars) by organising training sessions during Spring 2017 with a focus on specific media; such activities had already been introduced, but they need to be more directed towards training and coaching.

A first version of a sustainability and exploitation plan was then submitted at the end of January 2017 in deliverable D3.7.1 *Initial version of Sustainability and Exploitation Plan*.

Although, the Prototyping phase with functionally complete and stable releases of the PREFORMA prototypes has just ended, preliminary impact assessments indicate that the potential of the solution developed by the suppliers are of great interest for the community, and also that considerable impact has been achieved (already)on memory institutions, open source companies and standardisation activities.

In their final repots, the Evaluation Committee commented the supplier's plans for starting paid services around their products, saying that, although there seems to be a demand for these services, the target audience may still be new to open source business models and therefore less motivated to pay for open source. The Committee recommended the suppliers to start developing a sustainability and exploitation plan to be released in the final versions of their softwares at the end of the Testing and validation phase.

On **tactical level**, the PREFORMA Consortium focused on securing services and agreements to ensure the project was efficiently managed. Clear contracts with suppliers and relalistic routines for payment, a transparent and roadmap-based testing and evaluation process of the

results, and a well-established pattern for internal and external communication (a Communication manager in the PMT) are some of the success factors worth focussing on. An example of a secured service is the framework for governing the provision and management of files to be used for testing the prototypes, described in the internal document *Data Management Plan for training, testing and demonstration files in the PREFORMA project* (DMP) developed by the PREFORMA Consortium in cooperation with the suppliers.

On an **operational level**, the PREFORMA Consortium continously followed the supplier's struggle to develop their conformance checkers in accordance with the PREFORMA requirements. The Consortium was also on a high state of readiness for handling unexpected challanges. Formally, the PREFORMA team T6.1 and T6.2 respectively had the role of coordinating the work of the suppliers. In practice, much of the coordination was conducted in close relationship with the PREFORMA PMT, which was an effective way to handle up-coming issues of importance and sometimes urgent nature.

To ensure an efficient management of the PCP, an Evaluation Secretariat was established, composed by Claudio Prandoni (PROMOTER) and Bert Lemmens (PACKED). Their task was to:

- supervise and orchestrate the work of the PREFORMA Evaluation Committee for the evaluation of the results of the Prototyping phase;

- manage the communication with the suppliers during all the phases of the PCP;

- supervise the periodic meetings organised between the suppliers and the PREFORMA Consortium, including writing the minutes of each meeting.

When needed, specific task forces were set up, for example to prepare for the Testing and validation phase (by identifying "classes" for each media type targeted by PREFORMA). Another example was the setting up of the task force organised to support one of the suppliers who had not managed to match all requirements when the final release and end of phase report were examined by the Evaluation Committee.

Acquisition, storage, and handling of test files, based on the guidelines in the Data Management Plan, were recurring topics in PREFORMA´s internal discussions. To administer the handling of test files in the PREFORMA Vault, and later to guide memory institutions in the use of GiTHub, three dispatchers were appointed by the PREFORMA Consortium. A GitHub guidance document was produced to support the producer of testfiles (see Annex 2). Special contact persons (one per supplier) were also appointed to take part in the work of the Technical Working Group during the 2nd part of the Prototyping phase.

The interaction with the suppliers through monthly meetings, and separate meetings when needed, formal feedback reports, and open issues in the GitHub, was an important framework for the following up and the coordination of the work preformed by the suppliers. This underlying structure was then filled with a huge amount of mail contacts with the suppliers, periodically almost on daily basis.

5.2 COMPLIANCE WITH PREFORMA REQUIREMENTS

As stated in section 1.2 above, the objective of this deliverable is to provide a report on the improved versions of the three first software prototypes from the 1st part of the Prototyping phase. The report shall inform on how the suppliers have:

- provided required functionality;

- adhered to utilising best practices from open source development;

- established a process of feedback to standardisation organisations.

The issue of compliance has of course been of paramount importance for the PREFORMA Consortium when evaluating the supplier's software. The Consortium took a rather pragmatic approach towards compliance with PREFORMA requirements. More "hard-drawn" interpretations of the requirements must conform to the reality of memory institutions and other end users of the PREFORMA outcome. Otherwise there is a risk that PREFORMA will "throw out the baby with the bathwater". This discussion was most animated in relation to the open source approach. The excellent work of the PREFORMA expert partner on open source forced the Consortium to consider all aspects of open source before taking decisions on how toproceed. This will assist the PREFORMA after-life in avoiding dangers and unpleasant challenges that are not determined at present.

The overall impression of the improved versions of software prototypes is, as the Evaluation Committee expressed it in their final report (see Annex 8), that the three supplier's projects have dealt in an all-inclusive manner with the research and development issues addressed in the PREFORMA Challenge Brief.

The procurers as well as the Evaluation Committee appreciate the work done so far by the suppliers, convinced that what the they have achievied at the end of the Prototyping phase will be a good starting point for the succeeding Testing and validation phase. However, it is still room for improvments, particulary in relation to quality and open source aspects, as the strict review by the expert partner on open source has demonstrated.

In the following sub-sections, the compliance with the PREFORMA requirements will be examined with a starting-point in the three check-points for this deliverable, mentioned above: (1) Provide required functionalities, (2) Utilise best practise from open sours development, and (3) Establish a process for feedbacks.

5.2.1   Provide required funtionalities

During the Prototyping phase, the three suppliers awarded with contracts were expected to:

- provide software prototypes that fulfil the requirements of the PREFORMA project;

- demonstrate the results;

- provide explanations and documentation how the developed software can be effectively used in archiving scenarios at memory institutions.

Provide software prototypes

As discussed in section 5.1 above, the requirements of the PREFORMA project on the software prototypes are of different kind:

- Firstly, there are basic requirements, set out in the Call documents (Invitation to Tender, Challenge Brief, and Framework Agreement);

- Secondly, in deliverables D4.3 *Functions of the Open Source Portal* and D8.8 *Monitoring of the Open Source Project implementation* the requirements related to open source issues are further developed; these issues will be discussed in section 5.2.3 below;

- Thirdly, technical and functional specifications for the prototyping were provided by the suppliers as an outcome of the Design phase;

- Forthly, clarifications and explanations were constantly presented by the PREFORMA Consortium during the whole Prototyping phase, mainly based on the feedback process that involved the suppliers, the PREFORMA Consortium, Evaluation Committee, and external testers. These clarifications and expectations were sometimes rather detailed and aimed to influence how software should be developed, distributed and released by the suppliers throughout the Prototyping phase. Specific working groups were also set up by PREFORMA to handle specific tasks

The main observation is that the **functionalities** of the software have been evolving during the whole Prototyping phase, especially between the two intermediate releases. Overall, the Evaluation Committee found that the development of the Conformance Checkers had progressed well and that the software, in most cases, was easy to install, setup and use.

However, a great deal of the feedback on the software releases expressed a desire for more advanced and better performed features or pointed at specific needs not (yet) properly addressed by the suppliers. In that sense, the requirements gradually became more concrete and well adapted to the every-day work of the memory institutions. This continous process of clarifications, explanations, and interpretations of the basic requirements was important for the Evaluation Committee to balance, and the Committee did that by accepting the software prototypes as compliant with the PREFORMA requirements but, at the same time, giving rather detailed recommendations for further development in the Testing and validation phase. The supplier's capacity to adapt to these changes or clarifications in the requirements is a strong aspect of their work. This capacity improved substantially during the 2nd part of the Prototyping phase, as can be seen in their intermediate and final reports (see chapter 4.1 above).

Demonstrate the results

Regarding dissemination and demonstration of the results, the list of activities undertaken by the suppliers to promote and demonstrate their open source project grew substantially as the Prototyping phase progressed During its 1st part, demonstrations were not carried out in any significant way, but during the 2nd part an increasing number of software demonstrations took place, performed at events organised either by the PREFORMA Consortium, or by other projects and organisations, or by the suppliers themselves. Moreover, the Evaluation Committee underlined in its feedback reports on several releases that the suppliers had been very active at conferences and other fora.

New channels for software demonstrations were also introduced during the 2nd part of the Prototyping phase, the main ones being Webinars and training seminars/hands-on workshops for users.

Provide explanations and documentation

The main observation concerning explanations and documentation is that the suppliers have provided final releases that show proof of **usability**, in various degrees. The phrase "usability" is not used in the PREFORMA documents, but was an undercurrent theme in the feedback process during the 2nd part of the Prototyping phase. Guides, manuals and guidelines are crucial for improving the usability by making it easy for users to get started and to handle the software. Therefore, the Evaluation Committee emphsised in its feedback reports the importance of updating the documentation after each release, the installation guide as well as the GUI manual and the CLI manual. The Committee also required that explanations and documentation should be updated to an acceptable level before the Committee could approve the supplier's final releases.

5.2.2 Utilise best practices from open source development

Suppliers are expected to utilise best practices from open source development, which include the use of:

- an open source work practice for development;
- frequent open source releases;
- promotion activities aimed towards a sustainable community.

The PREFORMA web site contains a dedicated section (Open Source Portal, OSP) with references to each of the three open source projects in PREFORMA. The implementation of an open source approach in PREFORMA is specified in deliverable 4.3 D4.3 *Functions of the Open Source Portal*. This deliverable describes the functions of this OSP and clarifies the requirements for the open source projects in terms of executables, source code, and build environments that the suppliers need to make available on their specific development platforms. The PREFORMA expert partner on open source, University of Skovde, have followed up how the suppliers have progressed, and provided recommendations for further actions in two releases of deliverable D8.8 *Monitoring of the Open Source Project implementation.* These recommendations were carefully considered by the Evauation Committee before accepting the supplier's releases (with complementary recommendations for additional actions).

Use of an open source practice

The use of an open source practice involves adherence to established open source community norms and values to maximise transparency and acceptance amongst the broader community of volunteers and open source companies. It also entails publication of a roadmap and use of wikis, forums, issue trackers, software configuration management systems etc, to promote the open collaborative development process. A collaborative platform for open source software development (e.g. GitHub or equivalent) should also be used to support an open work practice.

Overall, the suppliers have showed adherence to established open source community norms and values as required. There are several examples in the intermediate and final reports describing ways in which they have tried to fulfil the PREFORMA requirements on open source.

PREFORMA requires that a supplier provides an up-to-date roadmap for the different versions of the software. During the reviews of the intermediate and final reports it was noted that details were lacking, especially on external contributions and development beyond the PREFORMA project. The feedback reports were very clear on what was required, but nevertheless the suppliers could not fully satisfy the Evaluation Committee in this regard. In the end, the Committee decided to accept their results, but they all received a recommendation to address this issue in the following tesing and validation phase.

Concerning a collaborative platform, all suppliers have actively been working with, and responding to user-requests on GitHub.

Use of frequent open source releases

The requirements set up for this activity were clearly defined from the start. Early and frequent open source releases must be used, and the developed software shall be provided from the very beginning of the development, with evolving functionality over time. Copyleft licenses shall be used for all developed software and associated digital artefacts. All developed software has to be provided under both Mozilla Public License (MPL) v2.0 or later and under GNU General

Public licence 3.0 (GPL v3) or later, and all associated digital artefacts (e.g. instructions, manuals, documentations, test cases, etc.) developed during the project have to be provided under the Creative Commons (CC) license Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0).

In deliverable D8.3 *First Prototype Report*, it was noted that the suppliers, during the 1st part of the Prototyping phase, had built substantial capacities for software releases in order to accommodate the requirement of PREFORMA for frequent, monthly releases. The information given did merit that the requirement on frequent releases was given the status "checked".

In the 2nd part of the Prototyping phase, the requirement on software releases, monthly and formal ones, was clearly announced in the work plan, agreed in the beginning of the 2nd part (see Annex 1). However, in the reviews of the formal releases, the PREFORMA expert partner on open source noted that not all suppliers have managed to fulfil the PREFORMA requirement for time based (monthly) stable releases. The Evaluation Committee recognised, however, that these gaps were minor kind and could be accepted. In the recommendations following the decision to accept the supplier's final releases, the Committee underlined that attention must be paid to the open source practices, including frequent (monthly) releases, which should be uploaded regularly on the PREFORMA server.

An issue discussed already in the Design phase, and raised again during the Prototyping phase, was which part of the software that must be provided under MPL3+ and GPL2+ licenses (i.e. the Implementation checker and the Policy checker), and which part can include third party libraries which are released under a generally recognised open source license compatible with MPL3+ and GPL2+ (i.e. Shell, Reporter and Metadata fixer). The PREFORMA Consortium sent a clarification letter to the suppliers during the Design phase, and at the plenary meeting in Berlin it was decided that this letter was still applicable, i.e. that the core part of the software should conform to the PREFORMA licensing requirements. To get their releases approved as final, the suppliers were requested to provide a "Core Distribution Package" defined by the PREFORMA Consortium.

Promotion activities aiming at a sustainable community

Promotion activities include participation in community events in order to network with other open source developers, but it also includes the provision of illustrative examples (source code, binaries, test files, screenshots, etc.) in order to demonstrate how developed software can be used.

All three suppliers made great efforts in promotion activities, which have been reported in their intermediate and final reports from the Prototyping phase and in their End of phase reports. Some examples to be mentioned are the following ones:

veraPDF has targeted both organisations that archive PDF/A files in accordance with institutional practice and those that are archiving PDF/A and other PDF files at scale. These two organisations were approached and they also received feedback from seven, ranging from national archives to university libraries. The veraPDF project has also established a dialog between the PDF industry and the cultural heritage/digital preservation communities.

EasyInnova reported that their most intensive testing came from Basel University in collaboration with the Swiss Federal Archive and its Coordination Committee for sustainable archiving of digital documents (KOST). A TI/A community had been built up around three online channels: the TI/A website, the TI/A twitter account, and the TI/A Intranet for the involved experts. EasyInnova could also count on a valuable collaboration with the Swiss Coordination

Centre for the Long-Term Preservation of Electronic Documents (KOST-CECO) and some of the bigger Swiss Archives.

MediaArea highlighted among other activities their cooperation with the CELLAR Working Group. The formation of CELLAR within IETF had brought together a strong community of open source developers and preservationists interested in the development of standards. The Evaluation Committee treasured the engagement of the CELLAR team in clarifying the existing varieties and working in tandem with market players (e.g. Google) to keep varieties in check. The fruitful collaboration with Artefactual and the Tate Museums, reported by MediaArea, is also a strong point, and especially the integration of MediaConch into Archivematica. This was highlighted by the Evauation Committee as a very important result.

In all, the Evaluation Committee found the suppliers results in this field of activities to be satisfactory and in line with the PREFORMA requirements. However, one of the suppliers had difficulties establishing a community of users actively engaged in developing and testing. This resulted in a recommendation for future work from the Committee.

5.2.3   Establish a process of feedback

Suppliers are expected to establish a process of feedback with the relevant standardisation organisations and other relevant stakeholder groups (e.g. legislators and other suppliers). This feedback process is critical for improving interoperability and long-term preservation of files, and creates a basis for continuous improvement of developed software in the open source project.

The standardisation efforts made by the suppliers in the Prototyping phase have been satisfactory and also progressed very well, according to the Evaluation Committee.

veraPDF has a successful engagement with the ISO community responsible for the development of both PDF/A and PDF (ISO 32000) specifications. This has resulted in a significant influence on the future development of both standards, especially through the suppliers PDF/-Next initiative.

EasyInnova has made good progress with their TI/A initiative; they have addressed issues with Adobe and will follow Adobe's advice to work on an ISO recommendation instead of a new file format. A draft of this recommendation has also been brought to a pre-final stage, with a final draft submitted to the ISO TC171 6 to start the official process of standardisation.

MediaArea has made a good job with the IETF standardisation process, which includes standard drafts for both Matroska and FFV1, based on technical reviews of as well as feedback on existing specification papers, and an open dialogue with users and developers.

ANNEX 1: WORK PLAN FOR THE SECOND PART OF THE PROTOTYPING PHASE AGREED AT THE FIRST SUPPLIER MEETING DURING THE 2ND PART OF THE PROTOTYPING PHASE

MEETINGS

1. Monthly web meetings

These meetings have a triple focus:

- To ensure that all relevant information will be provided to suppliers and Consortium members alike;

- To provide a transparent forum for discussion of issues brought up by various parties (whether technical, organisational or regulatory);

- To discuss the results of the software releases in a way that would engage both the PREFORMA Consortium and the suppliers in a dialogue about the releases, their usability and adherence to the overall goals and requirements of the PREFORMA project

2. Individual meetings with suppliers as part of the evaluation process (feed back, see below)

3. Meetings on specific issues

Specific technical issues will be discussed more in depth with the suppliers in a special Technical Working Group led by RA. This group will also cover interoperability issues which during the 1st part of the Prototyping phase were discussed in an Interoperability Working Group. The outcomes from the Technical Working Group will be reported and commented at monthly web meetings.

The Technical Working Group represents a good communication channel between the PREFORMA Consortium and the suppliers in technical issues, but also in other issues like policies, etc. Therefore, a contact person from the PREFORMA will be appointed for each supplier.

REPORTING

1. Software releases

- Monthly stable releases

- Two formal releases based on the functional and technical specifications from the previous Design phase: one *Intermediate release* in July 2016 and *one Final release* in the end of October 2016.

2. Documentation

Two complementary reports were requested together with the formal releases, one with the Intermediate release and one with the Final release.

3. Formal reports

Each supplier had to provide a formal End of phase report. These reports should contain the basic data for deciding if the results of the Prototyping phase are good enough for inviting the

suppliers to take part in the call for tenders for the Testing and evaluation phase. In order to start the Testing and evaluation phase as soon as possible in 2017; these End of phase reports need to be submitted in mid November 2016.

FEEDBACK

1. Daily feedback

Memory institutions will be recommended to test the prototypes as soon as a new release is available and to report upcoming issues directly to the suppliers using GitHub. The suppliers will then be asked for help in summarising the activity in GitHub.

2. Feedback on Intermediate releases

Feedback reports on Intermediate software releases and documentations will be provided by the PREFORMA Consortium to the suppliers and further discussed at separate meetings.

3. Feedback on final releases

After the completion of the Final software releases and the complementary reports, an informal evaluation will take place with the purpose to give the Consortium a first indication of the final outcome of the Prototyping phase. Feedback will be provided to the suppliers and further discussed with them in separate meetings.

CALL FOR TENDERS AND EVALUATION

A formal call for Tender for the Testing and evaluation phase will be sent out to invited suppliers in the end of November – beginning of December 2016. These tenders will be formally evaluated by PREFORMA Evaluation Committee. Evaluation criteria and methodology was included in a second version of deliverable D8.1 "*Competitive Evaluation Strategy*" (submitted by the end October 2016).

DEMONSTRATIONS

- Two demonstrations of the prototypes were foreseen as part of the activities contracted for the Prototyping phase: one in month 24 (moved to 7th of April 2016 in Stockholm) and one in month 36 (moved to 23rd of November 2016 in Berlin).

- One demonstration of the final software is foreseen in the Description of Work in month 42, as part of the activities contracted for the Testing and evaluation phase. This is planned to take place in Spring 2017 in Padua, jointly with the third review of the project.

- Suppliers were requested to attend and demonstrate their software at the final conference, to be held in in Autumn 2017 (preliminary in Stockholm).

- In addition to these main events, two more things are planned for:

  1. Training events / tutorials organised in the premises of the PREFORMA memory institutions for which time has been set for three such events during Spring 2017;

  2. Webinars, for which three was planned during May to September.

PREPARATION OF THE TESTING AND VALIDATION PHASE

The Cranfield paradigm based on experimental test collections has been adopted by the Consortium for testing and validation. An experimental test collection consists of a triple (documents, classes, ground-truth), where documents are real or synthetic documents for the

various format from memory institutions, suppliers, and the open source PREFORMA community. Classes are defined by the PREFORMA Consortium to identify the different conformance checks that are expected; ground-truth determines which documents belong to which classes.

Preliminary work will start in April, e.g. preparing excel sheets for the identification of classes and ground truth, preparing instructions, establish task forces, etc. The suppliers will be asked to identify experts and also people from the standardisation organisations to help in the definition of the classes.

UPDATING THE WORK PLAN

During autumn 2016 it became more and more obvious that the schedule of WP6 would become very tight in the end of 2016. Therefore, a more formalised "roadmap", agreed with the suppliers, is needed for the last part of the Prototyping phase (see Annex 7). This "roadmap" has, then, to be checked and (when needed) updated at the monthly web meetings with suppliers.

ANNEX 2: THE GITHUB GUIDANCE DOCUMENT

Today, open source projects has become a vital part of the software industry, and magazines and others that monitor developments issue yearly awards of top open source projects. It has become, according to those who work in it, become a "world". Open source has several connotations. One is that the source code is available on-line for anyone to see, and contribute too. In fact, some of the major platforms that we have identified in the report are open source projects themselves (an example is Jenkins[23]). Another aspect of open source is specific open source licenses which "allow software to be freely used, modified, and shared".[24]

With respect to follow up, open source projects are available "in the open". But how do we actually follow the PREFORMA projects of the veraPDF consortium, MediaConch and DPF Manager?

One key is to follow the continuous integration pathways and links that are available on GitHub. GitHub provides you with the opportunity to create an account free of charge, which makes it a little easier to search for the three projects.[25] An alternative is to use the links to each of the three projects that the suppliers have provided in the intermediate and final reports.[26] This provides an overview of the various repositories of each of the three projects.

veraPDF consortium

The main repository of the veraPDF consortium is the veraPDF library.[27] The README-file provides basic information about licensing, how you can get the veraPDF software, how to build it from source, and the CI-status of the various subprojects. Here, you are being acquainted with the continuous integration tools which the consortium uses: Travis-CI, and Jenkins. Both of these two platforms are linked to GitHub.

The CI-status is clickable, providing direct links to ongoing developments within the project.

By clicking on Travis-CI[28], you are provided with a pathway to the current build[29] jobs, branches that have been set up, including the master branch, the build history, as well as pull requests meaning contributions to the software (provided mostly by the members of the consortium).[30] At

---

[23] https://github.com/jenkinsci/jenkins

[24] https://opensource.org/licenses

[25] https://github.com/

[26]   https://github.com/veraPDF,   https://github.com/MediaArea/MediaConch, https://github.com/EasyinnovaSL/DPFManager

[27] https://github.com/veraPDF/veraPDF-library

[28] https://travis-ci.org/veraPDF

[29] https://en.wikipedia.org/wiki/Software_build

[30] https://travis-ci.org/veraPDF/veraPDF-library

the moment of writing this text, the consortium had set up branches like "default or master branch", "active branches", and "inactive branches".[31] If you click on build history, a list of these is produced.[32]

You can also take a glance of the GitHub-activities of the open source project by clicking on the GitHub-icon on top.

Now, let's go back to the GitHub README and discover which information that is available by clicking on the Jenkins tab under "CI-status". Here, you enter a webpage which at centre lists recent changes, latest test results and last successful deployed artefacts. Two links to Jenkins activities are set up by the consortium, one concerning the 0.8 release while the other displays activities having to with the 0.9 release.

Now, the 0.8 release is deployed at the open source portal (release date 11 December 2015), and brief information about this particular release can be utilised by the release notes.[33] At the Jenkins platform you are provided with full information however.[34] You can also jot back to the GitHub page by clicking on the GitHub icon in the left column (status, changes, modules, GitHub, embeddable build status, Git Hook Log).

If you are interested in test results, click on "latest test result".[35] Here, you are provided with information about core tests, tests which concern the PDF feature report, GUI, legacy types, metadata fixer and so on. The implementation checker module has been tested a vast majority of times, thus pinpointing the PREFORMA Consortium's priorities in the 1st prototyping phase.

For a full overview of the Jenkins related activities of the consortium and the people that are involved just click on "back to dashboard". [36]The status of the activities is marked by colours such as blue and red. Some information about the persons who have contributed is available as well through the "people" column.

MediaArea

The MediaConch project is written in C++, while the veraPDF and DPF Manager are written in Java. The Jenkins and Travis-CI platforms are well fitted (as is the Maven) for Java projects but there are possibilities to use plug-ins for C++ with Jenkins. MediaArea uses Travis-CI for build related activities.

---

[31] https://travis-ci.org/veraPDF/veraPDF-library/branches

[32] https://travis-ci.org/veraPDF/veraPDF-library/builds

[33] https://github.com/veraPDF/veraPDF-library/blob/master/RELEASENOTES.md

[34] http://jenkins.opf-labs.org/job/veraPDF-library-0.8/

[35] http://jenkins.opf-labs.org/job/veraPDF-library-0.8/lastBuild/testReport/

[36] http://jenkins.opf-labs.org/

We will follow the same pathway, going back to the README – but this time under the source repository of GitHub, which belongs to MediaArea.[37] The MediaConch source code readme contains a link to the Travis-CI under the headline "How to Build".

By clicking on the link, you enter the MediaConch build activities on the Travis-CI.[38] Here, you can get information about current jobs, branches, build history and pull requests (meaning contributions to the source code made either by team members, or contributors outside of the project). If you want to go back and have a look into the GitHub repositories, just click on the GitHub icon.

On the Travis-CI, you can also follow ongoing activities regarding the dependencies of the MediaConch project, most notably MediaInfo.[39]

EasyInnova

EasyInnova has three projects on GitHub: DPF Manager, TIFF Library and easyTIFF. The README belonging to the DPF Manager provides a link to Travis-CI[40], while the TIFF library provides access to the Maven repository of the project[41], as well as to its particular Travis-CI activities.[42]

At the DPF Manager and TIFF Library Travis-CI, you obviously find content organised under the same columns as in the other two projects (current builds, branches, build history and pull requests). By downloading the log under the current column you find details such as tests.[43] The inter-relationship between Travis-CI and GitHub is made obvious by clicking on the GitHub icon on top. This brings you back to either the DPF Manager project[44], or to the TIFF Library project.[45]

---

[37]   https://github.com/MediaArea/MediaConch_SourceCode/blob/master/README.md#how-to-build-mediaconch

[38] https://travis-ci.org/MediaArea/MediaConch_SourceCode

[39] https://travis-ci.org/MediaArea

[40] https://github.com/EasyinnovaSL/DPFManager/blob/develop/README.md#ci-status

[41] http://mvnrepository.com/artifact/com.easyinnova/tifflibrary4java

[42] https://travis-ci.org/EasyinnovaSL/Tiff-Library-4J

[43] https://s3.amazonaws.com/archive.travis-ci.org/jobs/98642242/log.txt (The downloading log feature is of course available at the Travis-CI sites of the other two open source projects.

[44] https://github.com/EasyinnovaSL/DPFManager

[45] https://github.com/EasyinnovaSL/Tiff-Library-4J

Lastly, the information provided at the Maven[46] repository regards the TIFF library.[47] The columns found here are: versions, usages, type and date. Under the "version" column you find facts concerning the artefact itself, the file, date and homepage.[48]

---

[46] https://maven.apache.org/index.html

[47] http://mvnrepository.com/artifact/com.easyinnova/tifflibrary4java

[48] https://github.com/EasyinnovaSL/Tiff-Library-4J

ANNEX 3: TEMPLATE FOR THE INTERMEDIATE REPORTS

PROTOTYPING PHASE 2 INTERMEDIATE REPORT

Project Acronym: **PREFORMA**

Grant Agreement number: **619568**

Project Title: **PREservation FORMATs for culture information/e-archives**

Name of the supplier's project

Revision: [draft, final]

Authors:

> **Name (Organisation)**
> **Name (Organisation)**
> **……**

Dissemination Level		
P	Public	X

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, if possible even more organised release compared with the respective predecessor versions
- A report which
  - Describes
    - More in 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 which 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.

PROTOTYPING PHASE 2 - INTERMEDIATE REPORT

1. Details
Type of Organisation:
Registered Name of Organisation:
Registered Address:
Town/ City:
Postcode:
County:
Country:
Report Author:
Telephone Number:
E-mail Address:
Project Name:
Report Type: Prototyping Phase 1 – Final Report
Total Contract Price [euro]:
Start Date:
End Date:
Sub-contractors:

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

Please provide the PREFORMA consortium with a concise overview of the releases developed so far, and of the functionalities that are available at the time of this report.

Please highlight

- *which is the progress compared to the October 2015 release (final release of the first prototyping phase)*
- *how are you addressing the comments received from the PREFORMA consortium*
- *which are your plans how to progress further.*

Feel free to refer to any other document you provided so far, when appropriate, by providing the link.

2. Testing

Please provide the PREFORMA consortium with a detailed description of the datasets that have been used to test the release (own, memory institutions, external, etc.), and the respective purpose of testing.

3. Dissemination and community building

Please provide the PREFORMA consortium with the list of dissemination activities that you have undertaken to promote your open source project (webpages, blogs, newsletters, press releases, papers, presentations, etc.).

Please describe any potential long-term collaborations/partnerships entered into, by listing the organisation/s and the role they played in the project.

4. Open Source approach

Please provide the PREFORMA consortium with a description of how you addressed the relevant open source topics, best practices, and licensing

How did you progress in setting up an open source community around the developed tools?

5. Standardisation efforts

Please provide the PREFORMA consortium with a description of how you are actively contributing to the standardisation process in your domain, by means of providing feedback on the existing standards contributing as well as the way on how to support emerging standards.

6. Impact assessment, sustainability, future use and exploitation

Please provide the PREFORMA consortium with a description of your ideas and plans related to the sustainability, future use and exploitation of the results of your project.

Please include some evidence of the impact that the project generated so far for the memory institutions and for any other relevant target group.

7. Gap analysis and next steps

Please provide the PREFORMA consortium with a description of the status of the work compared to what was planned in the functional and technical specification that you provided at the end of the design phase.

Please highlight critically what it is still missing in the current release and which are your plans to overcome the gaps.

Please include also an updated version of your work plan and a timeline, preferably in a graphical way (GANTT) in a way that the PREFOMA consortium members now and later can easily compare the status of fulfilling the requirements of the project as well as the level of compliance to your own technical and functional description.

PREFORMA - Future Memory Standards
PREservation FORMAts for culture information/e-archives
EC Grant agreement no: 619568

ANNEX 4: TEMPLATE FOR THE FINAL REPORTS

PROTOTYPING PHASE 2
FINAL REPORT

Project Acronym: **PREFORMA**

Grant Agreement number: **619568**

Project Title: **PREservation FORMAts for culture information/e-archives**

Name of the supplier's project

Revision: [draft, final]

Authors:

>   **Name (Organisation)**
>   **Name (Organisation)**
>   **……**

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

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.

PROTOTYPING PHASE 2 - FINAL REPORT

1. Details
Type of Organisation:
Registered Name of Organisation:
Registered Address:
Town/ City:
Postcode:
County:
Country:
Report Author:
Telephone Number:
E-mail Address:
Project Name:
Report Type: Prototyping Phase 1 – Final Report
Total Contract Price [euro]:
Start Date:
End Date:
Sub-contractors:

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

Please provide the PREFORMA consortium with a concise overview of the progress since the last intermediate release, and of the functionalities that are available at the time of this report.

Please highlight

- *The progress compared to the July 2016 release (intermediate release of the second prototyping phase)*

- *how you are addressing the comments received from the PREFORMA consortium*

- *what are your plans to progress further.*

2. Datasets used to test the release

Please provide the PREFORMA consortium with a detailed description of the datasets that have been used to test the release (own, memory institutions, external, etc.), and the respective purpose of testing.

3. Dissemination and community building

Please provide the PREFORMA consortium with the list of dissemination activities that you have undertaken to promote your open source project (webpages, blogs, newsletters, press releases, papers, presentations, etc.).

Please describe any potential long-term collaborations/partnerships entered into, by listing the organisation/s and the role they played in the project.

How did you progress in setting up an open source community around the developed tools?

4. Open Source approach

Please provide the PREFORMA consortium with a description of how you addressed the relevant open source topics, best practices, and licensing issues identified in the report of the University of Skövde.

5. Standardisation efforts

Please provide the PREFORMA consortium with a description of how you are actively contributing to the standardisation process in your domain, by means of providing feedback on existing standards as well as supporting emerging standards.

6. Impact assessment, sustainability, future use and exploitation

Please provide the PREFORMA consortium with a description of your ideas and plans related to the sustainability, future use and exploitation of the results of your project.

Please include some evidence of the impact that the project has generated to date for the memory institutions and for any other relevant target groups.

7. Gap analysis and next steps

Please provide the PREFORMA consortium with a description of the status of the work compared to what was planned in the functional and technical specification that you provided at the end of the design phase.

Please highlight critically what it is still missing in the current release and what are your plans to overcome the gaps.

Please include also an updated version of your work plan and a timeline, preferably in a graphical way (GANTT) so that the PREFOMA consortium members now and later can easily compare the status of fulfilling the requirements of the project as well as the level of compliance to your own technical and functional description.

ANNEX 5: TEMPLATE FOR THE END OF PHASE REPORTS

END OF PHASE 2 REPORT

Project Acronym: **PREFORMA**

Grant Agreement number: **619568**

Project Title: **PREservation FORMAts for culture information/e-archives**

Name of the supplier's project

Revision: [draft, final]

Authors:

Name (Organisation)
Name (Organisation)
……

Dissemination Level		
P	Public	X

INTRODUCTION

The purpose of the end of phase report is to ensure that contractors performed the procured R&D services as specified in the framework agreement.

This report must be submitted within 14 days from the completion / termination of the phase. Satisfactory completion of this report forms part of the contract.

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

- Börje Justrell: borje.justrell@riksarkivet.se
- Claudio Prandoni: prandoni@promoter.it

The objectives of the report are:

- To provide an overview of the work done in order to measure the results against the objectives and the requirements included in the PREFORMA Challenge Brief.
- To provide a comprehensive report to be shared with stakeholders to facilitate further commercialisation of the product.

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 comprehensively, but keep your answers succinct and no longer than necessary to clearly motivate them. When describing technical solutions, please regard your audience as being someone familiar with the technology, but not an expert. The report may be filled in using text only. However, diagrams or pictures may be added if appropriate within the restriction on the page limit of a total of 20 A4 pages.

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.

END OF PHASE 1 REPORT

1. Details
Type of Organisation:
Registered Name of Organisation:
Registered Address:
Town/ City:
Postcode:
County:
Country:
Report Author:
Telephone Number:
E-mail Address:
Project Name:
Report Type: End of Phase 2 Report
Total Contract Price [euro]:
Start Date:
End Date:
Sub-contractors:

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

Please provide a concise overview of the project objectives and of what was required to be undertaken during the prototyping phase as agreed at the end of the first design phase.

[limit: 1 page]

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

Please provide the PREFORMA consortium with a concise overview of the progress of the work required to be undertaken in the prototyping phase, relating it to the original objectives and to the requirements defined in the PREFORMA Challenge Brief and in the Technical and Functional Specifications submitted at the end of the first design phase.

[limit: 2 pages]

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.

If applicable, explain the reasons for any deviation from what was agreed at the end of the first design phase, or for failing to have achieved critical objectives, and the impact of such deviations on the project. If applicable, indicate the corrective actions that you put in place or that you planned for the testing phase, in case you will be awarded a contract.

[limit: 1 page]

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

Please provide the PREFORMA consortium with a concise overview of the main results achieved so far. Please refer to the release report where appropriate without repeating too much here.

[limit: 2 pages]

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

[limit: 1 page]

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

[limit: 1 page]

8. Please provide complete and clear information about the allocation of the money paid by the Authority taking into account 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).

[limit: 1 page]

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

[limit: 1 page]

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

[limit: 1 page]

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

In this sub-section, the supplier is expected to describe possible business models, business plans, and business cases based on use cases or scenarios relevant for the project. The business plan should not only cover the PREFORMA phases to come but may also give an indication on how exploitation could look like after the end of the PREFORMA project.

[limit: 2 pages]

12. Please add any additional information that you consider relevant to be reported. This may be in the form of text, pictures, diagrams, data, graphs that support the work done.

[limit: 2 pages]

13. Describe what ethical aspects you have identified and how this may influence your project.

[limit: 1 page]

FINANCIAL REPORT

	Unit price	Quantity	Quoted price (€)	Total Price (€)
Labour Price				
1.				
2.				
3.				
4.				
5.				
6.				
Materials				
Capital Equipment				
Sub Contract				
Travel and accommodation				
Other (specify)				
TOTAL PRICE (excluding VAT)				
TOTAL PRICE (including VAT)*				

ANNEX 6: TEMPLATES FOR THE EVALUATION

FEEDBACK ON THE INTERMEDIATE RELEASE - JULY 2016

<The software releases are available in the Open Source Portal (http://www.preforma-project.eu/open-source-portal.html).>

<The intermediate reports are available in the project's repository under the *Tender/Second prototyping phase/Reports/* folder.>

<If you notice something wrong while using the conformance checkers, we recommend you to report it using the issue trackers that the suppliers set up on GitHub and refer to the specific issues in the evaluation report:

- veraPDF: https://github.com/verapdf
- DPF Manager: https://github.com/EasyinnovaSL/DPFManager
- MediaConch: https://github.com/MediaArea/MediaConch>

1. GENERAL COMMENTS

<If you cannot find an appropriate category below>

2. RESULT FROM EXAMINATION

<questions concerning how the conformance checker validates, e.g., how did file x y z validate, good, poorly, bad, etc.>

3. THE CONFORMANCE CHECKER

<questions concerning the conformance checker itself, e.g., setup, usage, bugs, interface issues etc.)

4. INTERMEDIATE REPORT

<comments in relation to the subjects in the intermediate report>

DESCRIPTION OF THE RELEASE AND PROGRESS COMPARED TO THE LAST INTERMEDIATE RELEASE

<Suppliers shall provide the PREFORMA consortium with a concise overview of the releases developed so far, and of the functionalities that are available at the time of this report. Furthermore, suppliers shall highlight which is the progress compared to the October 2015 release (final release of the first prototyping phase), how they are addressing the comments received from the PREFORMA consortium, which are their plans how to progress further.>

TESTING

<Suppliers shall provide the PREFORMA consortium with a detailed description of the datasets that have been used to test the release (own, memory institutions, external, etc.), and the respective purpose of testing.>

DISSEMINATION AND COMMUNITY BUILDING

<Suppliers shall provide the PREFORMA consortium with the list of dissemination activities that they have undertaken to promote their open source project (webpages, blogs, newsletters, press releases, papers, presentations, etc.).

Furthermore, suppliers shall describe any potential long-term collaborations/partnerships entered into, by listing the organisation/s and the role they played in the project >

OPEN SOURCE APPROACH

<Suppliers shall provide the PREFORMA consortium with a description of how they addressed the relevant open source topics, best practices, and licensing and how did they progress in setting up an open source community around the developed tools.>

STANDARDISATION EFFORTS

<Suppliers shall provide the PREFORMA consortium with a description of how they are actively contributing to the standardisation process in their domain, by means of providing feedback on the existing standards contributing as well as the way on how to support emerging standards.>

IMPACT ASSESSMENT, SUSTAINABILITY, FUTURE USE AND EXPLOITATION

<Suppliers shall provide the PREFORMA consortium with a description of their ideas and plans related to the sustainability, future use and exploitation of the results of their project.

Furthermore, suppliers shall include some evidence of the impact that the project generated so far for the memory institutions and for any other relevant target group.>

GAP ANALYSIS AND NEXT STEPS

<Suppliers shall provide the PREFORMA consortium with a description of the status of the work compared to what was planned in the functional and technical specification that they provided at the end of the design phase, highlighting critically what it is still missing in the current release and which are their plans to overcome the gaps.>

PREFORMA - Future Memory Standards
PREservation FORMAts for culture information/e-archives
EC Grant agreement no: 619568

EVALUATION OF THE FINAL RELEASE - OCTOBER 2016

Supplier Name	
Reviewer Name	
Reviewer Type	<Technical | Domain | External> Expert

Category	Category Weght	Item		Item Weight	Score [1-5]	Notes
C_1 Impact on the Challenge	35%					
		$I_{1,1}$	Basic Research Questions	10%		
		$I_{1,2}$	Conformance Checker	20%		
		$I_{1,3}$	Reference Implementation	20%		
		$I_{1,4}$	Future Proof	5%		
		$I_{1,5}$	Commercial Feasibility	15%		
		$I_{1,6}$	Open Source Work Practices	5%		
		$I_{1,7}$	Delivery and Installation	5%		
		$I_{1,8}$	Open Source Interaction Practice	5%		
		$I_{1,9}$	Open Source IPR Distribution	5%		
		$I_{1,10}$	Negotiation Protocol	10%		
			Category Total		0,00	
C_2 Technical Approach	35%					
		$I_{2,1}$	Architecture	25%		
		$I_{2,2}$	Performance and Quality	10%		
		$I_{2,3}$	Shell Services and Features	25%		
		$I_{2,4}$	Implementation Checker Services and Features	10%		
		$I_{2,5}$	Policy Checker Services and Features	5%		
		$I_{2,6}$	Reporter Services and Features	10%		
		$I_{2,7}$	Metadata Fixer Service and Features	5%		
		$I_{2,8}$	Proposed Approach for Phase 3	10%		
			Category Total		0,00	
C_3 Quality of the Tender	15%					
		$I_{3,1}$	Project Plan	25%		
		$I_{3,2}$	Management Effectiveness	25%		
		$I_{3,3}$	Resource Allocation	25%		
		$I_{3,4}$	Risk Assessment / Risk Factors	25%		
			Category Total		0,00	
C_4 Costs	15%					
		$I_{4,1}$	Price / Cost	100%		
			Category Total		0,00	
			Supplier Grand Total		0,00	

ANNEX 7: ROADMAP FOR THE SECOND PART OF THE PROTOTYPING PHASE

Daeadline	Task	Responsible
15 July	Template for intermediate release report	Borje + Claudio
31 July	Submission of intermediate release report	Suppliers
2 September	Internal feedback on intermediate release	Evaluation Committee
9 September	Feedback sent to suppliers	Claudio
1-30 September	Webinars, one per media type	Becky (veraPDF)
30 September	First list of classes	Nicola
5 October	IPRES workshop, Bern	Borje + Erwin + Suppliers
14 October	Template for final release report	Borje + Claudio
14 October	Template for End of Phase Report	Borje + Claudio
18-19 October	Eafip Athens	Anna Kasimati
31 October	Submission of final release report	Suppliers
31 October	Submission of D8.1	Nicola
4 November	Kick-off of the evaluation • Evaluation report template / spreadsheet • Assignment of experts (minimum 3) to each proposal • Briefing the Evaluation Committee on the evaluation procedure and deadlines	Nicola + Claudio
7 November	Submission of End of Phase Report	Suppliers
8 November	EVA/Minerva workshop, Jerusalem	Borje + Bert
11 November	Admission to the call for tender for Phase 3	Borje
11 November	Tender Form template	Borje + Claudio + Per
14 November	Launch of the call for tender	Claudio
16 November	Image & Research Conference workshop, Girona	Antonella + Magnus + Peter + Sònia + EasyInnova
18 November	First individual evaluation reports submitted by experts	Evaluation Committee

23 November	Experience Workshop, Berlin	All
24 November	Plenary meeting and discussion among experts	PREFORMA partners
30 November	Revised evaluation reports submitted by experts highlighting issues to be solved	Evaluation Committee
2 December	Preliminary evaluation reports circulated to the suppliers asking them to fix all open issues in the December release	Claudio
4 December	Proposals submission	Suppliers
16 December	Updated evaluation reports taking into account the new proposals	Evaluation Committee
31 December	Release of the updated version of the conformance checkers	Suppliers
31 December	Final list of classes	Nicola
31 December	Update of D8.1	Nicola
31 December	Submission of D3.5, D3.6, D3.7.1, D4.6, D8.5, D8.8	Stefan + Erwin + Claudio + Borje + Bjorn
13 January	Final evaluation reports taking into account the latest releases	Evaluation Committee
16 January	Consensus meeting	Evaluation Committee
20 January	Final evaluation report submitted to the suppliers	Claudio
25 January	Negotiation completed (if needed)	Borje
27 January	Award decision	Borje
31 January	Signature of the contracts for Phase 3	Borje
1 February	Kick-off Phase 3	Nicola

ANNEX 8: FINAL EVALUATION REPORTS (CONFIDENTIAL)

ARTEFACT DOCUMENT

Project Acronym: **PREFORMA**

Grant Agreement number: **619568**

Project Title: **PREservation FORMAts for culture information/e-archives**

Consolidated Evaluation Report

Revision: ver 0.1

Date: **06 February 2017**

Proposal: **veraPDF**

PREFORMA - Future Memory Standards
PREservation FORMATs for culture information/e-archives
EC Grant agreement no: 619568

Criterion	Weight	Overall Score
Impact on the Challenge	35%	4,39
Basic Research Questions	10%	5,00
Conformance Checker	20%	4,32
Reference Implementation	20%	4,17
Future Proof	5%	4,13
Commercial Feasibility	15%	4,13
Open Source Work Practices	5%	4,73
Delivery and Installation	5%	4,43
Open Source Interaction Practice	5%	4,72
Open Source IPR Distribution	5%	4,45
Negotiation Protocol	10%	4,47
Technical Approach	35%	4,26
Architecture	25%	4,25
Performance and Quality	10%	3,92
Shell Services and Features	25%	4,10
Implementation Checker Services and Features	10%	4,62
Policy Checker Services and Features	5%	3,93
Reporter Services and Features	10%	4,42
Metadata Fixer Service and Features	5%	4,42
Proposed Approach for Phase 3	10%	4,62
Quality of the Tender	15%	4,38
Project Plan	25%	4,77
Management Effectiveness	25%	4,23

Resource Allocation	*25%*	4,12
Risk Assessment / Risk Factors	*25%*	4,42
Cost	*15%*	**4,20**
Price / Cost	100%	4,20
TOTAL		**4,32**

General remarks and recommendations

The Evaluation Committee considers veraPDF a project that deals comprehensively with the R&D issues addressed in the PREFORMA Challenge Brief.

The team is actively working with, and responding to user-requests on Github.

Good progress has been made with the PDF/A Next initiative.

Here below are listed the main recommendations concerning the future work. The Evaluation Committee expects them to be addressed during the next phase:

- Make sure to update all documentation after each release: installation guide, the GUI manual and the CLI manual.

- User Interface has room for improvement. Non-technical users are experiencing problems in understanding how to use it. The Evaluation Committee recommends investing on the improvement of the GUI during the next phase.

- Include the possibility to create a policy file through the GUI.

- Reviewers experienced different results in terms of performance and stability. Validation seems to take a lot of time or to freeze in certain cases, particularly when trying to validate files of big dimension (100 Mb). Please make sure to take this into account in the future releases.

- Pay attention to respect the open source practices with nightly builds and monthly releases. Monthly releases need to be uploaded regularly also on the PREFORMA server.

- Finalise the work on interoperability in order to be possible to integrate/plug other conformance checkers into veraPDF.

- Provide core distribution package releases together with the "standard" releases until the end of Phase 3. The CDP must only contain the source code for the Implementation and Policy Checker (no code for e.g. the GUI, plugins, or "optional features") and all source code files in the CDP must be licensed as "GPLv3 or later and MPLv2 or later" (no exceptions).

- Include OpenJDK and not OracleJDK in the next core distribution package releases.

- In large scale environments, a proper handling of the report output data (distributed, identification of reports, link to files) becomes more important. Although this cannot be done by the tool itself, use-cases or best practices should be provided that would help to set up a large scale solution, where the tools itself are components of the whole

system.

- Create and upload on the PREFORMA server an up-to-date roadmap with detailed milestones for different (development version, stable version, and deployed (LTS) version) releases covering the time frame until (at least) 2020, that demonstrates the ambition to achieve a sustainable open source project beyond the end of PREFORMA.

- Although there is demand for paid services around veraPDF, the target audience is still new to Open Source business models and therefore not so much motivated to consider paying for OS. We recommend to start to develop a sustainability and exploitation plan to be released in its final version at the end of the testing phase.

ARTEFACT DOCUMENT

Project Acronym:　　　　**PREFORMA**

Grant Agreement number:　　**619568**

Project Title:　　　　　　**PREservation FORMATs for culture information/e-archives**

Consolidated Evaluation Report

Revision: ver 0.1

Date:　　**06 February 2017**

Proposal:　　**DPF Manager**

PREFORMA - Future Memory Standards
PREservation FORMATs for culture information/e-archives
EC Grant agreement no: 619568

Criterion	Weight	Overall Score
Impact on the Challenge	**35%**	**3,58**
Basic Research Questions	*10%*	4,30
Conformance Checker	*20%*	3,50
Reference Implementation	*20%*	3,40
Future Proof	*5%*	3,13
Commercial Feasibility	*15%*	3,37
Open Source Work Practices	*5%*	3,73
Delivery and Installation	*5%*	4,30
Open Source Interaction Practice	*5%*	3,60
Open Source IPR Distribution	*5%*	3,70
Negotiation Protocol	*10%*	3,40
Technical Approach	**35%**	**4,05**
Architecture	*25%*	4,17
Performance and Quality	*10%*	3,83
Shell Services and Features	*25%*	3,77
Implementation Checker Services and Features	*10%*	4,63
Policy Checker Services and Features	*5%*	4,30
Reporter Services and Features	*10%*	4,00
Metadata Fixer Service and Features	*5%*	3,63
Proposed Approach for Phase 3	*10%*	4,23
Quality of the Tender	**15%**	**3,47**
Project Plan	*25%*	3,47
Management Effectiveness	*25%*	3,58

Resource Allocation	*25%*	3,47
Risk Assessment / Risk Factors	*25%*	3,35
Cost	*15%*	**4,74**
Price / Cost	100%	4,74
TOTAL		**3,90**

General remarks and recommendations

The Evaluation Committee considers DPF Manager a project that deals comprehensively with the R&D issues addressed in the PREFORMA Challenge Brief.

The team is actively working with, and responding to user-requests on Github.

Good progress has been made with the TI/A initiative.

Here below are listed the main recommendations concerning the future work. The Evaluation Committee expects them to be addressed during the next phase:

- There seems to be problems of compatibility with OS versions (32/64 bit) and/or Java versions. Please double check and address this.

- Pay attention to keep separate the functionality of the implementation checker (which checks a file against the standard specifications) and policy checker (which allows a user to add or relax rules) and make sure that the policy checker validates all the rules that have been created by the user.

- Reviewers experienced different results in terms of performance, which seems to be a bit low particularly when checking batch of images with some policy rules. Please make sure to take this into account in the future releases.

- Modularity, stability and scalabity of the software have been improved even if sometimes the software still freezes. Please make sure to take this into account in the future releases.

- These seems to be some difficulties in establishing a community of users who are actively engaged in developing and testing the DPF Manager, in order to get real time feedback on possible issues and bugs. Since it's hard to build an active OS community from scratch (it takes a while to get people seriously involved), in order to save time trying to solve all the bugs, the Evaluation Committee suggests to add much more automated install and runtime testing along with a comprehensive set of test data (for each compliance check an example (good / bad) should be provided).

- User Interface has room for improvement. Non-technical users are experiencing problems in understanding how to use it.

- Pay attention to respect the open source practices with nightly builds and monthly releases. In particular, nightly builds seem not to be available.

- Finalise the work on interoperability in order to be possible to integrate/plug other conformance checkers into DPF Manager.

- Provide core distribution package releases together with the "standard" releases until the end of Phase 3. The CDP must only contain the source code for the Implementation and Policy Checker (no code for e.g. the GUI, plugins, or "optional features") and all source code files in the CDP must be licensed as "GPLv3 or later and MPLv2 or later" (no exceptions).

- Make sure to provide in the next core distribution package release all the required licensing information (GPLv3+ and MPLv2+) as two files in the source root directory (gpl.txt and mpl.txt respectively) and in the header of all the source code files.

- In large scale environments, a proper handling of the report output data (distributed, identification of reports, link to files) becomes more important. Although this cannot be done by the tool itself, use-cases or best practices should be provided that would help to set up a large scale solution, where the tools itself are components of the whole system.

- Create and upload on the PREFORMA server an up-to-date roadmap with detailed milestones for different (development version, stable version, and deployed (LTS) version) releases covering the time frame until (at least) 2020, that demonstrates the ambition to achieve a sustainable open source project beyond the end of PREFORMA.

- Although there is demand for paid services around DPF Manager, the target audience is still new to Open Source business models and therefore not so much motivated to consider paying for OS. We recommend to start to develop a sustainability and exploitation plan to be released in its final version at the end of the testing phase.

ARTEFACT DOCUMENT

Project Acronym: **PREFORMA**

Grant Agreement number: **619568**

Project Title: **PREservation FORMAts for culture information/e-archives**

Consolidated Evaluation Report

Revision: ver 0.1

Date: **06 February 2017**

Proposal: **MediaConch**

PREFORMA - Future Memory Standards
PREservation FORMATs for culture information/e-archives
EC Grant agreement no: 619568

Criterion	Weight	Overall Score
Impact on the Challenge	*35%*	4,55
Basic Research Questions	*10%*	4,80
Conformance Checker	*20%*	4,50
Reference Implementation	*20%*	4,60
Future Proof	*5%*	4,60
Commercial Feasibility	*15%*	4,30
Open Source Work Practices	*5%*	5,00
Delivery and Installation	*5%*	4,80
Open Source Interaction Practice	*5%*	5,00
Open Source IPR Distribution	*5%*	4,50
Negotiation Protocol	*10%*	4,10
Technical Approach	*35%*	4,23
Architecture	*25%*	4,53
Performance and Quality	*10%*	3,85
Shell Services and Features	*25%*	3,68
Implementation Checker Services and Features	*10%*	4,83
Policy Checker Services and Features	*5%*	4,90
Reporter Services and Features	*10%*	4,58
Metadata Fixer Service and Features	*5%*	3,53
Proposed Approach for Phase 3	*10%*	4,25
Quality of the Tender	*15%*	3,95
Project Plan	*25%*	4,18
Management Effectiveness	*25%*	4,21

Resource Allocation	*25%*	4,09
Risk Assessment / Risk Factors	*25%*	3,33
Cost	*15%*	**5,00**
Price / Cost	100%	5,00
TOTAL		**4,41**

General remarks and recommendations

The Evaluation Committee considers MediaConch a project that deals comprehensively with the R&D issues addressed in the PREFORMA Challenge Brief.

The team is actively working with, and responding to user-requests on Github. Nightly builds and monthly releases are available and working.

A good job has been done with IETF standardization: technical review of, and feedback on existing specification papers, open dialogue with users and developers; practical feasibility and usability of implementations are being considered.

Excellent work engaging the CELLAR team in clarifying the existing varieties and working in tandem with market players (e.g. Google) to keep varieties in check.

The integration of MediaConch into Archivematica is another very important result.

The work on the policy checker has been impressive. Additional goodie: option to publish policies online allow easy exchange of policies between institutions/users.

Here below are listed the main recommendations concerning the future work. The Evaluation Committee expects them to be addressed during the next phase:

- The metadata fixer is shortly behind schedule.

- Update and clarify the license file to ensure that all third party libraries - required as well as optional - are declared, such as libpng and sqlite.

- Document how to: verify that the code/library is not present in the core distribution packages (where are files stored or supposed to be stored), enable or disable the optional third party library before compilation, verify if the third party library is enabled or disabled after compilation.

- Create a table where all third party libraries are listed with e.g. the following column layout: library, author, license, GPLv3++/MPLv2++ Compatible (bool), optional (bool), (how to) enable/disable, (how to) verify.

- User Interface has room for improvement. Non-technical users are sometimes experiencing problems in understanding how to use it.

- Ongoing improvements in usability and stability. So far there are still some bugs that affect the reliability of the GUI for everyday work.

- Provide core distribution package releases together with the "standard" releases until the end of Phase 3. The CDP must only contain the source code for the

Implementation and Policy Checker (no code for e.g. the GUI, plugins, or "optional features") and all source code files in the CDP must be licensed as "GPLv3 or later and MPLv2 or later" (no exceptions).

- In large scale environments, a proper handling of the report output data (distributed, identification of reports, link to files) becomes more important. Although this cannot be done by the tool itself, use-cases or best practices should be provided that would help to set up a large scale solution, where the tools itself are components of the whole system.

- Create and upload on the PREFORMA server an up-to-date roadmap with detailed milestones for different (development version, stable version, and deployed (LTS) version) releases covering the time frame until (at least) 2020, that demonstrates the ambition to achieve a sustainable open source project beyond the end of PREFORMA.

- Although there is demand for paid services around MediaConch, the target audience is still new to Open Source business models and therefore not so much motivated to consider paying for OS. We recommend to start to develop a sustainability and exploitation plan to be released in its final version at the end of the testing phase.