DELIVERABLE

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

Deliverable D8.3 First Prototype Report

Revision: final v 1.1

Authors:

  Per Elfner, Riksarkivet
  Borje Justrell, Riksarkivet

Contributors:

  Magnus Geber, Riksarkivet
  Benjamin Yousefi, Riksarkivet

Reviewers:

  Peter Pharow, FRAUNHOFER
  Erwin Verbruggen, Sound and Vision
  Ruth Montague, LGMA

Revision History

Revision	Date	Author	Organisation	Description
0.1	25/11/2015	Borje Justrell	Riksarkivet	Structure of the document
0.2	14/12/2015	Per Elfner, Borje Justrell	Riksarkivet	Restructuring of the documents; new text in all chapters except Summary
0.3	19/12/2015	Per Elfner, Borje Justrell	Riksarkivet	Complementary text in chapter 1 and new text most of the other chapters
0.4	21/12/2015	Per Elfner, Borje Justrell	Riksarkivet	New text in chapter 4 and a new chapter 5 Points on Progress; section 4.6 moved to the new chapter 5; section 2.1.3 has been taken out; changes in the text in chapter one due to changes in other chapters
0.5	28/12/2015	Per Elfner, Borje Justrell	Riksarkivet	Language corrections in chapter 1 - 3; update of the text in section 1.3; section 3.2 re-written; new text in chapter 5
0.6 (1,2,3)	31/12/2015	Per Elfner, Magnus Geber. Börje Justrell, Benjamin Yousefi	Riksarkivet	Language corrections in all chapters; updates in almost all chapters; a new annex 3 added and also a new chapter 6 (Conclusions)
0.7	11/01/2016	Claudio Prandoni	Promoter	Added annex 1 and 2
0.8	28/01/2016	Borje Justrell, Per Elfner	Riksarkivet	Corrections based (1) on comments made by the reviewers and (2) on the outcome of a language check.
1.0	02/02/2016	Claudio Prandoni	Promoter	Final check and submission
1.1	03/02/2016	Claudio Prandoni	Promoter	Small corrections in the footnotes

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This deliverable contains original unpublished work except where clearly indicated otherwise.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 1st part of the Prototyping phase in the PREFORMA project. It consists of four distinguishable parts.

The first part (chapter 1 to 3) gives the overall context, including aims and objectives, documents particularly important for the work preformed, and formal background. It also summarises the discussions with the suppliers and internally in the PREFORMA Consortium, and presents complementary issues in more depth.

The second part (chapter 4) focuses on the results achieved by the suppliers, based on their software releases and submitted reports.

The third part (chapter 5) analyses the progress made by the suppliers, including what is left to be done.

The fourth part (chapter 6) resumes the outcome of the 1st part of the Prototyping phase in the form of six main conclusions and a brief charter of success. The results achieved indicate that the project is moving in the right direction. The interplay between theories, discussions, releases and testing has created an exciting situation for the coming development with huge potential for further interest and involvement as well as enhanced communication between different stakeholders.

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

1 INTRODUCTION

1.1 DOCUMENT OVERVIEW

This document presents the results of the 1st part of the Prototyping phase. It consists of six chapters:

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

2. **Formal Procedures** – This chapter clarifies the starting position of the 1st part of the Prototyping phase as well as the internal process of managing it, by summarizing the discussions at meetings with the suppliers and within the PREFORMA Consortium;

3. **Complementary Issues** - This chapter discusses complementary issues not targeted in the objectives of the 1st part of the Prototyping phase;

4. **Open Releases** - This chapter gives an overview of the results that the suppliers have achieved as presented through their software releases and the reports they provided during the 1st part of the Prototyping phase;

5. **Points on Progress** - This chapter analyses the progress that suppliers have made with regards to the project objectives as well as things that may still be missing.

6. **Conclusions** - This chapter presents the main outcomes and what remains to be done.

The document also provides three annexes:

Annex 1: Template for Intermediate Reports

Annex 2: Template for Final Reports

Annex 3: Following Open Source Projects – This annex gives details on how to follow the suppliers´ work on their open development platforms in real time.

1.2 AIMS AND OBJECTIVES

Deliverable D8.2 "*Design – First report*" reports on the process of evaluating the bids for PREFORMA phase 2 (Prototyping) and its outcomes. Three suppliers were invited to sign 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 media types for phase 2 (Prototyping) 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).

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

According to the Description of Work of the project (DoW), suppliers are expected to provide software prototypes that fulfil the requirements of the project, to demonstrate the results and to 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.

Formally, WP 6 is composed of two tasks: T6.1 Prototyping Step 1 and T6.2 Prototyping Step 2. The Work Package leader as well as leader of both tasks is Riksarkivet.

Task T6.1 Prototyping Step 1 is executed by the suppliers and aims to deliver and demonstrate the first three prototypes, which include:

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

The objective of this deliverable D8.3 is to provide a report on these first 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 the deliverable D8.8 "*Monitoring of the Open Source Project Implementation*". This provides more in-depth feedback on whether the suppliers have adhered to the requirements in deliverable D4.3 "*Functions of the Open Source Portal*". D8.8 also evaluates further how each open source project has progressed in the area of implementing open work practices for software development, such as frequent open releases and promoting activities aimed at a sustainable community.

The aim is that this deliverable will also serve as a basis for the Re-design phase and for the coming development in the 2nd part of the Prototyping phase. Therefore, it goes rather deep into details.

The general approach used for this deliverable is to adapt to a situation assessment of the open source projects working on PREFORMA. This task was carried out through:

- an investigation into the self-reporting of the suppliers, apparent in reports submitted at two stages (an intermediate and a final) during the 1st part of the Prototyping phase;

- an investigation into the web presence of the projects;

- a brief consideration of those areas where visible, noticeable progress is evident..

1.3 STEERING DOCUMENTS FOR THE 1ST PART OF THE PROTOTYPING PHASE

Aside from the Call Documents (Invitation to Tender, Challenge Brief, and Framework Agreement) a few others can be identified as particularly important for the work preformed during the 1st part of the Prototyping phase.

First are the deliverables submitted by PREFORMA, among which the following have been directional ones :

- 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. In the 1st part of the Prototyping phase this 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, PREFORMA aims 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.

  This "check list" is used to evaluate whether the requirements communicated by PREFORMA are fulfilled in the software releases produced by the Suppliers.

- D8.1 "*Competitive Evaluation Strategy*", where the evaluation framework is defined based on contributions from the technical partners and the 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 was also 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 the 1st part of the Prototyping phase is that D8.2 also describes the way (methods, measures, principles) the PREFORMA Consortium and its external reviewers carried out the formal review, the evaluation, and the preparatory work to decide from which suppliers to invite submissions on the bid for the Prototyping phase.

- D4.3 "*Functions of the Open Source Portal*", which reports the functions of the Open Source Portal; presents the requirements for each associated open source project website, and describes the template layout for the portal pages. It also sets out the direction for how the work in the Prototyping phase (WP6) will be conducted. The Open

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

Source Portal is a dedicated section within the PREFORMA website that provides references to each open source project's development platform. The portal includes references to collaboration tools, source code, build environment, executables, test files, and other information related to each open source project. Each open source project focuses on one type of file format and all developments of software, and associated digital assets, related to the project (e.g. roadmaps, instructions, issues, email communication, forum dialogues, documentation, synthetic test files related to the file format handled by the project, etc.) will be available on an open development platform (GitHub or equivalent).

Second, PREFORMA started during the Prototyping phase 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 shall also 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), and describes the workflow for the data provision, storage and orchestration (including responsibilities) of the different actors involved: external file providers, PREFORMA Consortium members, PREFORMA dispatchers, and suppliers. DMP also analyses the possible legal and copyright implications. The DMP is 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. The document is as a work in progress, and is updated as the project advances. It consists of two parts: (1) Legal Background, which explains the primary source regulating the rights and obligations of the parties of the PREFORMA project, and the secondary sources, where the interpretation and understanding of the primary source are expressed; (2) Conceptual Framework, which conceptualizes the PREFORMA project and examines the legal issues that may arise in that context.

2   FORMAL PROCEDURES

The purpose of this chapter is twofold: firstly, to clarify some basic conditions for the Prototyping phase and especially its 1st part and, secondly, to present overall management by summarising the discussions held at meetings with the suppliers as well as internally 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: the Design phase, followed by a Prototyping phase, and ending with a Testing 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, knowing that the PREFORMA Consortium would make an evaluation based selection from six to three Suppliers when the phase was concluded.

The major phase 2, Prototyping, is intended for the development of prototypes based on the functional and technical specifications which the suppliers provided during the Design phase. The Prototyping phase is subdivided into three distinct stages:

- Prototyping phase, part 1, which took place between April and October 2015;
- Re-design , planned to take place between November 2015 and February 2016;
- Prototyping phase, part 2, which will begin on the 1st of March 2016 and will last until December 2016.

During this major phase 2, the three chosen suppliers will prove to the interested parties of the project that their respective validation products are able to meet the original PREFORMA Challenge (set forth in the Challenge Brief).

Finally, in the major phase 3, Testing, which is planned to take place for six months from January to June 2017, the validated applications will be tested by the memory institutions of the PREFORMA Consortium. Contracts for phase 3 will be based on the successful completion of phase 2, which means that the Consortium will evaluate 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 phase (Design) a signed contract for the Prototyping phase (including Re-design), which supplements the Framework Agreement. All the invited suppliers returned signed copies.

A payment plan was set for the Prototyping phase: 30% in advance, 30% intermediate, and 40% as final payment. The first 30% was paid out during April and May. Invoices for the intermediate payment were requested by Riksarkivet before Christmas and were due to be paid in January 2016.

An additional sum of 30.000 EURO remained available in the budget after the contracts had been signed with the three suppliers. This sum is planned to be used during the 2nd part of the

Prototyping phase to improve the quality of the conformance checkers and/or to add additional functionalities. The PREFORMA Consortium will, together with the suppliers, evaluate the best way to invest them.

2.2 MEETINGS WITH THE SUPPLIERS

2.2.1 Intentions

The Prototyping phase started with a formal (virtual) kick-off meeting on April 14, 2015, where a number of issues were discussed in order to prepare the selected suppliers and the PREFORMA Consortium members for the work ahead. The kick-off meeting was the first of a number of Prototyping phase meetings with the suppliers, to ensure that all of the relevant information was provided to suppliers and Consortium members alike.

Another measure installed and used by the PREFORMA Consortium during the 1st part of the Prototyping phase was the organisation of a review process, with special focus on the software releases which were to take place throughout its duration. Besides providing the Consortium with both an intermediate and a final release of software, the suppliers were required to prepare reports that described their releases, the progress compared to their previous releases and several other issues of interest. Also included in the review process was feedback from those within the Consortium that tested the software. Their feedback was compiled into separate reports, which were submitted to the suppliers.

The supplier meetings, therefore, had a double focus throughout this 1st part of the Prototyping phase. Firstly, to provide a transparent forum for discussion of the relevant issues brought up by various parties (whether technical, organisational or regulatory). Secondly, 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 (see section 1.3 for further discussion about steering documents for the 1st part of the Prototyping phase).

2.2.2 Common meetings

Kick-off meeting

During the virtual kick-off meeting, a number of issues were brought up to establish work procedures for the 1st part of the Prototyping phase.

The Project Management Team (PMT) presented the result of the evaluation that took place at the end of the Design phase, based on the final evaluation report dated 10th of April 2015. The report outlined the four evaluation criteria used by the Evaluation Committee: a) Impact on Challenge, b) Technical Approach, c) Quality of the Tender and d) Cost, resulting in a ranking from one to six between the six Suppliers that applied to take part in the Prototyping phase.

In first place was the proposal provided for a project in the area of PDF-file validation by the veraPDF consortium. Second place went to the MediaArea proposal for an open source project to validate files in the audiovisual area, followed in third place by the proposal presented by EasyInnova for their project in the area of validation of image files.

The ranking from one to three by the members of the PREFORMA Evaluation Committee formed the basis for a decision to award contracts to veraPDF, MediaArea and EasyInnova for the duration of Phase 2, providing the members of the PREFORMA Consortium with innovative

projects purposing to develop validator applications in all the three areas that were originally of interest: text files, image files and audiovisual files.

The Prototyping phase itself was presented. Firstly, communication channels were established, in the form of a general mailing list (prototyping-phase@preforma.project.eu), where suppliers could post questions to persons responsible for the phase itself. Secondly, important dates were set for planning purposes, including the connection between the 1st part of the Prototyping phase and the Re-design phase. Thirdly, a schedule for the release of software was mentioned, pointing out that intermediate and final releases were expected as well as regular or frequent releases, to be discussed with each individual supplier. Fourthly, a plan was set to arrange periodic virtual meetings. Fifthly, the issue of interoperability between different conformance checkers was raised, as well as the set up of a common Open Source Portal on the PREFORMA website.

The kick-off meeting also included plans of a practical nature concerning contracts and payments, as well as the evaluation that will take place following the end of the Prototyping phase.

First virtual meeting

A first virtual meeting with the PREFORMA suppliers was held in the Adobe Connect room used for PREFORMA on the 25th of May. Issues on the agenda for the meeting were: a) interoperability, b) test files, c) monthly releases, and d) use cases.

With regard to interoperability, a date was set for a special face-to-face supplier meeting in Brussels for the 11th of June, to be arranged by the PREFORMA partner PACKED (see below). For the first time, the issue of test files was raised, including many aspects such as licensing, storage of files, and the number of files that would be appropriate for the suppliers to use throughout the Prototyping phase. Clearance and direction regarding test files were provided through the first version of the Data Management Plan. Further, there was optimism among the suppliers with regard to the possibility of releasing new versions on a monthly basis, with the distinction that these should be stable ones.

At the meeting the three suppliers also asked the PREFORMA Consortium to define the use case for making the conformance checkers interoperable with each other. The relevance of this issue had wide ranging implications for the development work during the 1st part of the Prototyping phase. Should there be a shell component that can enable the wrapping up of the different – three – checkers into one single conformance checker? This requirement was part of the functional architecture of the Challenge Brief.

First guidelines on this issue were penned during the 1st part of the Prototyping phase by members of the Consortium in an internal "*Use Case for Interoperability*"-document (dated June 10).[1] This was then discussed with the suppliers in the face-to-face meeting in Brussels the following day (June 11; see above). The minutes from this supplier meeting in Brussels contains

---

[1] https://docs.google.com/document/d/1wDY4qDPTYRxjLuFRHq6Fis8frbjm_A8Rhdq9YbJEgKo/edit#

six core aspects of interoperability (see section 3.3) which were reported on during the next virtual meeting.[2]

Second virtual meeting

A second virtual meeting with the PREFORMA suppliers was conducted on the 15th of June. This meeting addressed the initial preparations of the Prototyping phase. Points discussed included:

- interoperability, bringing up each of the six aspects of interoperability discussed at the meeting in Brussels (June 11; see above);

- test files: Feedback on a previous email on this phase was requested; the suppliers highlighted the importance of accessing policy requirements as part of their testing activities;

- status report on monthly releases;.

Those participating at the meeting were also informed that a PREFORMA GitHub account had been set up (https://github.com/preforma), and that each of the suppliers had been asked to begin supplying information, links and resources to the Open Source Portal on the project website.

Third virtual meeting

Before the release of software at the end of July, a third virtual meeting with the suppliers was conducted. They were instructed to provide a short report to accompany the release, describing it, underlining the gaps and things that are missing compared to the concepts provided during the previous phase. A template for the report was distributed (see Annex 1).

The main issue that was discussed during the third supplier meeting was the test files. It was reported that a first draft of the Data Management Plan had been sent out. Based on this report, a need to clarify definitions, roles, responsibilities, workflows and requirements for the provision and storage of test files was raised.

During the meeting, it was also mentioned that contact had been made with the memory institutions concerning unique organisation related policy requirements to be used in the software development throughout the Prototyping phase. This is of vital importance to the project, since checking policy requirements arguably belongs to the core functionality of the validator/validators to be developed in PREFORMA.

Fourth virtual meeting

After the first release of stable software at the end of July, a fourth supplier meeting was held on the 19th of August. The PREFORMA Consortium noted that both releases and the accompanying short reports had been well received and that they provided the project management team with a good picture of the current development. A next step announced was

---

[2] https://github.com/preforma/meeting-minutes

the arrangement of individual virtual meetings with each of the three suppliers to further refine the Consortium's input and feedback.

The issue of the test files had a recurring role during the fourth supplier meeting as well. The participants were informed of the progress of the Data Management Plan that Riksarkivet were developing, and of Riksarkivet´s plan to enter an agreement with a Service Provider for a cloud based storage solution (the PREFORMA Vault). Legal issues were raised, for instance concerning the need to ask file providers for permissions to use the files for technical purposes only and how files will be licensed. Members of the PREFORMA Consortium were asked to estimate the number of test files they intended to supply.

At the meeting, the Project Management Team also reported that requests for policy requirements from external stakeholders had been prepared within Work Package 3, and that the team would follow up on its progress.

Fifth virtual meeting

A fifth suppliers meeting was held one month later on the 14th of September. A plan for individual meetings with the suppliers was set, and for those meetings the PREFORMA Consortium prepared three separate documents containing specific comments related to the three projects.

The Project Management Team reminded the suppliers that at the end of the 1st part of the Prototyping phase (in October) they are expected to provide the second stable release of the conformance checkers with an accompanying report. This report will follow more or less the same structure as the one delivered in July plus some sections for answering to the feedback received by the PREFORMA partners and to compare the outcomes with their plans in the beginning of the Prototyping phase. A template for the report was sent out as for the Intermediate Report (see Annex 2).

 It was also noted that the Re-design phase is scheduled from November 1st until February 28th and that it will follow the same approach as the first Design phase in the previous year. At the end of the Re-design phase there will not be a formal evaluation and selection. All the Suppliers will enter into the 2nd part of the Prototyping phase.

The issue of policy requirements was also brought up. PREFORMA received a suggestion from its Community to use the catalogue of policy elements created by the SCAPE project, which was forwarded to the Suppliers.

Sixth virtual meeting

A last and sixth Supplier meeting was conducted virtually on the 28th of October in conjunction with the second and final release of software that took place at the end of October. The template for the complementary report was sent out with a deadline for submission on October 31, 2015. It is similar to the July 2015 report template but on this occasion the suppliers had to clarify both progress since the intermediate release in July 2015, and progress compared to what was delivered at the end of the Design phase in March 2015 (end of Design Phase Report with Functional and Technical Specification)

The Project Management Team would like to see:

- the status of the development outlining the functionalities that have been planned for each module (implementation checker, policy checker, reporter, fixer);

- items that are still missing;
- the plan to complete these functionalities.

It was also stated that the work would continue further after the October release, leading into the Re-design phase planned to commence in November. Development is still expected to continue throughout the Re-design phase, with the goal that the suppliers can present and demonstrate stable software versions at the forthcoming Open Source Workshop in Stockholm in April 2016.

Further points at the meeting included a report on the progress of the PREFORMA Vault. Dispatchers had been designated to facilitate the distribution of files. A form for describing the technical properties and metadata of the files were in progress as well. This form was deemed necessary because the Project Management Team did not anticipate that content providers would fully describe the files. If needed the files received by contributors could be analysed before further distribution to suppliers as training files. Such analyses will definitely be done with files used for testing by PREFORMA WP7 (Validation and testing).

Another major point addressed at the sixth supplier meeting regarded the standard setting issues brought up in the deliverable D4.3 "*Functions of the Open Source Portal*". Here, a report from a parallel virtual meeting with suppliers conducted the day before was delivered and discussed. At the parallel meeting it was suggested that the Debian package management system could be utilized to manage all build/compile and runtime dependencies for conformance checkers written in C++ and Java. Carl Wilson from veraPDF volunteered to take the lead in organising and assisting, with information and other resources, the suppliers to create Debian packages for conformance checkers written in Java and C++.

2.2.3 Individual meetings with suppliers

During September, separate meetings were held with each of the three suppliers at which feedback on the software release and the reports submitted in July were discussed. The main content of the meetings was a discussion about the PREFORMA requirements, specifically those supplied in the deliverable D4.3 "*Functions of the Open Source Portal*" to clarify potential misunderstandings.

What emerged from the discussion was that the latest version of this deliverable had not been sent to the suppliers. The deliverable was updated in April 2015 following a request by the EC reviewers to further develop some parts of the text, particularly the sections related to source code, build environment and executables. This revision was only formal and not substantial, i.e. the content and the requirements did not change from the version that was sent out in December 2014.

The new Appendix A (of the version 2.0) explicitly lists all the different combinations which were before summarised in the main text (Sections 3.4 – 3.5 – 3.6). It includes 24 different unique combinations, of which six (1, 5, 9, 13, 17, 21) refers to "non-integrated combinations" and 18 (2, 3, 4, 6, 7, 8, 10, 11, 12, 14, 15, 16, 18, 19, 20, 22, 23, 24) refers to "integrated combinations".

However, both the deliverable and the updated version were produced before the discussions were held about the interoperability of the different conformance checkers and before the meeting with the suppliers in Brussels in June (see above). Therefore, the "integrated combinations" contained in Appendix A are not mandatory requirements anymore, the Project Management Team explained to each of the suppliers.

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

The Project Management Team (PMT) expressed concern that eventual misunderstandings regarding expectations on technical details need to be brought up on the Supplier meetings.

In addition to the main discussion concerning the requirements of the D4.3, each meeting provided the suppliers with opportunities for individual dialogues with the PMT. MediaArea reported that it had interacted with format providers to initiate a dialogue about file format standards. Further, they had presented their future application at conferences as a means to disseminate early results, and to create interest in PREFORMA as a means to ensure digital preservation.

EasyInnova reported on efforts to build a community around their product. They also informed on a related parallel activity from their side to create an "archival" version of the TIFF standard. On the 15th of September, a first meeting within the so-called TI/A standard initiative was held. The purpose of the TI/A standard initiative is to prepare a standard proposal to submit to the ISO (International Organization for Standardization). For more details, see section 4.5.3.

The veraPDF consortium presented the issues outlined in their report, and gave a broad presentation of the current status for PDF/A (23rd of September).

2.3 PREFORMA CONSORTIUM MEETINGS

2.3.1 Project Management Team meetings

In addition to organising supplier meetings, the Project Management Team (PMT) met to discuss issues, especially in the beginning of the Prototyping phase. In May, a number of strategic issues were mentioned, among them:

- Interoperability
- Frequent releases
- Test files
- Usage and role of D4.3
- Further development of use cases

The way in which these issues have been elaborated upon throughout the 1st part of the Prototyping phase is addressed further in this deliverable.

The PMT also met virtually in July to discuss the imminent intermediate release, which took place later that month. Further preparation for the distribution and use of test files was discussed. PMT meetings addressing the same issues were also held in August and September.

2.3.2 PREFORMA plenary meeting

On October 8th and 9th 2015 a plenary meeting was held in Pisa, Italy. In addition to reports on the overall status of the project and review recommendations, it focused on:

- Reports and discussions concerning the on-going 1st part of the Prototyping phase, mainly how to review and evaluate the final releases and reports, managing the PREFORMA Vault, and handle the issue of interoperability;

- Planning of the Re-design phase and the Testing phase, in the latter case including starting to share training files with the suppliers during the Prototyping phase;

- Planning of the Open Source Workshop and other events in April 2016 in Stockholm, by setting the structure of the work shop and additional activities (a plenary meeting and the second review of PREFORMA);

- Involvement of external stakeholders, which includes demonstrations of PREFORMA prototypes and planning for the Experience Work shop in Berlin in December 2016.

2.4 FORMAL OUTCOMES OF THE 1ST PART OF THE PROTOTYPING PHASE

Although no formal decisions were made with regard to the suppliers during the April to November period, a number of structures were either already established before the Prototyping phase began, or put in place to facilitate the development of the PREFORMA application.

First of all, directional input in the form of expectations placed on Suppliers were provided, and discussed. Input to this was partly provided in the Call documents, partly outlined in the Work Package plan, but further developed in the Deliverable D4.3 "*Functions of the Open Source Portal*".. A major aspect of this framework of expectations is that suppliers should not focus merely on the open source software development, but also take part in dissemination and standardisation activities, according to their own sets of priorities. Expectations and goals were also placed on the three suppliers regarding the way in which their software would be developed, distributed and released throughout the period (most notably in the D4.3, but aspects relating to this were mentioned in the Call documents). Some of these expectations were addressed at supplier meetings (in September and October).

Second, a structure of software releases was set in place, with immediate and final releases planned for and delivered by the suppliers based on the functional and technical specifications that each supplier presented during the Design phase. This meant that suppliers were to show continuity with respect to the concepts and plans that they had previously presented, and to show that their work was within project budget and to the reasonable satisfaction of those that tested the software.

A third type of formal outcome regarded separate meetings with each of the suppliers, common supplier meetings and other avenues to facilitate and distribute feedback after each of the two releases.

A fourth type of outcome from the Prototyping phase had to do with acquisition and storage of test files, based on the guidelines within the Data Management Plan.

2.5 PROTOTYPING PHASE EVALUATION

At the end of the Prototyping phase (end of December 2016), the results of the three open source projects will be evaluated following an approach similar to that of the Design phase.

Although the official evaluation will start in January 2017 it is of crucial importance to start planning it in advance:

- Firstly, to give the suppliers time to improve their prototypes as testing needs to start during the Prototyping phase;

- Secondly, to ensure that the conformance checkers meet all the requirements that have been identified in the PREFORMA Challenge Brief, work needs to start early both on the testing methodology and on the collection of test files, involving memory institutions inside and outside the PREFORMA Consortium. A basic requirement is the ability to

distinguish between files that are distributed in advance to the suppliers to test their prototypes, and files that are used in the official evaluation following the Prototyping phase.

Two extra check points have been 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;

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

3 COMPLEMENTARY ISSUES

This chapter summarises some issues that were raised during the 1st part of the Prototyping phase.

3.1 TEST FILES

To facilitate several aspects of work in the PREFORMA project, test files have been given a lot of consideration (i.e. organic or real-world files and not synthetic test files, which the suppliers either develop or collect). In PREFORMA, test files are intended to be used for different purposes:

- Examination: files used for testing by the PREFORMA members/partner network during the Prototyping phase. It distinguishes tests done locally, "in-house", or in an external environment but not made public. These files can later either be provided for training or for evaluation, but not both.

- Training: files used for testing by the suppliers during the Prototyping phase (WP 6). These files cannot be used as evaluation files.

- Evaluation: files used for testing by the PREFORMA Consortium during the evaluation in the Testing phase (WP 7). These files cannot have been used for training, but need to correspond to the type of files used for training.

- Dissemination: files used for presentation purposes by the PREFORMA Consortium and suppliers during the outreach phase (WP 4) or later by third parties for miscellaneous purposes, such as evaluating the PREFORMA conformance checkers.

The Data Management Plan (DMP), described in section 1.3 above, was created to fill the function of a framework for handling files to be used in testing the prototypes. It outlines components in the process of testing like workflow for the data provision, storage and orchestration (including responsibilities) of the different actors involved.

The DMP also describes the setting up of the PREFORMA Vault, the central location where providers of test files will submit their files. The purpose of the vault is to have an effective and practical method for handling all submissions of incoming and outgoing test files from providers to suppliers, in order to:

- guarantee the best possible distribution of training and evaluation files;

- keep track of the content that has been used for the training files, to ensure that the test files correspond to the training files and that they have not been released in advance to the suppliers.

The PREFORMA Vault makes use of the Amazon S3 service. Since PREFORMA is not a legal entity, Riksarkivet has entered into an agreement with the Service Provider.

The idea is that anyone within or outside of the PREFORMA Consortium can provide files for the project, and that these files will be sent to the PREFORMA Vault. For the administration of the Vault and to monitor the flow of incoming files, three PREFORMA dispatchers have been designated: Bengt Neiss (KB), Erwin Verbruggen (NISV) and Sònia Oliveras i Artau (AJGI). They are responsible for receiving files, and organising and allocating which files are to be used for training and evaluation, keeping them separate and distributing the training files to the suppliers.

Files will be collected under a simple agreement, making a clear distinction between internal uses of files for testing and other development related purposes, and external use of files for dissemination.

A call for participation has been launched with the aim of involving memory institutions and other content providers outside the PREFORMA Consortium in sending test files and participating in the testing phase. So far 16 institutions and organisations have expressed their interest to contribute.[3] A mailing list (*preforma-testing@promoter.it*) has been created for those external stakeholders and a dedicated webpage on the PREFORMA website (*http://www.preforma-project.eu/external-partners.html*).

In the short term perspective, a dedicated metadata form has been developed in cooperation with the suppliers to make it possible to start collecting information about training files for testing as soon as possible. This form includes a limited amount of information such as descriptive metadata, some technical properties of the files, copyright restrictions and expected behavior. Most of this information is not mandatory, as we don't want to discourage the providers by asking them to spend too much time filling in metadata.

Targeting the collection of test files during part 2 of the Prototyping phase, a specific task force, led by the University of Padua, has been organised to identify all the information that needs to be attached to the test files for a proper analysis to be made. This includes:

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

- Establishment of the so called ground truth.

3.2 SOFTWARE DEMONSTRATIONS

According to the DoW, two demonstrations will be organised by the Suppliers and reported in this deliverable D8.3 "*First Prototype Report*" and in the coming deliverable D8.4 "*Design – Final Report*". During this 1st part of the Prototyping phase, demonstrations have not been carried out by the suppliers in any pronounced way.

The first public demonstration to be organised by the PREFORMA Consortium is the Open Source Work Shop in Stockholm in April 2016. As stated by the Project Management Team at the sixth supplier meeting in October, the development of the validators is expected to continue throughout the Re-design phase, with the goal that suppliers will present and demonstrate stable software versions at this workshop.

The announcement of the workshop started when the first versions of prototypes were released by the suppliers. Different communities received information about the PREFORMA project, were invited to the workshop, and informed about the first results and the possibility of participating in the Testing phase. As recommended by the reviewers in the first PREFORMA review, communities outside of PREFORMA's main target groups have been addressed, according to their specific needs (e.g. digital preservation, standardisation, interoperability, security, etc.).

---

[3] See deliverable D3.3 "*Network Report Year 2*"

However, the PREFORMA software has been demonstrated by a project partner at three conferences/workshops in October and November 2015 in Sweden (for a full overview of dissemination activities, see deliverable D4.5 "*Dissemination Report Year 2*").

3.3 INTEROPERABILITY OF DIFFERENT CONFORMANCE CHECKERS

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

At an internal level

The PREFORMA Consortium and the suppliers agreed that this firstly is an issue for the suppliers to handle. The suppliers met on the 11th of June in the venue of PACKED in Brussels. The initial discussion focused on an internal use case document for a common shell "*Use Case for Interoperability*" prepared by the Consortium. The main outcome of this meeting was a timeline indicating how and in which order it would be sensible to tackle the issue of interoperability, namely:

1. Common XML Report format - implementation checker, policy checker results

    a. Need to agree on common structure / schema for the report (top-level part)

2. Agreement on common report structure needed by the end of July

3. Common XML Configuration - Policy checking profiles

4. Common XML Configuration - Metadata fixer

5. Service discovery

6. Common API

    a. will need a few iterations

    b. address in Re-design phase

The PREFORMA partner FRAUNHOFER has taken on the responsibility to coordinate this activity.

The aim is that the suppliers will have time to work on interoperability at the beginning of the Re-design phase, when the PREFORMA Consortium is evaluating the results of the 1st part of the Prototyping phase.

At an external level

PREFORMA has signed a Memorandum of Understanding with the EU project Europeana Space. One of the coordinated actions mentioned in this document is the integration of the open source tools developed in PREFORMA (that control if a file complies with standard specifications and with other, acceptance criteria specified by memory institutions) into the Technical space of Europeana Space. The benefit for PREFORMA would be that this experiment could represent a proof of concept / test case where we try to embed PREFORMA software into a specific environment (external interoperability).

At the beginning of December 2015, representatives of PREFORMA and Europeana Space met in Tallinn and developed a plan for integration.

PREFORMA has also had initial contact with the EU project E-ARK, with the intention that the PREFORMA conformance checker should be integrated as a tool in their pilot archival services.

3.4 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. At the second virtual meeting (see section 2.2.2) the suppliers underlined the importance to have complex policy requirements in place, if possible by the time of the Re-design phase.

In the call for participation to external stakeholders, a request for valid policy requirements was added. The feedback, however, was minimal and this issue was, therefore, dropped with the advice to the suppliers to use the SCAPE Catalogue of Policy Elements instead. [4]

---

[4] 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   OPEN RELEASES

The suppliers are, according to the DoW, expected to provide two kinds of releases during the 1st part of the Prototyping phase:

- frequent (if possible, monthly) releases;

- two intermediate releases, one at the end of July 2015 and the other at the end of October 2015; the last one being also the final release of the 1st part of this phase..

These releases are expected to show that the requirements communicated by the PREFORMA Consortium are fulfilled. Requirements particular to this phase are first of all provided in Deliverable 4.3 "*Functions of the Open Source Portal*", while additional, basic requirements regarding software releases are provided in the Call documents (Invitation to Tender, Challenge Brief and most notably the Framework Agreement) and in deliverables D2.1 "*Overall Roadmap*" and D2.2 "*Tender Specifications*".

In order for these requirements to be fulfilled, they must be discussed and deliberated according to the formal procedures in place within the project. If and when there are misunderstandings concerning any particular group of requests made by the PREFORMA Consortium, these have to be addressed as early as possible. Indeed, many basic requirements are being discussed. Whether they are subject to a feedback process or not, they all include open source best practices, intended to characterize each of the three respective projects. This will ensure that the overall goals of PREFORMA are reached satisfactorily.

In the following sections, the focus is on five different activities of the suppliers: a) software releases, b) software testing, c) dissemination and community building, d) open source approach and e) standardisation efforts. This categorization follows the outline of the template of the reports, which the suppliers submitted in conjunction with their two releases. These reports, called the Immediate Report (July) and the Final Report (October), are both internal working documents. The suppliers describe in more detail the respective release, and how they are providing the required functionalities. At the time of the releases, members of the PREFORMA Consortium were invited to provide feedback both on the software and on the self-reporting of the suppliers. The releases, the reports, and the feedback provided by the Consortium comprise the source materials for this chapter.

4.1   ACTIVITY 1: SOFTWARE RELEASES

In the reporting structure established by the PREFORMA Consortium, suppliers were asked to provide an overview of the releases, and of the functionalities that were available at the time of the respective Intermediate and Final Reports. Software releases are the results of work within the respective open source projects assigned to each of the three suppliers within the text, image, and audiovisual file formats. The file formats addressed are a particular set of standard

file formats that are a) open standards[5], b) considered appropriate for long term preservation, and c) relevant for the memory institutions participating in PREFORMA.[6]

In the official PREFORMA documents, many references are made to what the PREFORMA Consortium requires of suppliers concerning software release and deployment. Issues included are open source licensing, provision of executables, the delivery of executables for platforms or operating systems; the provision of source code, up-to-date roadmap with detailed milestones for different releases, the frequency of releases, the provision of open source tools for creation of the executable from source and so on.[7] A vital part of the feedback process during the 1st part of the Prototyping phase was to follow up with regard to these particular requirements, based on the first experiences of downloading, installing, and using the software.

4.1.1 veraPDF consortium

The overall plan for the components of the veraPDF conformance checker is described in the following way on the project's website[8]:

- The Implementation Checker parses and analyses PDF documents. It outputs a report describing the PDF document and its metadata, and a Validation Report describing conformance to PDF/A flavours;

- The Metadata Fixer makes a limited set of fixes to metadata within PDF documents. The Metadata Fixer produces a fixed version of the original document and a Metadata Fixing Report, which describes the fixes attempted, and their success or failure;

- The Policy Checker parses and analyzes a PDF Features Report and generates a Policy Report stating whether the PDF document complies with institutional policy as expressed in a Policy Profile;

- The Reporter transforms veraPDF machine-readable reports as generated by the Implementation Checker, Policy Checker, and Metadata Fixer, into other forms for downstream use;

- The Shell manages the veraPDF consortium's other components, and ensures interaction in a coordinated sequences of actions. Users interact with the Shell through the Command Line Interface (CLI), Desktop Graphical User Interface, or Web Graphical User Interface.

The veraPDF consortium made its first public release on 15 July 2015. The main motivation for the early release was to test the continuous build and release capabilities (as is explained

---

[5] Cf. http://ec.europa.eu/idabc/servlets/Docd552.pdf?id=19529

[6] http://www.digitalmeetsculture.net/wp-content/uploads/2014/06/PREFORMA_Challenge-Brief_v1.0.pdf (p.6)

[7] D2.1 "Open Roadmap", D4.3 "Functions of the Open Source Portal", "Invitation to Tender", "Challenge Brief" and the "Framework Agreement" all contain various requirements.

[8] http://verapdf.org/home/#about

further in the open source approach section in this chapter). In the veraPDF roadmap, the consortium communicated to users and other target groups that they should not expect too much from the first release, since it contained very few features of a conformance checker as well as the possibility of bugs.[9]

The first release was/is a prototype of the veraPDF validation library.[10] The following points were made by veraPDF concerning its first release:

- (It is) a prototype GUI application;

- (It has) an unfinished implementation of the generic validation model;

- (It is characterised by) incomplete PDF parsing (or analysis[11]);

- (It has) a small set of validation rules as atomic profiles; and

- (It contains) prototype PDF feature and validation reporting.

In conjunction with the first release, the veraPDF consortium established a download area for both the development and public releases.[12] At this stage, veraPDF were making daily development releases, which is why it also established permanent links to the latest releases:

- http://downloads.verapdf.org/rel/veraPDF-library-GUI-latest.zip

- http://downloads.verapdf.org/dev/veraPDF-library-GUI-dev-latest.zip

The veraPDF consortium also made release notes for the version 0.2 available on GitHub, where the following is reported under the "features" headline:[13]

- The formal PDF model for PDF/A Level B validation[14]

- The set of validation rules covering ISO 19005-1:2005, 19005-1:2005/Cor.1:2011, 19005-1:2005/Cor.1:2007, 19005-1:2005/Cor.2:2011, Level B

- Implementation of the rules covering the following sections of ISO 19005-1[15]:

  o 6.1 File structure

  o 6.2 Graphics

  o 6.4 Transparency

  o 6.5 Annotations

---

[9] http://verapdf.org/roadmap/

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

[11] https://en.wikipedia.org/wiki/Parsing

[12] http://downloads.verapdf.org

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

[14] http://www.pdflib.com/knowledge-base/pdfa/

[15] Cf. PREFORMA Challenge Brief, p.6

- o 6.6 Actions
- o 6.7 Metadata
- o 6.9 Interactive Forms
- Initial implementation of the PDF Feature Report generation
- Minor improvements in the (graphical user interface) GUI and the Human-readable Report in HTML format

In the Final Report, the veraPDF consortium reports of two releases: the veraPDF 0.4 release and the veraPDF 0.6 release, with the second one being a "PDF/A1-b Release Candidate" according to the roadmap..

The veraPDF 0.4 release was made available on September 2015 and included the following features:[16]

- A number of bug fixes in the implementation of the formal model for PDF/A Level B validation[17]
- Added missing validation rules for full coverage of ISO 19005-1:2005, 19005-1:2005/Cor.1:2011, 19005-1:2005/Cor.1:2007, 19005-1:2005/Cor.2:2011, Level B
- Complete implementation of the PDF Feature Report generation
- Minor improvements in the GUI and the Human-readable Report in HTML format
- Added extra parameters to limit the number of rule failures and the number of reported errors
- Optimized performance

The veraPDF 0.6 release was made available on November 2nd 2015. The release is regarded by the veraPDF consortium as a stable or beta version of the software. Features mentioned in the final report as well as the release notes include:

- Stable (beta version) implementation of the formal PDF model for PDF/A-1b
- Prototype the formal PDF model for PDF/A-1a and PDF/A-2b, 3b
- Minor refactoring[18] and stricter naming conventions in validation rules for PDF/A-1b
- Prototype validation rules for PDF/A-1a and PDF/A-2b, 3b
- Prototype implementation of the Metadata Fixer
- Prototype implementation of the plug-in architecture for PDF Feature Report generation

---

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

[17] http://www.pdflib.com/knowledge-base/pdfa/

[18] Cf. https://en.wikipedia.org/wiki/Code_refactoring

- Optimized performance for PDF/A font rules validation (glyphs presence, widths consistency)

At the veraPDF website, downloads are available for Windows, Mac and Linux (no mentions of available distributions). Instructions on how to install software are published.[19]

4.1.2 MediaArea

In the Intermediate Report, the MediaArea Company explains that initial releases of MediaConch have been designed, built and released upon existing MediaInfo software. On the PREFORMA Open Source Portal, this open source project is described in the following way:[20]

> "MediaConch (CONformance CHecking for audiovisual files) is an extensible, open source software project consisting of an implementation checker, policy checker, reporter and fixer that targets preservation-level audiovisual files (specifically Matroska, Linear Pulse Code Modulation (LPCM) and FF Video Codec 1 (FFV1)) for use in memory institutions, providing conformance checking via an adaptable and flexible application program interface accessible by the command line, a graphical user interface, or a web-based shell. MediaConch is currently being developed by the MediaArea team, notable for the creation of open source media checker software, MediaInfo. Furthermore, the MediaArea team is dedicated to the further development of the standardization of the Matroska and FFV1 formats to ensure their longevity as a recommended digital preservation file format."

The core infrastructure of MediaConch scales to work on three different shells: a command line interface (CLI), a graphical user interface (GUI) and a web interface. Initial releases of the three are available for download or access at the MediaConch website, according to the report[21]. (The release notes at the website make mention of five 0-marked releases: 15.05 (both GUI and CLI) 15.06 (GUI and CLI), 15.07 (GUI, CLI and on-line), 15.08 (GUI, CLI, online), 15.09, same shells, called September release in Newsletter no 1[22]). The web interface version is the MediaConch Online[23], which is "a web interface that allows a user to generate policy sets for reporting on files available either online or via upload without ever having to directly download or install any software"[24].

Another feature of the September release of MediaConch (15.09) was the integration of the MediaTrace XML reporting tool.[25]

---

[19] https://github.com/veraPDF/veraPDF-library#install-from-zip-package

[20] http://www.preforma-project.eu/mediaconch.html

[21] https://mediaarea.net/MediaConch/

[22] https://mediaarea.net/MediaConch/2015/10/09/mediaconch-newsletter-no-1/

[23] https://mediaarea.net/MediaConchOnline/

[24] https://mediaarea.net/MediaConch/2015/10/09/mediaconch-newsletter-no-1/

[25] https://mediaarea.net/MediaConch/2015/09/10/annoncing-mediatrace/

According to the Intermediate Report, the MediaConch software verifies that files conform to the standards set forth by Matroska, FFV1[26] and LPCM (Linear Pulse Code Modulation[27]) files. Results are exported as XML. Two options of XML were given for export, according to MediaArea: standard, descriptive XML, or extensive, trace XML. Both XML standards have been manufactured to be easily read by both "humans and computers" according to the report.

The MediaConch release (ver.15.10, beta[28]) of October has CLI, GUI and online versions as well. It features a brand new implementation checker concentrating on Matroska and EBML conformance checks, as well as illustrative policy sets allowing users to check conformance on preservation master files[29], among other workflows.

Further inquiry into the MediaConch website also reveals that the latest release of software (official release) is available for different operating systems, the main one being Windows, followed by Mac, and several Linux distributions (or distros: Debian, Ubuntu, Mint, redhat, CentOS, Fedora, openSuse, Suse).[30] It is also possible to test ongoing developments by a snapshots section.

According to the Final Report, metadata development has been the focus of the phase leading up to the release of the beta version. Four separate conformance checking schemas for file reporting were developed:

- MediaConch (a container of the results of implementation and policy checks);
- MediaInfo version 2 (a technical metadata report focused on audiovisual file analysis and categorization);
- MediaTrace (a report on the structure of a digital file format);
- MediaArea (which is a container XML for the other 3).

Metadata schemas and documentation for these can be found at the Media Conch GitHub.[31]

4.1.3 EasyInnova

With the intermediate release of July, EasyInnova provided an Alpha version of the DPF Manager.[32] This Alpha version was distributed for different operating systems, according to the Intermediate Report:

---

[26] https://en.wikipedia.org/wiki/FFV1

[27] https://en.wikipedia.org/wiki/Pulse-code_modulation

[28] MediaArea does not use terms such as "alpha" or "beta" while reporting. Traditional software versioning do ascribe the number "0" to alpha releases, and "1" to beta. https://en.wikipedia.org/wiki/Software_versioning

[29] https://www.archives.gov/preservation/products/definitions/preservation-copy.html

[30] https://mediaarea.net/MediaConch/download.html

[31] https://github.com/MediaArea/MediaareaXML

[32] http://dpfmanager.org/

- Windows, with the typical executable installer (.exe) and also a package installer (.msl);
- Linux, here an installer was provided for different Linux distributions:
    - Devian, Ubuntu, Min and derivatives with Devian installer packages (.deb)
    - Rehad, Fedora, CentOs and derivative with a Redhead package manager (.rpm)
    - OpenSuSe and SLES the executable a shared object (.so);
- Mac OSX, here the typical Mac installer was provided (.dmg) as well as the Apple software package (.pkg).

At the DPF Manager GitHub website, EasyInnova provides the following information in the README.md:[33]

> "DPF Manager is an open source modular TIFF conformance checker that is extremely easy to use, to integrate with existing and new projects, and to deploy in a multitude of different scenarios. It is designed to help archivists and digital content producers ensure that TIFF files are fit for long term preservation, and it is also able to automatically suggest improvements and correct preservation issues. The team (…) has leveraged the support of 60+ memory institutions to draft a new ISO standard proposal (TIFF/A) specifically designed for long-term preservation of still-images. An open source community will be created and grown through the project lifetime to ensure its continuous development and success. Additional commercial services will be offered to make DPF Manager self-sustainable and increase its adoption."

The first release was focused on showing the capabilities of DPF Manager working with TIFF files. Overall, the conformance checker (CC) is able to detect the tags inside the file, and handle the embedded metadata.

According to the Intermediate Report the implementation checker is able to confirm that a TIFF image is following the current TIFF Baseline 6.0 and check the TIFF/EP (Tag Image File Format/Electronic Photography) ISO 12234-2 compliance[34], the Intermediate Report says.

The report module generates a machine-readable report in XML[35] and JSON[36] format, with information about the TIFF internal structure and metadata. Furthermore, the report includes the result of the implementation checker. Experts and non-experts alike should be able to analyse the result of the validation process, according to EasyInnova Intermediate Report.

The release also included the first version of the TIFF library for Java.[37] According to the README at GitHub, this particular library is set up to import and export TIFF files, and to check

---

[33] https://github.com/EasyinnovaSL/DPFManager, see also: http://www.preforma-project.eu/dpf-manager.html

[34] http://www.digitalmeetsculture.net/wp-content/uploads/2014/06/PREFORMA_Challenge-Brief_v1.0.pdf (p.7)

[35] https://sv.wikipedia.org/wiki/XML

[36] https://en.wikipedia.org/wiki/JSON

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

compliance with TIFF Baseline 6.0, TIFF/EP and TIFF/IT. It is licensed under the open source project licenses required by PREFORMA.

In its Final Report, EasyInnova commented that the Alpha version of the software showed CC capabilities reading TIFF files. The TIFF CC was able to detect the tags inside the file (baseline tags, extension tags and widely used private tags[38]) and handle with embedded metadata such as ICC Profiles, XMP, IPTC and EXIF.[39] Now, EasyInnova reports, the first Prototype phase release was able to include all of the features which were missing in the CC intermediate release (version number of the beta version, according to DPF Manager News on twitter is 1.2[40]).

The implementation checker should now validate:

- TIFF baseline 6.0

- TIFF/EP

- TIFF IT (Graphic technology – Prepress digital data exchange – Tag image file format for image technology) ISO 12639:2004E.

The policy checker module was included in the release, too. It uses Schematron[41] to provide interoperability between CCs (in PREFORMA). Memory institutions can now create custom rules to define their own acceptance criteria, according to the report. The beta release included only some rules, but EasyInnova is expecting to introduce new ones once feedback is coming in from memory institutions.

EasyInnova also mentions that a newer version of the TIFF library for Java is used in the metadata fixer component of the CC. The metadata fixer is able to add and delete metadata inside the file following the guidelines for handling image metadata[42], provided by the metadata working group.

The reporter module of the CC provides a human readable report (PDF and HTML format) and a machine-readable report (XML and JSON format) for each file checked. If multiple files are checked, a global report is generated. Moreover, the report includes a comparison between the original image and the new file generated by the metadata fixer module.

The release also included the graphical user interface (GUI) with a wizard[43] to create and save configurations, check files and directories selecting a saved configuration and view the reports generated by the tool.

---

[38] http://www.awaresystems.be/imaging/tiff/tifftags.html

[39] http://fotoforensics.com/tutorial-meta.php

[40] https://twitter.com/DPFManager/status/660005871372673024

[41] http://www.schematron.com/

[42] http://www.metadataworkinggroup.org/pdf/mwg_guidance.pdf

[43] https://en.wikipedia.org/wiki/Wizard_%28software%29

In the feedback process related to the beta release, comments were made that the development seemed to be well under way, and that features had been implemented in a good way. It would have been good however if the different development stories would have been published in a roadmap through GitHub, one commentator said (cf. D4.3, section 2.3).

4.2   ACTIVITY 2: SOFTWARE TESTING

Software testing is an activity that is broad in scope within each of the open source projects. Testing as an activity has many different purposes and is described by different concepts. For instance, tests may be run by developers as code is compiled, with the purpose of verifying that functions are working the way they should (Unit Tests[44]). The results of this type of test may be enforced automatically into a new development version. Another type of test may be used to test the behaviour of the individual software components and their interaction with each other. This particular type of test takes place in a process called continuous integration (CI)[45], and is carried out by all suppliers. Both of these types of tests are mentioned in Intermediate and Final reports.

More central to the interaction between the PREFORMA Consortium and the suppliers, is the type of testing which purpose is to the test specific functionalities of the validator. A supplier may test the reporting component of the CC, for instance, by use of a certain type of corpus of test files. Another (or the same) supplier may use another corpus to test the validation or CC-component (a validation of the validator), to check whether the software detects violations in the files of the standard.[46] Here, the suppliers use synthetic files. This particular file type is explained in a working document of the 1st part of the Prototyping phase[47]:

> "Synthetic files are synthetic in the sense that they are purposefully created to test a specific functionality of the validator. They work as a frame of reference of what the validator is to validate as a "correct" or "incorrect" implementation of a file format". (…) Synthetic files are used to set the accuracy of the (CC)."

Provision of synthetic test files must be licensed under the open source licenses of the project and distributed in an accessible fashion on the open development platform.[48]

Software testing is also a user related activity in PREFORMA. Usability testing has been carried out to check if the user interface is easy to understand and use, and members of the PREFORMA Consortium have been asked to report in writing on their experiences of downloading and installing the first releases. This type of end-user testing of software functions

---

[44] https://en.wikipedia.org/wiki/Software_testing

[45] https://travis-ci.org/ (a tool used by the suppliers to achieve this purpose)

[46] http://www.pdfa.org/2011/08/isartor-test-suite/

[47] "Data management plan for training, testing and demonstration files in the PREFORMA project", p. 1

[48] Cf. "Functions of the open source portal", Deliverable 4.3, pg.12-13.

necessitates access to various types of test files.[49] These must be in the open file formats focused by the respective projects[50]. They may be developed by the suppliers, or be test suites or collections that were developed external to the project.

A third group of interest, outside of suppliers and users in the PREFORMA Consortium/network, is early adopters within the respective communities that have tested the software, either by downloading and installing, or by compiling it. Some comments from this group are included under each of the projects below.

The suppliers were instructed to provide descriptions of the datasets that have been used to test the release (own, memory institutions, external etc.) and the respective purpose of testing in their reports (Intermediate and Final).

4.2.1   veraPDF consortium

In the Intermediate Report, the veraPDF consortium reported early concentration on PDF/A validation functionality, which was tested against the synthetic test corpora under production. The corpora were under revision control on GitHub.[51] The README contains the following information about this particular collection of synthetic test files:

> "The repository contains the veraPDF test corpus for PDF/A specifications (Versions 1B, 1A, 2B, 2U, 2A, 3B, 3U, 3A) as well as a number of additional tests files for ISO 32000-1". (…) This test suite complements Isartor and Bavaria test suites and follows their test file pattern:
>
> - all test files are atomic;
> - they are self-documented via the document outlines; and
> - the naming pattern and the directory structure indicate relevant parts of ISO 19005-1 specification"

The Isartor test suite[52] is also used by the veraPDF consortium. This test suite for PDF/A-1b deliberately violates each requirement of the PDF/A-1b standard in order to check whether validation software detects all possible kinds of violation of the standard requirements, and whether it provides appropriate explanation for rejected non-conforming documents. A PDF/A validator which completely passes the Isartor test suite is known to implement all required checks.

---

[49] Cf. "Data management plan for training, testing and demonstration files in the PREFORMA project", p. 4 calls these files "organic", developed or used for the purpose of testing the consistency of the conformance checker.

[50] Cf. http://www.digitalmeetsculture.net/wp-content/uploads/2014/06/PREFORMA_Challenge-Brief_v1.0.pdf, chapter 3 "Open Source Projects".

[51] https://github.com/veraPDF/veraPDF-corpus

[52] http://www.pdfa.org/2011/08/isartor-test-suite/

For the sake of simplicity, only a single aspect of the standard is violated at a time. However, in a few cases this is not possible because of interdependencies among multiple aspects of the standard.

The test requirements (assertions) are the inverse of requirements in the standard. For example, PDF/A-1 mandates that fonts be embedded, so the test suite contains several documents where the fonts are deliberately not embedded. PDF/A validators must report this standard violation in order to pass the test.

The other test suite mentioned in the README, the Bavara test suite[53] contains both conforming and non-conforming PDF/A documents from a variety of sources, created with a variety of PDF generation products. The Bavaria package contains a descriptive file called bavaria.xml which contains comments regarding interesting validation aspects of the test documents.

In the Final Report, the veraPDF consortium reported that the collection which had been under production had grown to consist of more than 200 synthetic test files, complementing the Isartor and Bavaria test suites mentioned above. In total, more than 400 test cases are now available for developers during the Prototyping phase.

In addition to this, the veraPDF consortium reports that it carries out reliability or consistency testing of the prototype software by using a large set of random real-world PDF files from Digital Corpora.[54]

In the feedback process, one of the PREFORMA partners remarked that the test corpus was not licensed properly, and as a result corresponding actions were taken by the veraPDF consortium in time for the second, 31/10/2015 release.

The usability testing by members of the PREFORMA Consortium showed overall that the tool was functioning according to expectations. Other aspects mentioned in the feedback regarded for example: the installation process, portable installation, size of installation files, and interface issues. Several feature reports were also provided concerning issues such as conformance checking of files with regards to metadata standards.

Early adopters who want to get involved can do so, for instance by posting issues on the GitHub platform, whether it is about test files, the veraPDF source code library and so on. [55]

4.2.2   MediaArea

In its Intermediate Report, MediaArea reports that test files have been created for Matroska-wrapped FFV1 video files and LPCM files.[56] One category relates to "implementation test files", while another relates to "policy test files". The MediaArea team explains that "…these test files

---

[53] http://www.pdflib.com/knowledge-base/pdfa/validation-report/

[54] http://digitalcorpora.org/

[55] For instance: https://github.com/veraPDF/veraPDF-library/issues

[56] https://github.com/MediaArea/MediaConch_SampleFiles

exist using various combinations of specifications in order to allow for comprehensive tests for policy checks…" and are explained further in the Final Report.

The implementation test files focus on synthetic Matroska files. Here, synthetic Matroska files that either adhere to or violate Matroska specifications are used to provide samples to all implementation checker reporting outputs. The company wants to set up a diverse FFV1, Matroska and PCM collection in order to support large scale testing of the policy checker.

In addition to this in the 1st part of the Prototyping phase, a set of synthetic Matroska test files was developed with the purpose of presenting both an XML and EBML representation of Matroska files.[57] In the report, MediaArea mentions that it uses several other collections of test files as well.[58]

In the feedback provided by the PREFORMA Consortium, one of the partners remarks that:

"…besides provision of 'real' test files it is clear that provision of synthetic test files is also key. With this, the relevance for other open source project would also increase"[59].

In the Final Report, MediaArea reports further on its testing activities. It has established several collections of test files and datasets for three reasons:

1. to aid in the research and experimentation associated with the development of a conformance checker;

2. to inform the establishment of archival best practices for the use of Matroska and FFV1;

3. to be used in standardisation efforts in these formats.

The collections include files that were specifically created in order to trigger detection of certain implementation issues (synthetic), as well as large real-world sample tests for testing the implementation checker and policy checker prototypes (organic).[60] The largest dataset collection is made up of MediaArea XML reports from a large online collection of media files available at the project's GitHub.[61] This repository contains a research corpus used in the development of the MediaConch. At the time of writing their report, additional work was done by the MediaArea team to increase collections of test files, for instance by incorporating reporting on online Matroska collections such as those within http://samples.ffmpeg.org.[62]

---

[57] http://www.matroska.org/technical/specs/index.html

[58] Several links presented in the Intermediate Report under "testing" are now invalid. The files mentioned in the report that were provided in association with release of a demo of the software, are probably available at another location at the project´s GitHub: https://github.com/MediaArea/MediaConch/tree/master/Demo/files

[59] Feedback on the Intermediate Release – July 2015, p.3.

[60] Distinctions between synthetic files and organic files are made in "Data management plan for training, testing and demonstration files in the PREFORMA-project", pg 3-5.

[61] https://github.com/MediaArea/MediaConch_MKVSurvey/

[62] See the 00-README in this online library for overall information.

Progress on testing and provision of test files was noticeable by the PREFORMA Consortium (according to the Final Report), who's members also participated in testing by downloading and installing the prototype MediaConch software.

It is commendable that MediaArea has provided reports of demonstrations of its own to test various features that the software is able to perform.[63] Here, files are tested to demonstrate policies or rules designed in the conformance checker for strict conformance to a particular video standard (NTSC or PAL)[64] against a test file that has the characteristics of both. The result of this particular demonstration was a "standards mismatch" with regard to video width, height or general frame rate.

Some early adopters have also tested the prototype software. Their comments and questions can be read under the "issues" tab at GitHub.[65]

4.2.3   EasyInnova

EasyInnova reports of testing to ensure that they are building a robust application, one of the principles adopted during the development. In order to create quality code JUnit[66] tests (a type of unit testing)[67] were carried out. New functionalities added to the DPF Manager are tested automatically before they are added to the repository.

The JUnit tests were carried out by using image test files. These files have been created taking into account three aspects: image data representation, TIFF internal structure and TIFF with embedded metadata. While testing, EasyInnova included all of the possibilities defined in the TIFF baseline 6.0 specification and extensions. The image tests also covered the byte order of data stored in TIFF files (Little-Endian or Big-Endian[68]).

The ambition of EasyInnova is also that the tool covers all the possible TIFF internal structures which require the employment of different kinds of image tests. Another set of images was used to test embedded metadata inside the TIFF.

The image tests were completed with a set of images containing errors. The final report lists these errors: "incorrect headers identification", "incorrect byte order", "bad block alignments", "circular references", "re-used references", "duplicate tags", "unsorted tags", "incorrect number of tags", "premature end of files" and "unknown private tags".

These synthetic images were provided by the University of Basel, and distributed under the Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0). The image files were released

---

[63] https://mediaarea.net/MediaConch/demo.html

[64] https://en.wikipedia.org/wiki/PAL#PAL_vs._NTSC

[65] https://github.com/MediaArea/MediaConch/issues

[66] http://junit.org/

[67] https://github.com/junit-team/junit/wiki/Getting-started

[68] https://www.cs.umd.edu/class/sum2003/cmsc311/Notes/Data/endian.html

into the public domain under the Creative Commons license (both in line with requirements posed by PREFORMA).

One of the functionalities EasyInnova has prepared via the project website is the creation of a community of early adopters that perform some tests with their respective collections. For EasyInnova this is a very important channel to get feedback from. Samples of issues brought up by these early adopters can be found on the project's GitHub platform.[69] For the same purpose as user forum is set up at the project website[70], as is an IRC Channel.[71]

A comment made by a member of the PREFORMA Consortium within the feedback process is interesting:

> "For software testing we chose TIFF files that were inside the source code repository[72] for unit testing. Since these files are used for testing the software they contain errors detectable by the software. A real function test (usability test) should contain files from outside the scope of the development process. Apart from some long processing times, the tool was able to perform analysis on several TIFF Images".

Members of the Consortium which have downloaded and installed the software have overall positive impressions, reporting that it performs well.

4.3 ACTIVITY 3: DISSEMINATION AND COMMUNITY BUILDING

The open source projects contracted by PREFORMA should have a clear plan for outreach, and seek to perform different dissemination activities, and also to be actively involved in the building of open source communities around their GitHub-based projects. In fact, projects on GitHub can operate in many different roles, ranging from those that own the projects to contributors, and to community members that "care deeply about the project and are active in discussions for features and pull requests".[73]

The request that suppliers should be active in building open source communities is central to the regulatory documents – it's even mentioned in the Framework Agreement.[74] Further references to the requirements for suppliers can be found in the Invitation to Tender, and in the Challenge Brief.[75]. It is associated with the open source approach as such with its emphasis on development by openness, availability of code, tools such as documentation, e-mailing lists, and

---

[69] https://github.com/EasyinnovaSL/DPFManager/issues

[70] http://www.dpfmanager.org/Usersforum.html

[71] http://www.dpfmanager.org/UsersIRC.html

[72] https://github.com/EasyinnovaSL/DPFManager/tree/master/src/test/resources

[73] https://guides.github.com/activities/contributing-to-open-source/index.html. Pull requests are explained in this bit: https://help.github.com/articles/using-pull-requests/

[74] PPFA, ch.18, page 16.

[75] ITT,p.15; Challenge Brief, p.14.

collaboration between owners and contributors.[76] Another opportunity to interact presents itself by the supplier seeking to interact with organisations, if these can play a vital role in their respective open source projects.

4.3.1 veraPDF consortium

VeraPDF has a dedicated web page, which forms the hub of its dissemination activities.[77] At the starting page you can choose either to subscribe to the veraPDF newsletter or to explore what is going on in the project. Then, the website is organised through a number of tabs which point to various topics of interest: veraPDF, project, software, roadmap, community and news. This means that people who want to contribute to the project by providing software can get help (information on how to do so is found under the software tab). Further, there is information about the veraPDF consortium as such, plans for the development (a roadmap), the project and its purposes; and also to the groups that the consortium considers its primary target groups.

At the dedicated project website, you can access the latest press releases disseminated by veraPDF. For instance, they informed the press about being selected to Phase 2 of the PREFORMA project, and of its first public software release in July 2015.

Members of the veraPDF team are also actively blogging about PREFORMA-related issues such as how it approaches PDF/A validation, or a presentation that took place in summer of 2015.

Furthermore, the veraPDF consortium has been active promoting its project through articles, participation at conferences, by holding open consultation sessions with the respective memberships of the organisations which cooperate in the veraPDF project.

With regard to the community building of the consortium, one particular aspect focuses on outreach to memory institutions. Under the "community" tab of the project website, you read that the activities are focused on the gathering of policy and functional requirements, the building of collections of test files, and on the building of a large community of practice around the veraPDF to ensure long term sustainability. Here, we find two groups of interest to the consortium.

One group is the memory institutions. They are invited to provide test files, to carry out early testing of software releases and to cooperate with the consortium in collecting requirements which may be operational in the development of the policy checker component of the tool.

The other group consists of those interested in the building of a community of practice; that is either professional or layman software developers. In fact, under the "software" tab at the web site, you are provided with instructions on how to contribute software to the project (given that contributions pass the screening of the project owner). The information is provided under three separate headlines:

---

[76] Deliverable D4.3, p. 9. This also means that the third and fourth activities are connected to each other.

[77] http://verapdf.org/

- First, those interested in the project are made aware of the software releases. Here, developers are instructed on topics such as versioning, the current status of releases, and are pointed to where they can find information about future plans.

- Second, interested persons are directed through links to the project's GitHub page.[78] Part of the information given to possible contributors under each GitHub project is a README-file, which explains what a particular folder contains, and what the purpose of the work is. Going back to the "community tab" on the project web site, you find a beginner's guide about all of the ongoing veraPDF development activities, which are organised into two GitHub projects.[79]

- Third, there is information about software quality. The goal of the project can not be reached unless software development best practices are used, the veraPDF consortium explains – it being the development of the "definitive PDF/A Validator". Here, the consortium mentions three types of platforms which are being used in developing the validator. The first is a platform used to ensure quality assurance of code (Sonar[80]). The second is the Travis-CI service[81] (CI standing for continuous integration), which is being used as work tool to organise, coordinate and sync work being done in the GitHub project. The third is the Jenkins open source integration server, which the veraPDF uses to facilitate goals such as testing and continuous deployment[82].

This means that the veraPDF consortium has provided the means whereby interested parties can follow on-going activities. Target groups have been identified, and provided with specific wish lists on what they can offer. Developers are provided with links if they want to make contributions. The web presence of the veraPDF consortium is found at very different sites, however, a fact which has caused early concern on behalf of the PREFORMA Consortium: Instructions ought to be found at just one place, clearly gathered in one single location.

4.3.2 MediaArea

The MediaArea team explains that it has made frequent efforts to disseminate project findings and results and to facilitate further collaboration. One area where such collaboration is found, regards the standardisation efforts, an activity which the company undertakes to create results which may be of broad interest to the public.

Noticeable at first is the web page which MediaArea has set up for its open source project.[83] Here the project is presented as an implementation checker, policy checker and reporter for

---

[78] https://github.com/verapdf

[79] http://verapdf.org/software/

[80] http://www.sonarqube.org/

[81] https://travis-ci.org/

[82] http://jenkins-ci.org/

[83] https://mediaarea.net/MediaConch/

Matroska, FFV1 and PCM[84]. The information about the project is organised in three tabs. In the first ("about"), interested parties can learn about the project, its team and community, and get updates by reading recent posts in a blog. The second tab ("documentation") hosts information meant for users, including instructions on how to download and use the software, whereas the third ("software") is designed to demonstrate various features of the conformance checking. Those interested can also register with the project, to receive updated news via e-mail.

A number of dissemination activities are addressed by MediaArea in the Intermediate and Final reports to the PREFORMA consortium.

Several presentations have been made at conferences and meetings, for instance at the International Association of Sound and Audiovisual Archives (IASA)[85] in autumn 2015, and the Association of Moving Image Archivists, AMIA in autumn 2015. For a full list of presentations, see deliverable D4.4.

Collaborations have been established with various institutions, such as the UK National Archive and Indiana University. MediaArea has joined the IETF-Dispatch list hosted by the UK National Archive to view and contribute to the active discussions about the future of Matroska and FFVI for archival standards. According to the Intermediate Report, contacts have been initiated with the Indiana University in the USA. This collaboration is motivated by the PREFORMA project, and the university is very interested in following the development of MediaConch and its use of Matroska-wrapped FFV1 files in a large scale video digitization project. A major emphasis in this particular field of outreach is standardisation. MediaArea has established a working group charter with the IETF (Internet Engineering Task Force).

Those interested in making contributions to the project, can utilize the MediaConch GitHub platform.[86] The MediaConch README presents the purpose of the project. For those who want to follow on-going activities, release dates are published. A guide has also been developed for those wishing to make contributions. It states that MediaArea welcomes and encourages open source contributions to the software throughout the development stage. The team also instructs interested developers that all "…contributions should be clear, concise, and follow the standardised, applicable coding and naming conventions within the project's style guidelines."[87].

During the feedback process following the immediate and final releases of the 1st part of the Prototyping phase, the PREFORMA Consortium determined that the overall measures to disseminate the project and build an associated community were good.

4.3.3 EasyInnova

EasyInnova has developed a strategy to build a community around the DPF Manager open source tool. This strategy has several components to it. First of all, the creation of a community

---

[84] See PREFORMA Challenge Brief for more information about the file formats mentioned, p.8

[85] http://www.iasa-web.org/

[86] https://github.com/MediaArea/MediaConch

[87] https://github.com/MediaArea/MediaConch

around the TI/A initiative.[88] This initiative (Tagged Image for Archival) is a standards initiative in which a group of experts seek to focus on the definition of a specification built on the TIFF file format.[89] This community is building quickly with strong interest around it. EasyInnova anticipates that the DPF Manager will be the first tool that can validate this new format.

The TI/A community is built around three online channels: the website (see above); a Twitter account as well as an intranet service for the involved experts. The EasyInnova website has acquired a certain number of visitors, accompanied also by a good following on Twitter. A significant number of experts have registered on the intranet, and the discussion among these is focused on the first draft of the TI/A specification.

The company's dissemination activities also include newsletters, publishing a white paper to raise awareness in the scientific community, and presentations at various conferences.

For those who wish to make contributions, thus being active in the open source community around the tool under development, there are instructions and links available at the website dedicated to the project.[90] Activities are organised under three headlines: discussion, learning and development. Under the first headline ("discussion") notable activities include the developers' forum and the GitHub issues list for those who wish to take part in discussions about DPF Manager features and bugs. The user documentation and the API documentation is also available on the website ("learning"). For those who want to take an active part in software development, a contributor's guide has been published. Potential community members can find a link to the project's GitHub page.[91]

First instructions to contributors are directed to those who either have identified an issue, or wish to develop a new feature. The next step is that contributors acquire basic information and knowledge about what it means to work on GitHub (a link to an available resource, the git-scm book is provided[92]); what it means to make a pull request[93], and learn how they can build and test software. Additional instructions are also available covering other topics.

The PREFORMA Consortium commented in the feedback process to the Final Report on the several instruments set in place. Overall, measures taken by EasyInnova to disseminate and build a community are good.

4.4   ACTIVITY 4: OPEN SOURCE APPROACH

How suppliers are addressing relevant open source topics, best practices and licensing, as well as how they are facilitating the building of open source communities around their tools, is one of

---

[88] http://www.ti-a.org/

[89] For further information, see the section in this chapter on standardisation efforts.

[90] http://www.dpfmanager.org/community.html

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

[92] http://git-scm.com/book/en/v2

[93] https://help.github.com/articles/using-pull-requests/

the major aspects of the intermediate and final releases. An open source approach is in fact supposed to characterize all of the work of the suppliers, whether it concerns development of software, testing, licensing, deployment or a preparedness to accept and assess contributions from members of their respective communities.

The entire work process of the suppliers is supposed to be built on open source best practices, and the PREFORMA Consortium has, therefore, communicated open source related requirements in different documents which can be subdivided into two groups: a) the Call documents of last year, more specifically the Invitation to Tender, Challenge Brief and Framework Agreement, b) the deliverables 2.1 and 4.3. In addition to this, certain issues pertaining to open source requirements have been addressed through the supplier meeting structure. How suppliers should concretise these requirements was a main topic during the feedback process of July and October.

4.4.1   veraPDF consortium

In the Immediate Report, the veraPDF consortium lists three open platforms when asked how they operate according to open source best practices (one of these platforms reoccurs in their Final Report): GitHub[94], Travis-CI and Jenkins. These platforms were listed, and chosen in the first place, because they are used to create a continuous deployment and release environment so that software is available as soon as it is ready.[95]

On the platforms we find correspondence between requirements made to suppliers, for instance those that involve frequent releases, assets for using different versions, nightly builds, issue and bug tracking, provision of documentation, instructions for those who want to contribute and so on. Below are the themes addressed by the veraPDF consortium in their reports.

First, the veraPDF consortium explains that it uses GitHub for revision control (or version control[96]), meaning that it is used to track and control changes to a project's files, in particular to source code and documentation, but in practice to attain control over each change made to a particular part of the software, whether from project owners and their developers or potential external contributors. Associated with version control systems is a certain terminology. A few of these concepts are used by this supplier in their reports:

- Release branches for patching public releases (a branch is a line of development, a copy of the project, under version control but isolated so that changes made to the branch do not affect other branches of the project and vice versa, except when changes are deliberately merged from one branch to another);

---

[94] https://github.com/verapdf

[95] http://verapdf.org/software/

[96] Although revision control, version control or change control often are used interchangeably, there is a difference of meaning between revision, meaning changes to an edition of something, and version. Versions are of different kinds: development versions, and official versions or releases (se for instance: http://verapdf.org/software/ under prototyping, where versioning is mentioned.

- All contributions as pull requests (a pull request is a request from a contributor, internal or external, to the project that a certain change to the code might be "pulled" into the project).

The veraPDF consortium also lists the use of the Travis-CI platform for continuous integration[97] (practices concerning merging or integrating contributions), more specifically it s use to test pull requests before they are merged into a particular branch.[98] When looking at this platform, you can get an overview of the full build history in the veraPDF library, showing how various pull requests or contributions have been merged into releases or versions.[99]

Thirdly, veraPDF consortium reports that "…after a successful build and test on Travis a pull request can be merged into the integration branch of the main repository", and names the two developers which have the authority to carry out this task (Carl Wilson and Timor Kamalov). Once a change is merged and pushed to GitHub, it is handled by Jenkins as a configuration management system. By means of the Jenkins tool, all development versions of the software are made available as zip files.[100]

veraPDF consortium also mentions the dual licensing requested by PREFORMA, and that it has made efforts to build a technical community (Intermediate Report); another phrase for an open source community.

In the Final Report further attempts to build a community around the software is mentioned. Since beginning software development in April 2015, the veraPDF consortium has focused on establishing the technical foundations for building an open source community. One of the reasons, according to the veraPDF consortium, is that "without these foundations, dealing with external contributions is labour intensive and error prone". The veraPDF consortium also wanted to establish the architecture and a working and tested code base. "Now that the infrastructure is in place, there is opportunity to make regular software releases" – a strategy which they believe is an effective way or raising awareness.

The Intermediate Report also mentions how the veraPDF consortium has implemented the open source related requirements on licensing. Project licensing is highlighted on the project home page (bottom of the page)[101] with its own section also on the website.[102] Vera PDF makes further comments on the issue by explaining that all active Java code in GitHub projects has dual license terms at the top of the README-files, and two license files in the project root. The

---

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

[98] https://docs.travis-ci.com/user/pull-requests

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

[100] http://downloads.verapdf.org/dev/ - see also http://verapdf.org/software/

[101] http://verapdf.org/home/

[102] http://verapdf.org/home/#licensing

validation profiles and corpus GitHub projects have clear Creative Commons licenses (4.0) in the READMEs.[103]

The final item on the open source approach related reporting of the consortium is documentation. Installation instructions on GitHub are outlined. User documentation in the form of a Quick Start Guide for end-users of the Graphic User Interface has been published. Developer documentation is available through the Maven reporter.[104]The developers also use this tool to publish site documentation through the Apache Maven site plug-in.[105]

4.4.2 MediaArea

In the Immediate Report, MediaArea reports that all of the MediaConch code is open and available on GitHub for anyone to review or comment. The MediaArea team monitors the GitHub Issues page, used for soliciting community feedback, for feature requests or other comments. Furthermore, the company mentions that it is established in the open source community via the MediaInfo software[106] and explains that MediaConch can be seen as an extended version of this software.

MediaArea has relicensed MediaInfo to comply with PREFORMA licensing requirements, The company reports further that Matroska recently adjusted their EBML license[107] to comply with the required PREFORMA standards.

In the Final Report, MediaArea explains that it uses GitHub´s features in order to secure a transparent development, social networking and collaboration, and that it can be used to foster feedback, discussion and dialogue with project stakeholders.[108] The activities on GitHub have been divided by MediaArea into several different repositories to focus on source code and project management, test file compilation and metadata definition.

The licensing for MediaConch, its source code, test files and documentation, is published on the project web page.[109]

4.4.3 EasyInnova

In the Final Report, which accompanied its October release, EasyInnova reported that setting up the DPF Manager Community website and defining tools and methodology were the first steps to start establishing a community around the DPF. At the end of the Re-design phase the

---

[103]See for instance: https://github.com/veraPDF/veraPDF-validation-profiles/blob/integration/README.md

[104] https://maven.apache.org/

[105] https://maven.apache.org/plugins/maven-site-plugin/

[106] https://mediaarea.net/en/MediaInfo

[107] http://www.matroska.org/node/47

[108] https://github.com/MediaArea/MediaConch

[109] https://mediaarea.net/MediaConch/about.html

company will start to promote the project in order to find developers interested in the area of digital preservation and who are willing to collaborate on the DPF Manager software.

In order to establish a community of developers, a set of methodologies and tools to facilitate the development has been set up by EasyInnova:

- First, GitHub is used as a "public control version repository". EasyInnova reports that it uses and takes advantage of all of the tools offered by GitHub. The GitHub issue tracker is used as a way to submit bugs, ask for new features or to receive requests by those who wish to make any contributions. The GitHub milestone is updated to follow the project development route map.

- Second, EasyInnova reports that it uses Maven as a "build automation tool".[110] Further uses of Maven mentioned by the company are: testing applications from the open source code, build, and management of reporting and documentation.

- Third, EasyInnova explains that the Travis-CI tool makes up part of its infrastructure that it uses to facilitate an open source community. By using this CI (continuous integration) tool, the developers can validate any contribution (push or pull request) to the project and ensure that these follow the code quality requirements in place.

- Lastly, the company says that in each monthly release, the source code and build environment is released as an executable ready to use even without the Internet, following guidelines from the D4.3 "*Functions of the Open Source Portal*" deliverable. (Licensing issues are brought up by the company in conjunction with its software releases.)

4.4.4 Feedback Process

Following the overview above of how the Suppliers have addressed open source topics, best practices, and reported on their progress in setting up open source communities, this section will briefly focus on the feedback which has been given with regard to open source.

Some of the issues brought to the attention of the suppliers, in the first place by the PREFORMA partner focusing on open source topics, have been discussed in supplier meetings (see chapter 2 "Formal Procedures"), while others are pending discussion and further deliberation in either this particular structure or the project's management structure.

The issues raised are:

- Issue # 1 – Provision of source code

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

- Issue # 3 – Time based provision of stable releases

- Issue # 4 – Identical software under both open source licenses (GPLv3 or later, MPLv2 or later)

---

[110] https://maven.apache.org/

- Issue # 5 – Provision of executable of the software on the open source portal

- Issue # 6 – Provision of executable of the software for use via web browsers

- Issue # 7 – Provision of detailed documentation concerning interpretation of the technical specification of each file format

- Issue # 8 – Provision of software which can be redistributed in a cascade (or by any individual)

Here, concerns were raised about issues which touch on several activities by the suppliers during the 1st part of the Prototyping phase, while laying considerable emphasis on software deployment (all issues pertaining to software release, installation, updates, versioning etc.) and licensing. Issues regarding documentation are also brought up.

The basis for the feedback is the deliverable D4.3 "*Functions of the Open Source Portal*", a document which is final in its 2.0 version (after the outcome of the PREFORMA review). The feedback is also based on observations that made during follow-up of the activities of the suppliers during the Prototyping phase. These observations are further developed in deliverable D8.8 "*Monitoring of open source project implementations*".

As previously mentioned, most of these issues are still pending further discussion and deliberation, but further insights to the requirements and how they relate to the overall workflow of the suppliers will be addressed in subsequent chapters.

One of the issues raised, when discussing the final reports, was that PREFORMA needs to change the method of making the software available, i.e. linking from the PREFORMA Open Source Portal to the relevant pages of the suppliers' websites/GitHub repositories where it is possible to download the packages.

In order to fulfill the PREFORMA requirements, the suppliers need to provide PREFORMA with all the zip files, both of the executables and of the source code (and when it will be available to complete the build environment), for all the various platforms (Mac, Win, different versions of Linux, as described in deliverable D4.3). It is fine of course that the suppliers also maintain them in their GitHub/website, but PREFORMA need an official/formal external distribution.

For this reason the suppliers websites were restructured in the Open Source Portal creating a sub-page where the periodically released zip files will be uploaded. This also answers to some of the comments of the PREFORMA reviewers, who were rather confused by the big number of links available in the Open Source Portal.

4.5   ACTIVITY 5: STANDARDISATION EFFORTS

Part of the PREFORMA Challenge is that the selected suppliers should engage in interacting with relevant standardisation organisations. Specifically, the interaction is supposed to be with those organisations that maintain the standard specifications used by the project. According to

the Invitation to Tender (ITT), the aim is "… to provide feedback, resolve technical issues and contribute to a dialogue for improvement of the technical specifications of standards".[111]

In the Challenge Brief, the issue of the suppliers´ interaction with standardisation bodies is further associated with the reference implementation. The conformance checkers should authorize an implementation of a standard specification that is to be used in all sectors as a "definitive interpretation" for that particular specific technical specification.[112]

Specifically, the standardisation efforts of the suppliers are supposed to be carried out through interplay with their respective open source projects. This is brought out in deliverable D2.1 *"Overall Roadmap"* of PREFORMA by explaining that the suppliers should establish a proactive approach for establishing effective process for interaction with the organisation that controls the file format. A role is sought after which may lead to resolution of issues in the specification of the file format in order to prevent further files to be generated in an incorrect way.[113]

This theme is supposed to characterize each of the three open source projects. This is why the suppliers are frequently asked to provide descriptions of what they are actively contributing to the standardisation process in their domain.

4.5.1 veraPDF

In its Immediate Report, the veraPDF consortium reports that they have performed a number of actions in relation to the ISO working groups (WG) that are working on standard specifications for the PDF/A format. Specifically, veraPDF mentions two WGs (ISO TC 171 SC 2 WG 5 for PDF/A – this working group "owns" the format – and WG 8 for PDF), both maintained by the International Standardisation Organisation, ISO. Since initiating the Prototyping phase, the following actions have been taken by the veraPDF consortium:

- Established WG policy on the point that existing parts of the PDF/A will not be amended via corrigenda or otherwise;

- Established that the PDF Validation Technical Working Group (TWG) may, as a body, supply proposed revisions to working text;

- Submitted proposed enhancements for a new part of PDF/A addressing ambiguities in existing specifications;

- Led the effort to initiate a new "PDF/A next" project in WG 5. The first formal ISO meeting to address PDF/A next was scheduled for late 2015.

One of the main actions is that the veraPDF consortium will develop a proposal to ISO to request permission to establish the PDF Validation TWG[114] and the veraPDF software it approves as normative references for "PDF/A next". The main background for initiating the

---

[111] Invitation to Tender, p.16

[112] PREFROMA Challenge Brief, p.14

[113] Deliverable D2.1, p.7.

[114] The TWG is hosted by the PDF Association: http://www.pdfa.org/pdf-association/

PDF/A next project is that the ISO WG 5 has resolved that the existing PDF/A specifications will not change; there will be no updates or corrigenda to ISO 19005-1, -2 or -3. This means that reputed long-standing ambiguities with regard to the standard specification have to be resolved in other ways. "PDF/A next" is the next step the ISO WG is taking in the development of the PDF/A specification, and is the current term of business for the WG.

The need for this next part of the PDF/A has arisen, in large part, as a function of veraPDF consortium activities, including the deliberations of the PDF Validation TWG. In fact, part of the normative references needed for the "next" project will come from the veraPDF software.

According to the PREFORMA Consortium's response to the Intermediate Report, the veraPDF consortium is well represented in the current development with regard to the PDF/A specifications, and they regard their degree of interaction with the relevant standardisation bodies as both high and grounded in current practice.

Further, the veraPDF consortium invites PREFORMA Consortium to discuss the issues of standardisation in order for both parties to agree on what may be the expectations in this regard.

The theme of standardisation is raised again in the Final Report of October. Here the veraPDF consortium reports that it seeks to facilitate collaboration between the PDF Validation TWG and ISO WG and in particular to achieve the objective of having the TWG and ISO WG jointly establish a mechanism for resolution of current technical questions enabling test file and software development by veraPDF to proceed on that basis.

4.5.2 MediaArea

Concerning the activities of the MediaArea Company, with regard to development of standards, they communicate that close collaboration is taking place with both the Matroska and FFV1 communities to standardize both formats. The Matroska wrapper is an open standard, which is supported by a non-profit organisation in France. In a presentation at a conference in Prague in July, MediaArea put forward a charter, stating that the

> "Organizations tasked with the preservation of audiovisual materials (such as governments, cultural heritage institutions, media organization faces obsolescence challenges as videotape technology is diminishing). They must digitize now to preserve materials and seek to use open, transparent, self descriptive, lossless formats, and there is an important role to be filled by open source standards. The PREFORMA project has selected FFV1 and Matroska as open formats to develop preservation tools around, but these formats are both in need of more formal standardization efforts".

MediaArea reports further that team members have been initiating conversations with contributors and interested parties via their GitHub page and through an email-list. The background is that the MediaArea project has been active in analysing existing standards for both formats, and has been active by presenting proposals for change, requesting feed-back from the communities.

Following the IETF conference, MediaArea spoke with working group chairs and area directors about the prospect of bringing both Matroska and FFV1 into the review process for standardisation which is hosted by the Internet Engineering Task Force (IETF).[115] The Intermediate Report states that both of the leading authors of Matroska and FFV1 stated their support publicly about such a development.

In their Final Report of October 2015, MediaArea reports on steps taken to form a working group around the open source formats, called CELLAR (Codec Encoding for Lossless Archiving and Real-time transmission).[116] After undergoing a review process by the Internet Engineering Steering Group (IESG), the IESG approved the CELLAR working group charter on October 22, meaning that the WG-process could begin.

The goal of CELLAR is to formalise specifications for the Matroska, FFV1 and FLAC formats. In order to provide standardised specifications for users and developers, the WG will seek consensus through a process of refining and formalising these standards. The specifications and GitHub-based development versions are made available at the CELLAR webpage, accompanied by a series of planned deliverables with already set dates. A Chair, Area Director and a mailing list for the WG have been established and the first face-to-face meeting is planned for June in Berlin.

The MediaArea team has also been actively participating in other discussions regarding the formats e.g. as pull requests and issues raised on GitHub and other issues raised through the Matroska mailing list. One issue of interest here is metadata format specification development. To further this work, MediaArea has been using a repository at GitHub set up to develop XML Schemas and documentation for the reporting formats used in the MediaConch project.

What kind of feedback has the PREFORMA Consortium provided to these developments by MediaArea? One note concerns the need for MediaArea to design their standardisation efforts based on current practice concerning how stakeholders may be involved and engage in open source projects and standardisation projects. Another note simply reports that the standardisation efforts of MediaArea seem to be fruitful and feasible given the time, noting especially that the Charter of Cellar has been approved, giving a clear signal for the working group to begin its work.

Further information about the standardisation efforts currently being carried out in CELLAR, as well through other channels, is continuously made available on the PREFORMA website.[117]

4.5.3   EasyInnova

EasyInnova started its TIFF-related open source project with the goal to create a new standard for long term digital still image preservation, based on the TIFF format. This was initiated as a result of advice received from different TIFF and digital preservation experts, from the University of Basel as well as from memory institutions.

---

[115] http://www.ietf.org/

[116] http://datatracker.ietf.org/wg/cellar/charter/

[117] http://www.preforma-project.eu/mediaconch.html

The new standard was originally named TIFF/A, with A standing for Archive, but due to copyright reasons it is now named TI/A. In their reports to the PREFORMA Consortium, EasyInnova has reported on a number of actions related to this new initiative. The purpose of TI/A is not to create a completely new image file format standard - the "new one" is defined on an existing standard. The ambition of the project is rather that TI/A will function as a version of the TIFF format that may be more suitable for long-term digital preservation.

According to a White Paper published on the subject by researchers from the University of Basel, the proposal includes a subset of TIFF which is compatible with the TIFF standard, but with certain measures in order to guarantee a correct rendering in the future.[118]

EasyInnova has undertaken the following in an effort to establish a new version of the TIFF file format standard:

- EasyInnova reports that it has joined AENOR, the Spanish standardisation body, to seek partnership in a working group responsible for digital formats.[119] (CNT50/SC1). As a standardisation body, AENOR seeks to improve and develop technical standards as well as work with issues pertaining to certification. EasyInnova reports that the members of this WG started informal discussions with other standardisation bodies on their behalf at the ISO/TC46 annual meeting that took place in June 2015.[120]

- EasyInnova reports that their partners at the University of Basel have joined the Swiss standardisation body, SVN, to lead the effort to create the new TI/A. Both AENOR and SVN are participating in their role as standardisation bodies in the work of ISO/TC 171. At the time of writing the Final Report, EasyInnova was accepted to participate in a meeting with the TC in order to secure approval and participation of the committee to create this new version of the standard.

- EasyInnova produced a draft specification. The objective is to get a core group of experts to further develop the standard based on the draft, and a group of interested people to provide feedback on it.

- A website for the TI/A initiative has been set up, to allow any interested party to follow the development of the standard. Here it is being reported that a submission to ISO for a new version of the TIFF file format standard will take place on March 1st 2016.

- EasyInnova have been in contact with Adobe, the company which owns the copyrights to the TIFF specification, but has not been given the right to use the TIFF acronym as part of the proposed name for the new standard version (TIFF/A). Therefore, the name has been changed to TI/A.

---

[118] http://www.ti-a.org/TIFF-A%20white%20paper.pdf

[119] http://www.en.aenor.es/aenor/aenor/perfil/perfil.asp#.VnEIMErhDIU

[120]

http://www.iso.org/iso/home/standards_development/list_of_iso_technical_committees/iso_technical_committee.htm?commid=48750

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

What kind of feedback have these activities spurred from the PREFORMA Consortium? First and foremost, there is approval among most members in the Consortium, formulated in words such as that the standardisation effort seems to be on track, and managed well according to expectations (Final Report). However, EasyInnova was advised that the standardisation efforts be grounded in current practice concerning how different stakeholders may be involved and engaged in open source projects and standardisation projects.

5   POINTS ON PROGRESS

This chapter summarises the progress that has been made throughout the 1st part of the Prototyping Phase. What are the visible, noticeable strong points of the development work?

There are several checkpoints here:

- the functional and technical specifications of the Design phase;
- the requirements communicated by PREFORMA;
- the development between the intermediate releases and reports in July and the final releases and reports in October 2015.

However, it is important to state, that an evaluation of any sort will not be undertaken in this chapter, and unless the issue at hand lends itself to comparison, the Suppliers will not be set against each other.

Six progress points are brought up:

1. Capabilities for software;
2. Evolving functionalities;
3. Usability;
4. Testing for quality assurance and accuracy;
5. Achieving reference Implementation;
6. Awareness of what is still missing.

The chapter concludes by examining the critical factor of compliance, focusing particularly on the documents which are directional for the Prototyping phase (see section 1.3), as well as on issues brought to the attention of the PREFORMA Consortium by one of the PREFORMA partners, and which have similar application to all three suppliers.

5.1   CAPABILITIES FOR SOFTWARE RELEASES

A primary observation of the 1st part of the Prototyping Phase is the releases: the sheer magnitude of them, their different status (development versions, stable versions and deployment versions), and the work that the suppliers have done to secure at least satisfactory quality. In other words, the suppliers have built substantial capacities for software releases, in order to accommodate the requirement of PREFORMA for frequent[121], monthly[122] releases. This is an interesting feature of the work of the suppliers, as well as a strong point.[123] In their

---

[121] Cf. PREFORMA ITT, and Challenge Brief.

[122] Cf. Deliverable 4.3.

[123] Some concept explanations: **Continuous Integration** refers to integrating, building, and testing code within the development environment. **Continuous Delivery** builds on this, dealing with the final stages required for production deployment. **Continuous Deployment** means that every change to code

Intermediate Report, the veraPDF consortium was asked to provide a description of how they address the relevant open source topics, best practices and licensing. They replied by a list of three particular systems in use: GitHub for revision control, Travis-CI for continuous integration, and Jenkins, an extensible open source server, for continuous deployment. The same information is mirrored more or less in the final report, now under the headline "source management" for GitHub.

The veraPDF consortium explains its use of the Jenkins tool further in the Final Report, by pointing that it believes that PREFORMA requirements concerning build environments can be satisfied by the use of automated tools and existing package management systems. Here, the veraPDF consortium refers to open source related requirements concerning the simultaneous release of the executable with its build environment (open source tools used to create the executable by compiling source code), meaning that the end-user should have access to all necessary files to run the application in the downloadable package.[124](There are more aspects to this discussion.)

MediaArea mentions the use of GitHub as an open development platform in its Intermediate Report. In the Final Report, the company explains further. Here, this platform is reported to be used for purposes such as source code management, project management, test file compilation, and metadata definition. In its functional and technical specification report, the MediaArea team provides further insights into its release capacities. Here, MediaArea explains that GitHub will be used to extract nightly builds of the code and deploy using continuous integration. The subsequent builds and software release can then be downloaded, built and run on any system.[125]

EasyInnova communicates use of GitHub not only as a "public control repository" (meaning that is secures availability to source code under the open source licenses) but also to take advantage of other functions of the platform such as issue trackers in order for contributors to follow the public development route map. Further, the company explains that it uses Maven as an automation tool, meaning that it is used to build and test the application from the open source code. EasyInnova uses Maven in conjunction with Travis-CI, their continuous integration platform to validate any contribution to the project and to ensure that such follows the code quality requirements.

The software release capabilities of the suppliers are much in line with current methodology and major trends in software release management.[126]

---

automatically gets put into production, resulting in many potential deployments every day. **Continuous Delivery** means that you are able to do frequent deployments but may choose not to do it, usually due to preferences for a slower rate of deployment. Continuous Deployment is necessitated by Continuous Delivery.

[124] Ibid.

[125] Conch – conformance checking for audiovisual files, p.28.

[126]   http://www.cmcrossroads.com/article/three-major-trends-software-release-management-you-should-adopt

The first trend is the adoption of agile release practices including automation (the veraPDF consortium mentions automated tools in the Final Report). A main reason is that developers are to focus on core development, leaving most of the deployment pipeline to be more or less automated. Quality checkpoints are included in the process and developers are notified only when something fails. As the software passes these checkpoints, new version releases are automatically pushed to the release repository from which new versions can be downloaded by end-users. There is still some manual work being done, but the idea is that software teams are to be relieved from the time-consuming manual configuration management tasks so they can focus on code and core product development. Continuous deployment is a concept associated with this.

A second trend is the use of version control systems such as GitHub. The suppliers communicate several uses of the GitHub open platform in their reports.

A third trend is continuous integration, or the merging of contributions. Each integrated contribution is verified by an automated build tool (including testing). The utilization of integration tools like Travis-CI is to advance the process of continuous delivery/deployment (mentioned by each of the Suppliers.)

The capacity building mentioned did not spur any comments from the PREFORMA Consortium – expect for the issue of time based provision of stable releases. It is unclear to some partners, whether the suppliers have exposed their monthly versions to a certain level of quality assurance. A background for the comment is that the deliverable D4.3 requires that suppliers do precisely that: monthly releases should be stable (meaning assured of their quality). Ambiguities in this regard had to do with concerns about lack of detailed information.

5.2 EVOLVING FUNCTIONALITIES

The main observation based on the reporting structure, is that the functionalities of the software are evolving during the 1st part of the Prototyping phase, and that suppliers have well defined plans how to progress further in the next phase. This is a strong aspect of the suppliers´ work. As such it is related to the overall infrastructure of the conformance checker, which was made public in the Call documents, to the functional specifications of the Design phase and to the ongoing feedback process involving both suppliers and the PREFORMA Consortium.

The veraPDF consortium shows proof of evolving functionalities. Examples from the release notes which accompanied the 0.2, 0.4, 0.6 releases (launched during the 1st part of the Prototyping phase) show us first of all, that specifications covering the ISO-standards have been implemented in code by increasing rate and quality. The 0.2 intermediate release covers implementation of a number of rules, while the 0.4 release had certain new rules being added, a development that seems to have reached stable conditions by the time of the 0.6 release. Secondly, we see work evolving with respect to the other components of the conformance checker as well, such as reporting and metadata fixing. The veraPDF software can be downloaded for all of the required operating systems, and users can choose between a CLI and a GUI interface. Although this is a sure indication of evolving functionalities the release notes do not allow us to outline it (the CLI and GUI interfaces were available with the 0.1 version already).

There is a positive relationship between the work of the veraPDF consortium, and the functional specifications that was submitted in the Design phase. Here, the conformance checker's

components are patterned by functional descriptions, functional architectures and use cases which provide a detailed overview of the process to be designed and developed. The development of the interfaces and their inter-relationship are described too.

The development of the veraPDF tool is based on the technical specifications, which were submitted in the Design Phase. Future development of the tool hinges on principles such as modularity, reliability, simplicity and the use of open standards. The long-term preservation ambition of PREFORMA is mirrored by an ambition to keep the design of the tool as simple as possible, so it is accessible to users and developers alike.

The next steps to be taken by the veraPDF consortium focus on the status of the functional components of the conformance checker and the user and developer interfaces. A noticeable fact regards the policy checker component. Progress on this component has been slower than first anticipated, the veraPDF consortium reports. Thus far, progress has been hindered by a lack (so far) of real policy requirements from the PREFORMA Consortium. Those interested in the continued progress of the veraPDF consortium are provided by the supplier with percentage checkpoints, which outline plans for future development.

MediaArea's MediaConch tool exhibits evolving functionalities, especially as steps from the early 15.0-releases to the 15.09 and 15.10 releases of September and October were taken. In September, the 15.09 release included the launch of MediaConchOnline, a web interface which allows implementation checking without the user having to download an installer. Echoing one of the main design ideas of the MediaArea team, a user can generate policy sets on their own for reporting on files. Another feature of the 15.09 release was the integration of a MediaTrace XML Reporting Tool. The 15.10 release features some brand new implementation checking functionalities which focus on Matroska and EBML. Apart from this, metadata development was a major theme for the MediaArea team in this phase as it saw the launch of four separate conformance checking schemas for reporting, which in the case of the MediaConch XML allowed for a presentation of a file's metadata elements in the form of easily intelligible sets of track information. The MediaConch software is available in the form of CLI and GUI. It can also be downloaded at all of the major operating systems.

The functional specifications of MediaArea form a blueprint for the achievements that have been made during the Prototyping phase, especially the functional requirements whereby each component of the tool is outlined. The team has noticed advantages of collaboration between suppliers and foresees an end-product where interfaces are made with the other conformance checkers – a priority which has not reoccurred in the reporting structure. The team has prioritized the implementation checker in both the Design phase and 1st part of the Prototyping phase. The registry of checks, which is mentioned in the functional specifications, is addressed in the Final Report.

The next steps to be taken by the MediaArea team include further work on all of the functional components. The team is convinced that the implementation checker module is stable. The policy checker component if the MediaConch is according to the team operational in the sense that testing has been productive. Currently the policy check samples only use data provided by the suppliers. The reporter aspect of the project has seen much progress with development of XML Schemas. Plans for the future include expansion of reporting to "various refined reports (related to the above mentioned schemas) that focus on various objectives and use cases". The metadata fixer component is not fully operational, and is high on the list of future activities.

The EasyInnova efforts have resulted in evolving functionalities. The intermediate release of July 2015 (1.0) showed the capabilities of the DPF Manager conformance checker to read TIFF files, to detect tags and handle embedded metadata. In its first release, the tool was able to validate the TIFF Baseline specification as well as the TIFF/EP. Other features of the early release included a prototype reporter and metadata fixer modules. The release (1.1.1) of October now included all of the missing features in the intermediate release, meaning that the conformance checker was able to take on the TIFF/IT standard as well. The release also included the graphical user interface. The policy checker module can be used by the memory institutions to create custom rules based on their own requirements (this module is the one of the five well-publicised conformance checker components).

The evolving functionalities seem to be in line with the functional and technical specifications of the Design phase. Here are both functional as well as so-called non-functional requirements (such as interoperability, modularity and deployment) outlined. Analysis of the reports does not indicate any major deviations from the course set in Design phase. Plans for the next phase include a re-evaluation of the current design and architecture. The goal is for the DPF Manager to be able to process multiple checks at the same time. EasyInnova also plans to make a major effort with regard to the interoperability issue. According to them, a common API is a requirement to start the development of some missing functionalities such as the server interface (one of the forms of deployment mentioned in the Challenge Brief) and the DPF Manager Web version.

5.3 USABILITY

The main observation is that the suppliers have provided releases that show proof of usability, albeit in various degrees, and users have been provided with guides to get started. The phrase "usability" is not mentioned in the PREFORMA documents and only on a few instances by the suppliers in their functional and technical specifications, but was an undercurrent theme of the feedback process during the 1st part of the Prototyping phase. Through it, suppliers have received reports that they can incorporate in future design and development.

According to the ISO 9241 standard on human-computer interaction, usability is defined as:

> "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use".

Basically, usability means that suppliers and the PREFORMA Consortium alike give thought to how and why people use a product. Feedback through evaluation, even through the means of simple user tests, can provide reports that supplement software testing. Open source software which is most frequently referenced as a means to correct validation of preservation files is simply an application that is used the most, first and foremost by the memory institutions of the PREFORMA Consortium, and secondly by other interested parties. Usability is a factor for many.

The veraPDF consortium has provided several releases (0.6 by the end of the phase) and has made efforts to make the tool accessible for users. The PDF/A conformance checker is implemented using the veraPDF Library, a software library that provides the functionality and APIs for PDF/A validation, policy checking, metadata fixing, and reporting. A test file corpus covering PDF/A 1b and 2b is included in the package.

The main controls of the user interface are a) "Choose PDF Button" which is used to specify the input PDF document; b) the "Choose PDF Profile Button", used to specify the validation profile,

c) the "Validate Button" which starts validating the specified PDF document for the conformance to the requirements of the specified validation profile; d) the "Generate Reports" dropdown list which specifies the types of the machine-readable reports included in the report with three available options. The "fix metadata checkbox" specifies whether the file shall be automatically fixed so that it is compliant with the standard specification. At last, several reporting options are available.

When validation is performed the restrictions from the rules are checked for the relevant objects from the PDF Document. A check may either fail or pass.

The veraPDF Desktop Graphical User Interface Quick Start Guide shows users how to install the software, and secondly how to use it. First, an introduction is made of the project, which includes a helpful glossary of terms used in conformance checking. Then, a how-to-use section follows the "installing" section.

Usability tests of the veraPDF software were positive overall and useful feedback given to the suppliers. One problem reported was that neither the public release nor the development release GUI worked on the Linux distribution Ubuntu, as well as some issues with an outcome of a conformance-checking test. However, the overall conclusion of those tests was that the application holds great potential and that it is being developed as outlined in the functional and technical specifications. At this stage, the users did not have anything to say with regard to the actual design of the application (GUI and CLI were tested), perhaps since overall functionalities were in focus instead.

The MediaArea releases are available for download at both their project website, and the open source portal of PREFORMA. The MediaConch tool currently consists of three main sections, "Checker," "Policies," and "Display." Test files are available on GitHub.[127]

In the "Checker" section, files may be checked for conformance using policies defined by the user. "Check local file" allows a user to select a file or files from a local computer. "Check online file" allows a user to select a file using a URL path. "Check local folder" allows a user to select a folder of files from a local computer or volume. Once a file or files are "checked" an implementation report declares whether a particular file is valid or non-valid according to specifications of Matroska, FFV1, and LPCM.

In the "Policies" section, a user can create policy tests as well as import previously generated policy sets in either XSL or Schematron format. Policy sets consist of individual rules and asserts. A policy may contain one or more rules, and rules may consist of one or more asserts. The "Display" section will allow a user to apply various display XSLs for use with policy and implementation check reports in the checker section. MediaConch has provided example HTML, XML and TXT displays.

The "How to Use" and "Getting Started" on the project website provides instructions for users.

---

[127] https://github.com/MediaArea/MediaConch_SampleFiles/tree/master/PolicyTestFiles

Reports from usability tests of the MediaConch software have been few, but informative. The company is perceived as a communicative and responsive Supplier. The installation has been easy, and functionalities have proved to work.

Lastly, we will report on the usability of the DPF Manager. EasyInnova releases are available at both their project website, as well as on the open source portal of PREFORMA. There are instructions on how to install either Windows or Mac or Linux versions of the software. The graphical user interface application is activated by double clicking on the desktop icon. Then, a user is invited to select (files are included in the package) or create a configuration file, with a four-step process which consists of selecting the profiles to be checked, defining the policy checker, specifying the report format/s and defining the metadata fixes and auto-fixes. Then files to process are selected, and the user can click the "Check files" button to see the results in the defined report format (HTML by default). EasyInnova has spent considerable effort into making the prototype GUI well designed and easy to use.

The DPF Manager User Manual is organised in three sections: a) overview and introduction, followed by b) installation procedure and c) application usage.

In the feedback process users reported that installation was easy. Further, issues concerning both the usability and stability of the first release (e.g. bugs) were addressed immediately with the result that most were resolved. The feedback process contains several well-detailed user reports, in itself an indication that PREFORMA Consortium members is interested in the DPF Manager, and also contributing to the further development of the tool.

To the PREFORMA Consortium it is evident that the feedback process has provided input that may be valuable to the suppliers. Continued thought about how and why the respective tools are used, as well as their being easy to use irrespective of which interface is chosen, is important as suppliers enter the Re-design phase.

5.4 TESTING FOR QUALITY ASSURANCE AND ACCURACY

The provision of datasets for the purpose of training and testing was part of the original plans for the PREFORMA project. In deliverable D2.1 "*Overall Roadmap*" a distinction is made between a) training datasets; which are aimed at driving and facilitating the design and development of software, and b) test datasets, which are aimed at evaluating and testing the software. D2.1 also makes an orthogonal distinction between synthetic and real data, where the former are data created with the purpose of pinpointing some specific compliance problem or critical issue for a given preservation format, whereas the latter are data managed by memory institutions as part of their preservation duties.

The "*Data management plan for training, testing and demonstration files in the PREFORMA project*" (see section 1.3) further exemplifies the use of training or synthetic data. According to this document, synthetic files work as a frame of reference for what the validator is to validate as a "correct" or "incorrect" implementation of a file format. In a very real sense, synthetic files can be said to represent a reference implementation of what constitutes the file format standard according to PREFORMA.

One way of validating synthetic data is via a formal mechanism consisting of experts on the particular file format. This type of data can come from several sources: the suppliers, by means of collections that have been verified by a formal mechanism[128] (or those that have proved their usability and worth through other means), from memory institutions, PREFORMA partners or other stakeholders. These files are supposed to be kept on open platforms, and also be made subject of scrutiny as reasons arise.

This particular aspect of the work of the suppliers is a point of progress, with checkpoints in a) the Design phase (plans to develop or collect test files are mentioned at several instances in the functional/technical specifications), and b) the 1st part of the Prototyping phase July-October. A good example can be found in the awareness displayed by the veraPDF consortium, but the other two suppliers show similar proofs.

Another type of testing mentioned in the reporting structure by the veraPDF consortium (first reference in the functional/technical report) and EasyInnova is unit testing. This is a type of software testing method by which individual units of source code, or sets of one or more modules or components are tested to determine whether they are fit for use. According to EasyInnova, functionalities are tested automatically through unit testing before added to the repository.

The veraPDF consortium mentions that it uses testing of this kind on compilation and enforced automatically on code check-in. The veraPDF consortium mentions integration tests also, signified by activities to test the individual behaviour of each software component, and their integration.

Mentioning these types of testing as a strong point, or progress point, is motivated given the emphasis on frequent, monthly releases in PREFORMA. Without the undertaking of measures such as those mentioned above, a continuous delivery – and deployment – of software would simply not be a feasible proposition. The PREFORMA Consortium continues to follow these developments with interest.

5.5 ACHIEVING REFERENCE IMPLEMENTATION

The PREFORMA project covers implementation of open file formats/standards in open source software, in order to create a reference implementation. Somewhat simplified, an open standard is a standard which possesses certain openness qualities; a reference implementation is an implementation of a standard which can be used as a definitive interpretation of the standard's specification. This belongs to the overarching purpose of the PREFORMA project.

An overview shows that the open source projects cover a range of open standards in the areas of text, image and audiovisual. These have been analysed in conjunction with the Bid for Tenders with regard to:

- the documents which contain the specifications of the standard;
- the organisations which maintain them;

---

[128] The PDF Validation TWG is a primary example of relevance to PREFORMA.

- the license conditions under which these standard specifications can be implemented (which are/should be open source).

A selection of standards is now being focused by the open source projects.

There can, and should, be an interplay between the open source projects and the standards adoption and implementation. One reason is that there is a vital relationship between open source projects and open standards:

- Open standards need implementations to provide: 1) confirmation of their suitability; i) a market presence; and 3) feedback from implementations and users;

- Open source development projects need guidance and direction regarding their interfaces for interoperability and portability.

Now, the on-going developments within the open standards covered in the project need to be monitored, and the open source development needs to be directed towards implementation of the specifications of the standard (as we have seen in section 4.5). This is the background of the standardisation efforts of the suppliers. The self-reporting of the suppliers indicates that a few significant steps have been taken.

A good example is the founding of the CELLAR charter by the MediaArea Company. The WG Charter states that:

"…the preservation of audiovisual materials faces challenges (…) by use of proprietary formats that lack formal open standards. The standardisation of open, transparent, self-descriptive, lossless formats remains an important mission to be undertaken by the open source community".

Interestingly, the Charter lists both the specifications of interest to the company as well as the development versions of the MediaConch software.

In order for MediaConch to provide definitive interpretations of the specifications, the WG will seek consensus and refinements for specifications for both FFV1 and Matroska in order to provide "authoritative, standardised specifications for users and developers".

The PDF/A Next and TI/A standard initiatives taken by the veraPDF consortium and the EasyInnova respectively have been mentioned in an earlier chapter. An ambition of the veraPDF is that the software development of the veraPDF validator should provide inputs of normative character to the development of this proposed standard. For EasyInnova a major strategy in reaching the project's goal is to build a community around the TI/A standard.

5.6 AWARENESS OF WHAT IS STILL MISSING

Through the reporting structure, suppliers have been requested to provide a description of the status of the work compared to that which was planned in the functional and technical specifications. They were also asked to provide updated versions of their work plans with a timeline included. The suppliers were asked to highlight what is still missing, and address how these gaps would be overcome. The main observation is that the suppliers show proof of awareness of the current status of the development and the next steps..

A good example is the veraPDF consortium's gap analysis in the Final Report. First, processes are identified and broken down under each functional component of the conformance checker. Second, a percentage number, indicating the gap between the present and the desired outcome, identifies the present outcome of these processes. Third, methods to achieve the desired outcome are listed and point out the work which needs to be done to reach the goal.

Steady progress of building a foundation for the MediaConch application is reported by the MediaArea team in the Final Report. In their gap assessment of its Design phase timeline, the company mentions activities such as standardisation, and development plans for each of the components of the conformance checker (expect for the shell). Lastly, an assessment is made with respect to which activities that must be prioritized as the Re-design phase begins.

5.7 THE CRITICAL FACTOR OF COMPLIANCE

To conclude, this section briefly addresses compliance with the requirements of PREFORMA. A couple of steering documents mentioned in section 1.3 are in focus: a) "*Legal Opinion*" (draft, date 2015-04-28), b) D4.3 "*Functions of the Open Source Portal*" (version 2.0 – FINAL). A number of issues regarding open source are also commented in particular to their current status (checked, yet to be communicated or resolved).

5.7.1 Requirements: the "Legal Opinion" document

This document purposes to:

- consolidate the legal sources of the PREFORMA project;
- provide an analysis and comments of the legal sources;
- conceptualize the project in order to identify possible problems and clarify possible concerns.

The document is to be viewed merely as advice or opinions, and should not be perceived as any binding status for the PREFORMA Consortium. The background for the writing of the *"Legal Opinion"* was as a response to a joint letter by the PREFORMA suppliers (dated 19th of February 2015). The document is divided into the sections of legal background and conceptual framework.

First of all, the *"Legal Opinion"* makes a reference to the documents which can be perceived as legal sources for the PREFORMA project.

The Framework Agreement is binding for the three suppliers that were chosen for the major phase 2 (Prototyping). In addition to the Framework Agreement, separate contracts for phase 2 were signed by the contracting authority and each of the three suppliers. In the event that the evaluation of the Prototyping phase (1st and 2nd parts) results in a selection, the Framework Agreement shall not have any effect upon the suppliers. It will expire on the date announced by the contracting authority for final award of phase 3.

The terms and conditions of the Framework Agreement that relate to intellectual property rights and license grants are given substantial place in the "*Legal Opinion*". (References: sections 17.1, 17.2, 17.3, 17.4, 17.5, 17.6, and 17.7 of the Framework Agreement.) These terms and conditions are also explained further in the "Conceptual Framework" section of the "*Legal Opinion*" (ch.12.2 "The source of the code").

The Invitation to Tender (ITT) is also discussed by the "*Legal Opinion*" document. Of course, the ITT contains a number of time-limited instructions that ceased to be binding once the Call for Bids was complete, such as the administrative instructions, but the minimum requirements and the exclusion criteria are still relevant. In the "*Legal Opinion*", those sections of the ITT which regulate distribution of intellectual property rights are highlighted by cross-reference to the Agreement.

The "*Legal Opinion*" also addresses the Challenge Brief, which sets the parameters for what each open source project is supposed to achieve long-term, and the conditions for the work of the suppliers.

A second vital point to bring from the "*Legal Opinion*" is its conceptualization of a number of issues relevant to the software releases of the suppliers (functionality, code, documentation, licenses and patent rights).

The very fact that the "*Legal Opinion*" was written in the first place is indicative of the fact that requirements do matter to the suppliers. Hence, the PREFORMA Consortium should have a preparedness to discuss and deliberate legal issues, and do it in a thoroughly through the formal procedures set in place. The agenda setting of PREFORMA encompasses a range of issues, some technical, others legal in character – and such has proved to be the case, a combination of those aspects.

5.7.2 Requirements: Deliverable D4.3 "Functions of the Open Source Portal"

The objective of the deliverable D4.3 "*Functions of the Open Source Portal*" is to report the functions of the Open Source Portal (of the PREFORMA project website), and the requirements for the associated open source project websites. Specifically, this deliverable sets out the direction for how the work of the suppliers in Work Package 6, and it should be conducted.

The PREFORMA website contains a dedicated section (portal) which provides references to each open source project that is maintained on individual development platforms. The portal includes references to source code, build environment, executables, test files and other information related to each open source project. Each open source project focuses on one type of file format and all developments of software are available on an open development platform.

The deliverable explains further that each open source project should utilise established work practices for community based open source projects. This includes iterative development with frequent releases of source code, executables and all associated development assets necessary for using different versions (development versions, stable versions, deployed versions (made available for use) developed by each project. Stable versions (provided on a monthly basis) have been exposed to a certain level of Quality Assurance (QA) in the development process.

In D4.3 a distinction is also made between stable versions and Long Term Support versions (LTS)[129], which are exposed to additional quality assurance (QA) and aim to be supported during a longer time window by the open source project.

All releases of software (whether in development version, stable version or deployed version) should be kept available at the open source project website during and after the project, so that the open source communities that have been built around each tool can contribute.

Further the D4.3 outlines documentation and coding practices, how development platform and tools are supposed to be set up, and further recommended practices regarding source code, build environment and executables. The development and provision of test files are discussed

---

[129] https://en.wikipedia.org/wiki/Long-term_support

as well. These guidelines are all directed at the suppliers. Instructions with regard to the Open Source Portal are then addressed, built on the premise that the portal should contain links to resources provided by each open source project.

PREFORMA is a project that funds the development of open source software reference implementations in the areas of text, image and AV. An open source approach is meant to be implemented in each stage of the development work process. This is evidenced by the fact that references to open source practices are also made in the Framework Agreement, ITT, and the Challenge Brief and also in the deliverable D2.1*"Overall Roadmap"*.

5.7.3   The Compliance Issues of WP6

Issues given "checked" status:

- PREFORMA requires that a supplier provides the complete source code (i.e. a single zip-file containing all necessary files) under two specific licenses on the open source portal. There must be one zip-file containing all of the source code necessary for each deployment platform. The current configuration of the Open Source Portal proves that this requirement is under control, given the status "checked".

- PREFORMA requires that the supplier provides frequent, and open releases (monthly) which have been exposed to a certain level of QA. The reports by the suppliers could have provided even more information about their procedures regarding capabilities for software releases, but the information do merit that this requirement should have "checked" status.

- PREFORMA requires that an executable shall be provided for each platform. Based on observation from the Open Source Portal, this requirement is "checked".

- PREFORMA requires that the supplier provides all code under open source licenses, and that all code can be distributed and re-distributed by any individual. "Checked" due to the introduction provided at the Open Source Portal.

- PREFORMA requires that the software can be used via standard web browsers. Here, web versions are under development by the suppliers (MediaArea has released its first version; while EasyInnova and vera PDF mention plans to release their web versions in the Final Reports. "Checked".

- PREFORMA requires that a supplier provides an up-to-date roadmap for the different versions of the software, targeted at external contributors, on the development platforms. An overview of the GitHub platforms of each supplier indicates that this requirement is understood and "checked".

Issues to be communicated:

- PREFORMA requires that the "MPLv2 or later" version and the "GPLv3 or later" version of the software that are developed and distributed shall always be identical. This is an issue which needs further discussion and deliberation, since formulations in the deliverable D4.3 seem to deviate from the Framework Agreement, ITT and Challenge

Brief which all regulate that all software is released under conditions set by both licenses (releases are dual-licensed).[130]

- Finally, issue number eight regards provision of detailed documentation of interpretation of the technical specification of each open file format. PREFORMA requires that the supplier provides documentation on precisely how technical specifications of file formats are being interpreted and implemented in software. The D4.3 states:

  "Such details is critical in feedback for organisations maintaining technical specifications in file formats and (…) the provision of precise interpretations of different parts of a technical specification has been interpreted (…) will constitute a very valuable resource for the open source and standards communities".

  This particular aspect is important for the PREFORMA consortium to follow up on.

---

[130] https://en.wikipedia.org/wiki/Multi-licensing

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

6 CONCLUSIONS

This chapter presents six main conclusions based on the analysis undertaken and documented in previous chapters. The chapter ends with a short "charter for success".

6.1 OVERALL GOAL ATTAINMENT

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

First, the prototyping activity is related to the software releases that took place during the 1st part of the Prototyping phase. In the reporting structure used in this 1st part, a plan for releases was expressed in terms of a) frequent releases (if possible monthly), b) intermediate releases (planned for July and October 2015 respectively).

The projects have evolved successfully from the Design phase, which was characterized by the formulation of the functional and technical requirements, into a phase where releases are being made regularly in conjunction with the prototype development. Based on observations made from updates at the Open Source Portal of the project, it can be quite easily deduced that the goal of frequent releases by the Suppliers is met.

At this particular stage, releases are managed through scheduling, planning, and end-users (for the moment in practice mostly represented by members of the PREFORMA Consortium) are invited to make usability tests for themselves. The self-reporting made by the Suppliers provides insights into how build and release capabilities have been set up in order to accommodate to the PREFORMA release requirements.

In addition to this, since the open source projects are out "in the open" on various homepages on the Internet, there are plenty of opportunities for the PREFORMA Consortium to follow the on-going activities closely. A key here is to follow developments not just on GitHub, but on the home pages to which the open source projects are connected to, and which display in practice the continuous integration and continuous deployment methods used by the suppliers.

A very important underlying guideline in the provision of software prototype releases is that PREFORMA requirements **should** be fulfilled. Still, an overview of the open source portal, where each open source project is described, demonstrates that there is awareness among suppliers about the PREFORMA requirements in areas such as the provision of source code, executables and build environment.

Second, the suppliers were required to provide explanations and documentation how the developed software can be effectively used in archiving scenarios at memory institutions (regardless of their size and the file type they make us of).

One of the 'unique selling points' of the PREFORMA project is that validation and policy checking functionalities are supposed to enhance long-term digital preservation. Therefore, the request is not merely for documentation (instructions, guides, issue trackers, release notes and so on), but also for explanations on how the software will be interoperable with legacy systems (via APIs). This request for explanations how the software can be used in archiving scenarios is still valid and important, not the least to create stakeholder engagement.

Third, the suppliers were required to provide demonstrations of the results. During this phase, suppliers have not carried out demonstrations in any pronounced way although numerous presentations of the projects have been made. A particular activity in this regard is the PREFORMA Open Source Workshop in April 2016 in Stockholm, where a presentation and demo of each of the three conformance checkers is scheduled.

Our conclusion is that goal attainment has been reached in the 1st part of the Prototyping phase: suppliers have delivered prototypes, provided information on their releases, how software testing is carried out, dissemination and community building efforts, the open source approach in use, and their standardisation efforts.

6.2   TRIPLE HELIX COORDINATION

The PREFORMA-PCP expresses an ambition to facilitate triple-helix collaboration between the project partners from academia and memory institutions, and industrial suppliers of solutions. This is essential, since the area of ensuring correct implementation and use of long-term sustainable file formats requires research on which formats to use, and how to ensure that software for using them will be available for the future. At the same time, only the suppliers have detailed knowledge concerning resources needed for the development of effective solutions that satisfy requirements for long-term archiving of digital assets as requested by PREFORMA.

Through the selection of the three suppliers in April 2015, we have made observations of genuine triple helix cooperation in PREFORMA.

We see civil servants taking part in research, cooperating with suppliers to solve technical issues; businesses commenting on requirements and other legal issues; and researchers being very much involved in legal interpretation of technical issues (an area often perceived to be reserved to civil servants employed at legal departments). Hence, the project is a unique area of cooperation.

This cross-fertilisation and crossing of sectional boundaries within the framework of the project shows that triple helix type of interactions are continuous, and perceived as normal to PREFORMA. The expected outcome of this is the creation of an environment in which the participating parties are involved in a learning process that enhances innovation. This takes place within the venues for cooperation: supplier meetings, follow-up activities including usability testing and feedback, informal discussions, by participating in the reporting structure, or in workshops and so on.

A major aspect of the Triple Helix Cooperation is technology- or knowledge transfers. This means that those parties who participate have the potential to gain up to date knowledge in a number of areas related to current software development as research and development occurs to enhance long-term preservation. Participating parties will not just gain from the software products – but from the process itself.

6.3   A LEARNING CURVE THROUGH PROTOTYPING

Software releases and usability testing have characterized the 1st part of the Prototyping phase. Through these two activities, suppliers and the PREFORMA Consortium have had learning experiences with positive effects on the future software development and on the work to be conducted in the next phase. Although we have not conducted a study to prove this, we can describe this conceptually as a learning curve.

A learning curve shows that learning increases through experience. Through the release of the first prototypes of the validators, issues such as implementation that may seem complicated to some have gained a practical dimension. By testing the products, stakeholders can see for themselves how various specifications have direct relevance to the functionalities of the conformance checkers. This may have a positive effect on their involvement in PREFORMA, which in turn may lead to new tests, new questions, and new information to the suppliers.

Through usability testing, the suppliers have gained insight into how the prototypes function, and how they are being perceived. In fact, this is very much in line with the original intentions of the use cases that were published in the Challenge Brief. It is also a process that may have effects on future design. A focus on usability – or on user centred design – suggests that the needs, wants and limitations of end-users should be given extensive attention at each stage of the development. Some foresight with regard to how the products will be used and be made interoperable with legacy systems should be included in this.

Now, our argument is not that the prototyping activities has provided new knowledge per se; there is plenty of expertise represented within the PREFORMA Consortium and amongst the suppliers, but rather that the real-world experiences of software releases and usability testing has provided a type of learning which could not have been amassed in any other way, and which is likely to continue throughout the duration of the project. A practical dimension has been added to the project. In fact most implementation issues in PREFORMA are totally practical in nature.

This learning curve may be summarised as follows:

1. The step from design to prototyping set in motion a series of events, most notably releases, testing and other issues included in the feedback process. In this sense PREFORMA entered familiar terrain to most software development projects. The suppliers got feedback from the users early in the project. The PREFORMA Consortium was made aware of whether the software matches the functional and technical specifications. Further, it provided a real-world test into the feasibility of having frequent, monthly releases. These experiences have been brought back to the Consortium as a foundation for the coming re-design phase.

2. Suppliers have been provided with valuable insights. One such insight regards how they are supposed to implement open source best practices. The deliverable 4.3 "Functions of the Open Source Portal" lays emphasis on two particular platforms where these practices can be followed: the open development platforms, and the PREFORMA open source portal. All three suppliers have been involved in the issues which the deliverable brings up, such as provision of source code, executables and build environment, documentation and so on. Another input has been given to the suppliers by those end-users that have tested the products. Here, issues such as installation, usability and functionalities have been raised. Third, the suppliers have been provided with input through external contributors that have begun testing the validators themselves.

3. The PREFORMA Consortium has had valuable experiences as well. These can be summarized by the introduction of two new windows: a) the launching of prototypes has opened an experiential window whereby the members of the Consortium can see for themselves how the technical and functional requirements are being implemented. This will probably make discussions more practical and relatable to the functionalities of the conformance checkers then before. b) The open development platforms and their

interoperability with the continuous integration tools which the suppliers use, make it possible for the PREFORMA Consortium to follow the open source projects in real-time. Through participation in testing and follow-up, the PREFORMA Consortium now has new perspectives and insights into the original intentions of the projects. These have been communicated in easy-to-understand language through the project website, and deliberated and developed further in a number of previously published deliverables (2.1, 2.2, 3.1, 4.3, and so on).

6.4   EXTERNAL PARTICIPATION CURVE: STANDSTILL OR INCREASE?

In this deliverable, the dissemination and community building efforts of the three suppliers are addressed in some detail. The background is that the PREFORMA suppliers should have a clear plan for outreach, and seek to partake in different dissemination activities, and seek to build open source communities around the respective tools. (However, dissemination and networking activities are not just supplier related, but also belong to the responsibilities of the members of the PREFORMA Consortium.[131])

Our observation is that the suppliers have done well disseminating the open source projects. Numerous means of outreach and participation have been set up by each of the suppliers, and there are reasons to believe that these will take off simultaneously as the tools reach further stages of development and take-up.

We are not as certain when it comes to the community building efforts, however. It is noticeable that capabilities have been set up thoroughly as have strategies (EasyInnova makes mention of one such strategy) and procedures (veraPDF reports of how it aims to proceed if and when pull requests reach developers.) We have become aware of methods whereby direct external involvement in the development can be measured (number of external pull requests, accepted commits, comments and questions etc.) but we have not undertaken any such measuring exercise in this deliverable. Research and a priori notions indicate that successful open source projects are highly collaborative, and a development in this particular area is anticipated.

The importance of external participation is further deliberated in deliverable D8.8 "*Monitoring of the open source project implementation*".

6.5   CONTINUOUS DELIVERY AND QUALITY ASSURANCE (QA)

A noticeable feature of the 1st part of the Prototyping phase is the sheer magnitude of software releases. We have also gained insights into the continuous integration and delivery practices used by the suppliers.

The on-going development in the three open source projects can be monitored "in the open" through the Internet platforms. The GitHub platform is directly linked to and operative with the Travis-CI and Jenkins platforms. This means that access to the most recent test reports are just a few mouse clicks away. The GitHub-Jenkins interoperability is especially helpful since it

---

131   http://www.digitalmeetsculture.net/wp-content/uploads/2015/04/PREFORMA_D3-1_Terms-of-Reference-for-the-Network-of-Common-Interest_v1.3.pdf

provides information not just about tests, but also informs of recent changes and the last successful deployed artefacts.

In addition to this, the suppliers were requested to report on their releases in intermediate and final reports. Issues such as software testing, fulfilment of requirements and relationship to the functional and technical specifications, as well as next steps to take are also included in the reporting structure.

A question for further discussion is whether the information that the suppliers provide is adapted adequately to satisfy stakeholder claims. Does the reporting structure need changes? Is it enough to use Internet links whereby the development can be followed more or less in real-time? The questions are not just related to quality assurance, but to interaction between all involved parties.

6.6 PCP WORKING CONDITIONS

It is important that PREFORMA continues to develop under PCP conditions. Pre-Commercial Procurement (PCP) is a competition-like procurement method, which enables public sector bodies to engage with innovative businesses and other interested parties in development projects to arrive at innovative solutions that address specific public sector challenges and needs. The innovative solutions are created through a phased procurement of development contracts to reduce risk. Why is this important?

First of all, the PCP framework continues to be important because of its incentive structure. A project which is funded by the European Commission, coordinated by a national government agency, and which has brought together both public sector organisations, private organisations, and universities from several countries in Europe to address long-term preservation through the development of software implementations of open file formats, is a major opportunity for the suppliers. Given successful outcomes, this may have an impact on how digital preservation is administered by memory institutions.

Second, the PCP framework continues to be important because of its emphasis on procuring research and development as a means to achieve public sector innovation. Some of the outer concerns and questions are these: Do government agencies stand to gain something from the employment of an open source approach to its IT-projects? What are the dynamics behind successful open source projects? Can these dynamics be utilised by the public sector as a means of counteracting vendor lock-in? In addition to this, there are some "inner" or topical concerns to the research related to the establishment of reference implementations.

Third, the PCP framework emphasizes competitiveness within a framework of values such as openness, equal treatment and fairness. This has several implications. Yes, the suppliers are competing, but may cooperate within a specific assignment if done in a transparent manner. Yes, the Suppliers are contracted, but are valued as collaborators able to provide innovative solutions supposed to reach beyond the duration of the Framework Agreement.

6.7 CHARTER FOR SUCCESS

Success point 1: Involvement

- The involved parties must realize that there is much to gain not just from the validator tools that make up the end-products of the project – but from the process itself. Through the means of triple helix coordination, participants can follow closely how frontline

software development issues are being raised, discussed and implemented in innovative tools.

- PREFORMA aspires to create open source software reference implementations in vital open file format areas, and all the while many implementation issues are characterized by either technical or judicial expertise, most of them are totally practical and (should) relate to end user contexts.

Success point 2: Communication

- Easier access to open development platforms through means such as wikis, blogs, forums, README-files, or by provision of recent links to platforms is essential to create further participation and interest.

- Given the importance of stakeholder interaction through on-line communication, it is essential that suppliers go the extra mile to secure that they are understood and that there is a "sell component" to what is being published.

Success point 3: Projects and Processes

- Finding out who is working in the open source projects, and what roles/functions they fulfil, takes some digging on the respective websites. Clarity can be important in this regard. Are these persons available to reply to comments or questions?

- From the reporting structure it is evident that the suppliers are involved in a number of processes and their related activities. These are described in writing but a preparedness to report on them with respect to their beginning, progress and anticipated result would be beneficial.

Success point 4: Seeking to Realize the PREFORMA Challenge

- Focus need to be on the following questions:

  o What are the main challenges to achieve long-term preservation of files within the open file formats in focus?

  o Are there major differing interpretations present in software implementations, and how does this eventual situation affect working conditions?

  o What measures are taken to achieve the objective frame of reference needed to interpret and implement the standard specifications?

  o What is the current relationship between the status of development and the OAIS[132] framework, and its goal for memory institutions to obtain sufficient control to the level needed to ensure long-term preservation?

---

[132] Concerning OAIS, see for instance PREFORMA Challenge Brief, p. 9.

ANNEX 1: TEMPLATE FOR THE INTERMEDIATE REPORTS

PROTOTYPING PHASE 1 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.

In Step 1 of the Prototyping phase, the plan for releases is as follows:

- Frequent releases: monthly;

- Intermediate releases: end of July 2015 and end of October 2015.

The intermediate release shall contain two parts:

- A more organised release compared with the respective predecessor version

- A report which

  o 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 but planned to be done toward the end of Step 1 in the prototyping phase.

  o 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 1 – 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 – Intermediate Report
Total Contract Price [euro]:
Start Date:
End Date:
Sub-contractors:

1. Description of the release

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

Please provide the PREFORMA consortium with a description of what it is still missing in the current release and which are your plans until the end of Prototyping Phase 1 (end of October 2015) on how 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.

ANNEX 2: TEMPLATE FOR THE FINAL REPORTS

PROTOTYPING PHASE 1 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.

In Step 1 of the Prototyping phase, the plan for releases is as follows:

- Frequent releases: monthly;
- Intermediate releases: end of July 2015 and end of October 2015.

The intermediate release shall contain two parts:

- A more organised release compared with the respective predecessor version
- A report which
    - o 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.
    - o 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 1 – 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 release 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 last intermediate release (July 2015) and how are you addressing the comments received from the PREFORMA consortium.

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. 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 design phase 1.

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.

ANNEX 3: FOLLOWING OPEN SOURCE PROJECTS

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[133]). Another aspect of open source is specific open source licenses which "allow software to be freely used, modified, and shared".[134]

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.[135] An alternative is to use the links to each of the three projects that the suppliers have provided in the intermediate and final reports.[136] 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.[137] 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[138], you are provided with a pathway to the current build[139] 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

---

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

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

[135] https://github.com/

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

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

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

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

consortium).[140] At the moment of writing this text, the consortium had set up branches like "default or master branch", "active branches", and "inactive branches".[141] If you click on build history, a list of these is produced.[142]

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.[143] At the Jenkins platform you are provided with full information however.[144] 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".[145] 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". [146]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.

---

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

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

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

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

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

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

[146] 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.[147] 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.[148] 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.[149]

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[150], while the TIFF library provides access to the Maven repository of the project[151], as well as to its particular Travis-CI activities.[152]

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.[153] 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[154], or to the TIFF Library project.[155]

---

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

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

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

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

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

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

[153] 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.

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

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

Lastly, the information provided at the Maven[156] repository regards the TIFF library.[157] 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.[158]

---

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

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

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