DELIVERABLE

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

D8.4 Design – Final Report

Revision: v1.0 FINAL

Authors:

Peter Pharow (FRAUNHOFER)
Claudio Prandoni (PROMOTER)
Antonella Fresa (PROMOTER)
Benjamin Yousefi (RA)

Contributors:

Uwe Kühhirt (FRAUNHOFER)
Christian Saul (FRAUNHOFER)
Christian Weigel (FRAUNHOFER)

Reviewers:

Nicola Ferro (UNIPD)

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	Public	X
C	Confidential, only for members of the consortium and the Commission Services	

Revision History

Revision	Date	Author	Organization	Description
0.1	07/03/2016	Claudio Prandoni Peter Pharow Uwe Kühhirt	PROMOTER FRAUNHOFER FRAUNHOFER	First outline
0.2	13/03/2016	Peter Pharow Uwe Kühhirt Claudio Prandoni	FRAUNHOFER FRAUNHOFER PROMOTER	First content ready to be shared with WP8
0.3	15/03/2016	Peter Pharow	FRAUNHOFER	Additions to the first sections and chapters
0.4	16/03/2016	Peter Pharow	FRAUNHOFER	Intro into the chapters without content
0.5	23/03/2016	Peter Pharow	FRAUNHOFER	Additional content to existing chapters
0.6	27/03/2016	Peter Pharow, Benjamin Yousefi	FRAUNHOFER, RA	Additional content to existing chapters, annexes
0.65	28/03/2016	Peter Pharow	FRAUNHOFER	Additional chapter and section headings for content after Stockholm events
0.7	29/03/2016	Claudio Prandoni	PROMOTER	General revision + added Annexes 4 and 5
0.8	01/07/2016	Peter Pharow	FRAUNHOFER	Additional content
0.85	21/07/2016	Peter Pharow	FRAUNHOFER	Summary
1.0	25/07/2016	Claudio Prandoni	PROMOTER	Final check

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This deliverable contains original unpublished work except where clearly indicated otherwise.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.

TABLE OF CONTENTS

EXECUTIVE SUMMARY 5

1 GENERAL APPROACH AND METHODOLOGY 7

1.1 D8.1 Overall evaluation procedure summary 7
1.2 D8.2 Design phase #1 evaluation procedure Lessons 7
1.3 D8.3 Summary of the findings 9
1.4 Informal Evaluation Process in design phase #2 10
1.5 Design phase #2 Informal Procedures 11
1.5.1 Information Procedures 12
1.5.2 Virtual Meetings 12
1.5.3 Informal Review Process 13
1.5.4 Visits to the suppliers' teams premises 13
1.6 Summary 14

2 ADDENDUM TO THE 2015 EVALUATION 15

2.1 evaluation procedure 15
2.2 Decisions 15

3 SUMMARY OF THE NEGOTIATION OF RESULTS 17

3.1 Negotiation Procedure 17
3.1.1 PREFORMA Consortium 17
3.1.2 veraPDF 18
3.1.3 EASY Innova 18
3.1.4 MediaArea 18
3.2 Results after negotiation 19

4 END OF DESIGN PHASE #2 REPORT 20

4.1 Structure of end of design phase #2 report 20
4.2 Suppliers' end of design phase #2 reports 22
4.2.1 PREFORMA Consortium 22
4.2.2 veraPDF 22
4.2.3 EASY Innova 23
4.2.4 MediaArea 24
4.3 Summary 25

5 THE SUPPLIERS' OPEN SOURCE WORKSHOP 26

5.1 Rehearsal and PREforma Pre-workshop 26
5.2 PREforma open source workshop results 26
5.2.1 PREFORMA Expectations and Intentions 26
5.2.2 PREFORMA Consortium 27
5.2.3 veraPDF 27
5.2.4 EASY Innova 28
5.2.5 MediaArea 28
5.3 Summary 29

6 LESSONS LEARNED FROM THE DESIGN PHASE #2 30

6.1 VIRTUAL MEETINGS 30
6.2 VISITS TO THE SUPPLIERS' TEAMS 33
6.2.1 veraPDF 33
6.2.2 EASY Innova 34
6.2.3 MediaArea 35
6.3 INTEROPERABILITY BETWEEN SUPPLIERS' WORK 35
6.4 END OF DESIGN PHASE #2 REPORTS 37
6.5 SUPPLIERS OPEN SOURCE WORKSHOP 37
6.6 SUMMARY 38
7 CONCLUSION 39
ANNEX 1: ASSIGNMENT OF REVIEWERS 40
ANNEX 2: FEEDBACK FORM ON FINAL RELEASE - OCT 2015 41
ANNEX 3: SUPPLIER VISIT AGENDA VERSIONS 43
ANNEX 4: NEGOTIATION REPORTS TO SUPPLIERS 46
ANNEX 5: SUPPLIERS' END OF DESIGN PHASE #2 REPORTS 96

EXECUTIVE SUMMARY

This deliverable is considered to be the final report on the activities related to the preparation and procedure of the design phase #2, internally also called the re-design phase. It is important to notice here already in the beginning of this report that the PREFORMA project, in its Description of Work (DoW) document, had not planned to go for a formal evaluation at the end of the design phase #2, like the one performed at the end of the design phase #1, but for an informal review allowing the project members as well as the suppliers to evaluate and rate the results of their work, their working procedure, the error finding mechanisms, the connection between the phases of design and prototyping, and the progress of the suppliers compared to their own Description of Work (DoW) documents. The final and concluding official and formal review will be performed toward the end of the prototyping phase #2, which eventually means in M36.

Reflected in this document, the second and final design phase of the suppliers' work to create and update functional and technical specification as well as interoperability documents and the related software modules started with the review of the design, the definition, and the specification of the functional and the technical part of their preparatory work according to the call for tender, the submissions (description of work) of the three winning supplier teams and consortia, the review results of the design phase #1, and the work results of the prototyping phase #1 that has followed the design phase #1 completion.

The document will thus include a short summary of the prototyping phase #1 description (originally compiled in D8.3) and the basic statements related to the two phases of WP5 including citations, references and methodologies for:

- PREFORMA evaluation strategy (from D8.1)
- PREFORMA lessons learned from the design phase #1 (from D8.2)
- Short summary of the findings of the prototyping phase #1 (from D8.3)
- Short summary of the informal review procedure
- Procedure of the evaluation of the suppliers' documentation
- Results of the PREFORMA consortium visits to the suppliers
- Individual meetings with the suppliers on results of evaluation
- Informal decision making process
- Statements of the end of design phase #2 report
- Open source workshop and the suppliers' performance
- Final decisions made by PREFORMA consortium

The previously completed tasks in WP8 laid the foundation for informal evaluation strategy for comparing the results of the suppliers, ranging from the end of the design phase #1 to the prototyping phase #1 to the beginning of the design phase #2. The formal evaluation framework, also applied to this informal evaluation for the design phase #2, has been defined in D8.1 and has been successfully applied in D8.2, based on contributions of the technical partners as well as of the memory institutions, either being partners in PREFORMA, or being invited as external experts. The strategy negotiated and established in T8.2, and consequently described in D8.2, too, was used as an input for evaluating the suppliers' results informally at the beginning of the

design phase #2 to value whether or not the suppliers have fulfilled the tasks they were expected to do.

The document D8.4 is thus intended to include all useful information for the internal and external work process for the design phase #2 as well as to give an idea on how a PCP project does follow and rate the progress of suppliers during the development process, and finally how the informal evaluation in PREFORMA has been performed.

After a brief introduction to the general approach and methodology (chapter 1), the first part (chapters 2, 3) summarizes informal evaluation criteria and procedures without detailing it too much. The main part of the evaluation of design phase #2 bases on design phase #1 and prototyping phase #1 evaluation.

The second part (chapters 4, 5) reflects and explains the results of the informal evaluation without giving a ranking, as this time it was not needed to rank but to compare the success of the suppliers with the plans of their respective work.

The last part (chapters 6, 7) is dedicated to the findings of the informal evaluation of the work of the suppliers, and to a summary comparing the design phase #1 lessons learned with the evaluation results of design phase #2, in order to draw a few conclusions on what to improve during the prototyping phase #2.

1 GENERAL APPROACH AND METHODOLOGY

The document presented here shortly repeats the description of general norms and methodologies for the informal evaluation (see explanation about formal and informal evaluation in the chapter before) and the related review process of the design phase #2 lasting from M23 (November 2015) to M26 (February 2016). Due to having planned and scheduled the open source workshop to be held in Stockholm in the beginning of M28, the delivery of D8.4 was postponed to the end of M28, in order to capture the results of the suppliers' presentations during the Stockholm week. The results of the former WP8 tasks, laid down in D8.1, D8.2, and D8.3, provided the design phase #2 team and thus also this document D8.4 with methods, methodologies, and procedure suggestions for the informal evaluation of the work performed by the three supplier teams. The following sections are to be seen as informal citations from D8.1 and D8.2 as well as D8.3 in order to provide the readers of D8.4 with the basis for understanding the chosen procedures, the underlying processes, the applied forms, the templates used to ask the suppliers for presenting their results, and the set of evaluation strategies as well as the decision making process at the end of design phase #2.

1.1 D8.1 OVERALL EVALUATION PROCEDURE SUMMARY

The aim of document D8.1 [D8.1] has been to develop a painstaking method to evaluate and compare different suppliers. In particular, the developed method shall be effectively employed in the evaluation and comparison of PREFORMA suppliers. The proposed competitive evaluation strategy will be used for the assessment of the suppliers at the end of the design phase 1 in order to choose those who will continue with the prototyping and testing phases.

The proposed method is a general evaluation framework, which is applied to the evaluation and comparison of suppliers as a relevant use case. This means that the proposed method can be successfully employed in other projects, in addition to the PREFORMA project. PREFORMA will provide a report explaining the respective lessons learned.

Evaluating and comparing suppliers requires the PREFORMA members as well as the external reviewers to identify two distinct processes: evaluation process: during this process each supplier is individually examined and it is scored according to its characteristics. The evaluation process is formalized through the evaluation matrix, as described in D8.1 section 2.1.

The outcome of the evaluation process is the supplier score that is a number representing the scoring achieved by the supplier; comparison process: once the suppliers have been scored, they are compared with each other on the basis of their supplier's score. The comparison process is formalized through the comparison matrix, as described in D8.1 section 2.2. The outcome of the comparison process is a ranking of the suppliers, based on their scorings.

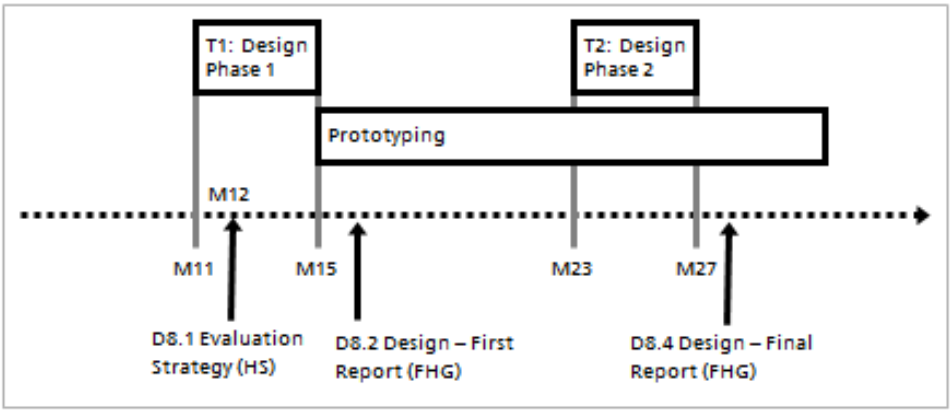
1.2 D8.2 DESIGN PHASE #1 EVALUATION PROCEDURE LESSONS

As one of the PCP projects currently running on EU level, the PREFORMA consortium needed to plan, prepare, and undergo the entire call for tender procedure. This meant learning a lot from the preparation of the process and from all the procedures related to the aspect of inviting suppliers to submit their description of work as part of their bid, from negotiating the bids along

the line of the PREFORMA requirements, from guiding the suppliers along this line, and fro evaluating the submitted specifications and administrative documents in order to be able to go for a fair decision making process.

The first lessons learned – of course there will be much more lessons that PREFORMA will learn from the prototype phase as well as the testing phase later on – contain the time from the end of the negotiation phase until the end of the decision making process for the supplier teams invited for entering the prototype phase. These lessons will shortly be summarized below. Of course there are much more lessons that the PREFORMA consortium did learn along the design phase #1 progress but they more or less address aspects beyond the evaluation and decision making process so they shall be reflected in a different place.

The design phase procedures in general were organized as shown in the figure below. The phase #1 ended with the submission of the suppliers' reports (end of design phase #1 report, functional specification, and technical specification) followed by the review and evaluation process. The results had been described, analyzed, and summarized in the PREFORMA report D8.2 Design – First Report [D8.2].



Before PREFORMA could have started analyzing the lessons learned from the evaluation process for the design phase #1 in particular, two generic lessons learned were found and taken into account already in the beginning. First of all, PREFORMA did find an appropriate and well-acknowledged way to intensively communicate with the suppliers. This mechanism has been applied during the course of the entire communication with the suppliers. Of course, things worked even better during the prototype phase #1. But both the suppliers and the PREFORMA consortium had to learn how to communicate with each under always taking the limited time and financial resources into account.

Secondly, it turned out to be a very good and wise decision to invite the suppliers for the workshop to Brussels. The aim of the workshop back in M15 (March 2015) was just to give all the suppliers a podium for not only sending their respective documents but also to illustrate the findings and definitions, to present themselves as a consortium rather than independent suppliers just forming a team, and to get in touch with them especially for the memory institutions that may have (had) several questions in terms of the implementation of the proposed modules and components.

Summarizing what has been found out during the design phase #1 evaluation procedures, at least five lessons learned could be identified. They will for sure be complemented by a series of other lessons learned that are of a more generic character, and will be presented and explained on a different occasion.

1. The definition of applicable evaluation criteria was found out to be a very critical part of the project because the evaluation process requires that – eventually – these criteria generate quantitative scores to be used to ranking.
2. The entire set of documents provided by the six suppliers has adequately been focused on the call for tender and the challenge brief topics so that it was comparably easy to evaluate the documents according to the quantitative evaluation criteria defined.
3. The assessment of additional qualitative criteria like, e.g., the “elegance” of the architectural approach could only be appropriately reflected for the evaluation scheme and the decision making process in an appropriate combination of quantitative and qualitative evaluation criteria.
4. The PREFORMA consortium will, with the beginning of the prototype phase 1, put very much effort and attention to the fact that all three selected supplier must focus their development processes on innovative R&D work.
5. All these preliminary lessons learned, and certainly several more, will seriously be incorporated into the planning and preparation of the re-design phase to start in November 2015 (M23 – M26).

First of all, in the summary chapter of this entire document, PREFORMA will compare the lessons found and learned after the design phase #1 with the findings of design phase #2, and draw conclusions to what extend the re-design of the specifications has led to appropriate answers to the lessons listed above.

Secondly, and as it had turned out to be very successful to get all suppliers together and to present results not only to the PREFORMA consortium but to the Open Source community, PREFORMA had planned to set up an Open Source Workshop in Stockholm at the beginning of M28, and the results will be incorporated in this document further down.

1.3 D8.3 SUMMARY OF THE FINDINGS

This chapter [mainly taken from D8.3 chapter 6] presents the main conclusions based on the analysis undertaken and documented in previous chapters.

During the prototyping phase #1, the three selected suppliers were expected to provide software prototypes that fulfil the requirements of the PREFORMA project as well as their own requirements derived from their own DoW, to demonstrate the results achieved so far, and to provide explanations and documentation how the developed software can effectively be used in archiving scenarios at different types of 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, and (b) the intermediate releases (planned for July and October 2015 respectively).

The projects have evolved successfully from the design phase #1, 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 prototyping phase #1, 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.

The PREFORMA core members’ conclusion is that goal attainment has been reached in the prototyping phase #1: 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 standardization efforts.

1.4 INFORMAL EVALUATION PROCESS IN DESIGN PHASE #2

Similar to the evaluation of the design phase #1 documentations, three reviewer types were considered which correspond to the three main stakeholders involved in the PREFORMA project. They were (a) the technical experts who analyzed the proposed solution by evaluating the solution from the technical point-of-view including the open source software code and its implementation, (b) the domain experts who analyzed the solution by verifying the solution regarding

whether it fits the requirement of the domain where it shall / will be used, and by implementing it in their own environments, and (c) the external experts who finally analyzed the proposed solution, external to the members of the PREFORMA consortium, to compensate for any possible biases in the respective domain especially also form the viewpoint of developments that may have happened outside PREFORMA.

In particular, each supplier solution has been reviewed by reviewers for each reviewer type but all with regard to the media file type of the respective solution. From the technical viewpoint, PREFOMA had allocated three reviewers for each proposal. From the domain view, three other reviewers were assigned to each proposal, and there was at least one external reviewer for each proposal. In total this means that each proposal had up to seven review matrix results at the end of the review period so that a comprehensive and legally valid / approved analysis and decision making process could be guaranteed. Moreover, the HS open source experts reviewed the achievements from that particular viewpoint, too.

In terms of details assignment of suppliers and reviewers (for a structured view see annex 1), PREFORMA almost kept the assignments that were used for the design phase #1, just adding the interoperability review expert to the list of the technical reviewers. In summary, the following appointments were made:

- RA, EVKM, and LGMA have evaluated VeraPDF project as domain experts.
- SPK, KB, and KIK-IRPA have evaluated Easy Innova's DPF as domain experts.
- S&V, AJGI, and GFC have evaluated MediaArea's MediaConch as domain experts.
- FRAUNHOFER and PACKED will have a look at all the three reports as the PREFORMA project's technical experts.
- RA (Benjamin Yousefi) reviewed all three sets of documentation mainly from an interoperability viewpoint.
- HS had a look at all documents from an open source viewpoint.
- Jozo Ivanović have evaluated the veraPDF project as external expert.
- Jan Dalsten Sørensen have evaluated the Easy Innova project as external expert.
- Peter Bubestinger and Hermann Lewetz have evaluated MediaArea as external expert.

As due to expertise and time constraints not all reviewers completely filled the review forms, it became necessary to on the one hand summarize the results as soon as possible in order to allow the suppliers to proceed – contrary to the design phase #1 review, the informal review in design phase #2 did not end up in a rejection of one or more teams but in an in-depth analysis of the achievements so far, and in a comparison with what the suppliers had planned to achieve at the end of the prototyping phase #1.

1.5 DESIGN PHASE #2 INFORMAL PROCEDURES

In order to prepare both the suppliers and the PREFORMA consortium members (technical partners and memory institutions) for the informal review process, various measures have been installed (1) to get all relevant information forwarded to the suppliers and the consortium members alike, (2) to organize virtual meetings between the consortium and the suppliers, (3) to or-

ganize the informal review process according to the delivery dates and the required documentation, and (4) provide the suppliers a basis for presenting their results to the consortium.

In the following, short description will be provided on the procedures for the four aspects along with a few results achieved.

1.5.1 Information Procedures

PREFORMA made a very good and promising experience in having installed virtual meetings with the suppliers throughout the entire phase of collaborating with the suppliers. As this procedural step started in the design phase #1, PREFORMA proceeded with it in the prototyping phase ‘1 as well as in the design phase #2, and it is planned to continue also for the prototyping phase #2. For keeping the consortium updated, the mailing list had been implemented right at the beginning of the PREFORMA project. This mailing list allowed all partners to send all relevant information to all partners at the same time. As PREFORMA figured out that this was a well-acknowledged way of spreading information, the project management implemented a mailing for all suppliers at once and a mailing list for all evaluators. Via these measures, the PREFORMA management has been able to distribute all necessary information timely and without preferring one or the other supplier. The PREFORMA management decided to provide relevant information / answers on suppliers’ questions not only to the supplier who asked a particular question but to provide all suppliers and all PREFORMA consortium members with the answers.

1.5.2 Virtual Meetings

As several of the suppliers had various questions also during the design phase 2, the PREFORMA consortium, as a first important focus, continued running the procedure of a virtual meeting with all suppliers and all PREFORMA consortium members invited to actively participate. The meetings for the design phase #2 started in early November, immediately after the virtual kick-off for the design phase #2, and ended right before the suppliers’ Open Source Workshop in early April. All suppliers were encouraged to raise questions before the meeting so that the consortium members were able to prepare for a comprehensive answer. A second important aspect for the virtual meetings – suppliers had the chance to exchange information among each other, so for the common work on interoperability aspects.

The procedure of establishing and continuing such virtual meetings allows the consortium, in addition, to answer all incoming questions to all suppliers so that all received the same message at the same time. As there was no competition but collaboration of all three supplier teams after the beginning of the prototyping phase #1, the suppliers mainly used the virtual meetings to get in touch with the memory institutions being consortium members. Question had relations to, e.g., training and test data, open source licenses, submission deadlines, the project interoperability and API development work, the role of suppliers and PREFORMA partners in the interoperability context, requirements documents and response during development phases and cycles, and the opportunity to re-submit the technical and functional specification during the design phase #2. The minutes provided after each meeting allowed those partners and suppliers who did not participate with a comprehensive overview of the discussion, and more importantly on the decisions taken.

1.5.3 Informal Review Process

As the design phase #2, contrary to design phase #1, did not interrupt the work of the suppliers, the PREFORMA consortium at a very early stage decided to ask the suppliers for possible updates in the technical and functional specification documents. All provided documentation including the intermediate and final report for the prototyping phase #1 were taken into account for the review. The appointed evaluators and reviewers had thus been able to review all the documents at once, and could provide information on how to improve the specifications, and what to additionally incorporate into the re-design procedure. The reviewers did not want to significantly influence the work process of the suppliers so they only informed about obvious gaps and misunderstandings in line with the DoW documents of the suppliers.

There were two main focus elements for the informal review process. First of all, as results of the prototyping process#1 were available, all reviewers were asked to install the respective software, either as a developer or as a user, and provide feedback on technical or content gaps and errors in a way that the suppliers could immediately deal with the findings. The GitHub platforms of all three suppliers offered experienced reviewers the opportunity to immediately provide the comments to the source code and the related procedures. For the memory institutions, it was mainly the implementation as a domain expert, so they were asked to provide their comments into a template documents that by the way was used by the technical experts, too.

The second element resulted in an analysis of the current status of work compared with the DoW documentation the suppliers had provided about one year ago, at the end of the negotiation phase. The reviewers were asked to compare the intermediate and final report to the prototyping phase #1 with the specifications and other reports provided by the suppliers. The analysis results were offered to the suppliers, and PREFORMA asked them to respond during the course of the virtual meetings but also, and mainly, during the PREFORMA delegation visits at the premises of the suppliers' teams (see following section).

1.5.4 Visits to the suppliers' teams premises

PREFORMA came up with the idea and plans of visiting the suppliers in their own premises right before the kick-off of the design phase #2. The idea to plan such visits had various reasons. Meeting the suppliers in their own environment (even if it is partly a virtual one) was very important for PREFORMA; and so was the chance to see not only the representatives of the suppliers' teams that appear in the virtual and real meetings but also other people in the back offices. Secondly, PREFORMA got more information about the internal procedures of work including error detection and handling. Third of all, as the character of the three suppliers' teams seemed to be different (which could even be seen as a positive effect), PREFORMA could, by means of intensively talking to them, get a respective impression of whether or not these expectations and impressions where right – an important factor for all future communication.

All in all it turned out that – more or less – only EASY INNOVA had a real home base as expected by the PREFORMA team members whereas the other two supplier teams, veraPDFa and MediaArea, had either at least a base to meet, or had decided to completely work virtually. So the PREFORMA representatives had to take these differences into account when planning and performing the visits to the suppliers.

January 23rd saw PREFORMA core team members showing up in Girona, Catalonia, Spain, for the visit to EASY INNOVA, the supplier responsible for the image media file type. In Girona, not

only the PREFORMA-related aspects became part of the agenda but also the interest of the Europeana Space / Technical Space team (in particular NTUA, Athens, Greece) for including and implementing the image conformance checker into their platform. So the agenda contained Iso possible collaboration aspects between PREFORMA and Europeana Space in addition to the dedicated PREFORMA project interests. Apart from the technical and administrative meeting items and topics to be met between PREFORMA and EASY INNOVA, the PREFORMA members got, the following day, the chance to visit the impressive Girona City Archive (one of the nine PREFORMA memory institution partners) to talk about expectations and results, and to learn about the work of the Girona archive staff including those representing the archive in the PREFORMA work.

On January 28th, PREFORMA visited the veraPDFa consortium in their premises in Brussels. Almost all members of the team where present, and PREFORMA had the chance to intensively talk to the veraPDFa members about, e.g., the comments to the veraPDFa work performed and provided so far, about the ongoing standardization process in PDF/A, and about the interoperability work between the suppliers and the status of that work. The aspect of the transparency of code and stakeholder community aspects became part of the agenda, too.

Last but not least, PREFORMA met the MediaArea consortium in a Brussels hotel on January 29th. As the Media Area mostly works virtual, there was no better place to meet. The team members all were prepared for talking to the PREFORMA representatives, and apart from the agenda items shared with the two supplier meetings before, MediaArea was very much interested in talking about technical details, details of the open source code compilation, and other aspects related to addressing the perfect match between the PREFDORMA technical expectations and the MediaArea development work.

The three versions of the proposed and agreed agenda of the respective visit to EASY INNOVA, veraPDF, and MediaArea (listed in the order the different visits took place) can be found in annex 2 to this document.

1.6 SUMMARY

It turned out to be a positive decision to frequently communicate with the suppliers, and to allow the suppliers a frequent communication with all PREFORMA consortium partners, especially the memory institutions. Bringing all three suppliers together independent of the media file type for the purpose of developing interoperable modules and solutions was another wise decision, as it seemed in the beginning that it wasn't that easy to get the suppliers in touch with each other independent of the PREFORAM communication channels. By means of mailing and virtual meetings, the consortium and the suppliers got in touch with each other, and the process of keeping each other informed started right after the formal but virtual kick-off of the design phase #2. And the decision to visit the three suppliers' teams (meaning to bring each supplier's team together in one place) appeared to not only bring information about the organization of work among the respective supplier team but provided a lot of information on their way of working, their internal procedures, and – last but not least – on their behavior as a supplier. More information about the visits and the results can be found in chapter 6 of this document as well all as in annex 3, as far as the agenda of the meetings is concerned.

2 ADDENDUM TO THE 2015 EVALUATION

The design phase #2, as explained before, had to be performed while the suppliers were expected to continue their work on the development of the solutions (conformance checkers, documentation, API, shells, interoperability aspects, standardization, etc.) So the evaluation as such, not only being an informal one compared with the formal one that took place at the end of the design phase #1, looked a bit different from what was planned in the PREFORMA DoW about two years ago. The project team decided to directly communicate to the suppliers, and this turned out to save time.

2.1 EVALUATION PROCEDURE

The procedure to evaluate the three suppliers' work results consisted informally of four different phases that had to interact with each other due of the fact that, as said before, the suppliers were allowed and obliged to continue the work on the prototypes while WP5 was informally evaluating the results achieved so far:

- **Phase 1. Individual evaluations:** Up to seven independent experts per medial file type evaluated separately each package of documents along with the provided versions of the software (a result of the prototyping phase #1) and provided an individual evaluation template based report.
- **Phase 2. Consolidated reports:** As both quality and quantity of the reviews differed, the technical management of PREFORMA undertook the work of summarizing the results of the individual reviews compiling a kind of consolidated report per supplier. The final version of the reports can be found in the PREFORMA repository, accessible from the reserved area of the PREFORMA website.
- **Phase 3. Requests for clarification:** A number of simple questions were asked to the suppliers to clarify missing or unclear information deriving from the analysis of the results (software and documentation) and from the outcome of the consolidated reports. PREFORMA had thus raised a few questions to the suppliers for clarification.
- **Phase 4. Clarifications:** The suppliers were asked to respond to the questions and findings either by means of messages, during the suppliers' virtual meetings, of eventually during the visits of the PREFORMA consortium members.

Finally, the PREFORMA partners waited for the "end of design phase #2" reports (due by mid of March 2016) and for the performance of the three suppliers' teams during the Open Source Workshop in Stockholm (April 2016), for finalizing the review.

2.2 DECISIONS

This section shortly presents and explains the decisions resulting from the assessment of all available information by the PREFORMA consortium. PREFORMA did not expect the suppliers to fail, what matters here is more about the progress of work, the continuation of work, and the interoperability between the three suppliers.

The PREFORMA CSA consortium partners could conclude that the three supplier teams were perfectly continuing to fulfill the needs of the project, in line with their own description of work, and get closer to each other in terms of the interoperability between the shells, the modules, the checkers, and the standardization efforts within and outside PREFORMA.

All suppliers performed very well with respect to the expected work they should individually have done, and in the context of collaborating with each other and with the PREFORMA team members. As PREFORMA partners were experienced in either technical procedures or memory institutions' aspects, the suppliers could benefit from support and clarifications as well as requirements updates in these two fields and even beyond.

In summary, the PREFORMA partners got the impression that the three supplier teams are able and willing to continue on the road to the prototyping phase #2, the final formal evaluation at the end of the third project year, and the transfer of the developed software modules to the testing phase in the last project year.

3 SUMMARY OF THE NEGOTIATION OF RESULTS

The negotiation of the informal evaluation and review results for the design phase #2 mainly took place during the virtual meetings with the suppliers and during the visit of the PREFORMA core team members at the three suppliers' premises during January 2016.

3.1 NEGOTIATION PROCEDURE

The final objective of the negotiation phase in general was allowing both the PREFORMA consortium and the supplier teams to discuss about the current status of their work, the deviations from the agreed road map, and the aspects of making the three different sets of modules interoperable to each other. A second aspect was the planning and preparation for the Open Source Workshop in Stockholm in early M28 (April 2016), and the work of the suppliers to be triggered and improved / enhanced / tailored toward this presentations of the results of not only the suppliers but also PREFORMA as a whole.

For achieving and assuring this, the WP5 and WP8 leaders established a bilateral negotiation process between the consortium and the suppliers, and undertook the attempt to sort out all aspects that remained unclear for a precise and consistent process of re-designing their approach, of continuing the prototyping and development process, and of adequately preparing for the Stockholm workshop as such.

The comments are normally taken into the document more or less in the shape they were sent in by the appointed reviewers, e.g. without filtering and editing, and relate both to specific and more general issues. Sometimes they might reflect different opinions but overall we hope that they provide a useful input for the next period. The negotiation reports contained aspects like the standardization efforts, the gap analysis as well as derived procedures and next steps, and the open source approach as well as community building and information activities.

Additional aspects of the reporting as well as the review of the documentation by the PREFORMA technical and domain experts became part of the agenda versions of the PREFORMA team to the suppliers' offices, and can be found in annex 3 (agenda) and 4 (negotiation reports).

The following sections contain short summaries of the review results and the negotiation protocols as well as of the aspects that PREFORMA considered important to discuss for the progress of the project as such but also for the benefit of the suppliers with regard to the requirements the technical partners as well as the memory institutions came across during the prototyping phase #1, and during various contacts with the stakeholders in the domain.

3.1.1 PREFORMA Consortium

The PREFORMA reviews were informal ones this time. Based on the reviewers' comments provided by filling in the respective form sheets, the PREFORMA core team created three reports for the supplier teams of veraPDFa, EASY INNOVA, and MediaArea. These reports were sent to the suppliers, and PREFORMA asked for their response. This happened by means of the virtual meetings with the suppliers, by having individual chats, and by visiting the suppliers for discussing the status and the progress of their respective work.

3.1.2 veraPDF

Being invited and obliged to submit the “end of prototyping phase #1” report due by the end of October 2015, the veraPDF team provided their documentation on time. The veraPDF consortium has created a platform for the open source project based on the GitHub platform. The different tools the GitHub approach provides for open source projects fills most of the different needs of communication within an open source project. However, the community is not tools and technology but people cooperating to reach common goals. In that respect the activity of the community is very low and it is important that the work to achieve this is increased. One of the main challenges for the project is to have a thriving community so the software will live on when the PREFORMA project ends.

3.1.3 EASY Innova

Being invited and obliged to submit the “end of prototyping phase #1” report due by the end of October 2015, the EASY INNOVA team provided their documentation on time. EasyInnova has created a platform for the open source project based on the GitHub platform. The different tools GitHub provides for open source projects fills most of the different needs of communication within an open source project. However, the community is not tools and technology but people cooperating to reach common goals. In that respect the activity of the community is very low and it is important that the work to achieve this is increased. One of the main challenges for the project is to have a thriving community so that the software will live on when the PREFORMA project ends. The downloaded version (1.2.1) of the software was not working as expected since some basic functions did not work. We therefore decide to check the software repository and build the software on our own. Since this is also part of the requirements we think reporting on this is valuable for both the evaluation as well as the software developers. On Nov. 22nd and updated version (1.2.2) was made available for download. It may have addressed some issues that were raised directly in the repository due to this report.

3.1.4 MediaArea

Being invited and obliged to submit the “end of prototyping phase #1” report due by the end of October 2015, the MediaArea team provided their documentation on time. MediaArea has created a platform for the open source project based on the GitHub platform. The different tools GitHub provides for open source projects fills most of the different needs of communication within an open source project. However, the community is not tools and technology but people cooperating to reach common goals. In that respect the activity of the community is very low and it is important that the work to achieve this is increased. One of the main challenges for the project is to have a thriving community so that the software will live on when the PREFORMA project ends. Throughout the First Design and the First Prototype phase, MediaArea has proven to be a communicative, critical and responsive partner in the project. I have personally gotten to know the team as a driven and dedicated group of people that see in the project the opportunity to contribute to the wider audio-visual preservation community.

The group has shown to be a stimulating and critical conversation partner that aims to improve project outcomes by voicing inefficiencies and uncertainties about common supplier topics, such

as licensing and the shared API. With the Supplier Response to Feedback on the Intermediate release – July 2015, the MediaArea team has responded adequately and in-depth to comments from the PREFORMA evaluation team. Throughout conference presentations to technical experts in the audio-visual archiving community and thanks to open discussions on message boards and mailing lists, the group's networking and standardization efforts are getting increasing attention.

3.2 RESULTS AFTER NEGOTIATION

PREFORMA finished the negotiation discussions and meetings for preparing the following prototyping phase #2. The updates in the reports and the descriptions of work showed the results of the negotiation process, and laid the foundation for completing the re-design, and for continuing the prototyping work on the requested software components.

As described before, the visits of PREFORMA core team members at the premised of three suppliers helped a lot sorting out discussion points, and pushed the process of developing a common agreement among the suppliers for the development of their respective shells and the interoperability of the modules to develop.

All participants agreed to follow the proposed procedures, to attend the monthly virtual meetings, to provide questions prior to these meetings, to follow the deadlines of M26 (February 2016) and M28 (April 2015), and to prepare as best as possible for the April 2016 suppliers' Open Source Workshop to be held in Stockholm. Not only PREFORMA but all attendees expect to learn not only more about the individual solutions but also about the interoperability of the modules to develop, and the continuation of the work process even beyond the end of the prototyping phase #2.

4 END OF DESIGN PHASE #2 REPORT

In addition to the review of the functional specification and the technical specification documents as well as the DoW and the “prototyping phase #1” documentation, provided by the three supplier teams of veraPDF, EASY INNOVA, and MediaArea, another document has to be incorporated into the completion of the review process. The so-called “end of design phase #2” report is foreseen as not only addressing the way the suppliers did make use of the budget. Another important contribution to the informal review provided by the “end of design phase #2” reports, is considered to be the way the suppliers aimed at dealing with the interoperability aspects, the shells, the API, and additional aspects related to open source approaches and the interoperability. In general, the report has to be seen as complementing the statements in the documents provided so far, and aiming at how to proceed in terms of a road map of the suppliers toward the “prototype phase #2” to follow.

The purpose of the “end of design phase #2” report is thus to ensure that contractors have performed the procured R&D services as specified in the framework agreement. The description of work undertaken during the design phase #2 should include what work was completed and why this was important. The suppliers had to complete the template form as fully as possible. The report must be submitted within 14 days of the completion/ termination of the phase. The three supplier teams were advised that satisfactory completion of the “end of phase #2” report formed a part of the contract.

4.1 STRUCTURE OF END OF DESIGN PHASE #2 REPORT

The aim of the administrative and partly also technical report to be provided at the “end of the design phase #2” is to review the results of the three suppliers. In particular, the report contains the following questions to be answered by all suppliers in a way that questions that were purposely left open in functional and technical specification documents should find answers in the “end of design phase #2” reports. Specifications in their respective revised version shall be added to the “end of design phase #2” report as attachments making sure to have all relevant documents for evaluation as one single document with various references from the main part to the annexes and vice versa.

With regard to some of the aforementioned non-technical aspects, the template of the “end of design phase #2” report contained the following questions with the respective explanation what the PREFORMA consortium did want to see in terms of answers.

1. Provision of the administrative Details

Please provide us with all necessary administrative details that PREFORMA needs in order to proceed with the procedures.

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

Please provide the PREFORMA consortium with a concise overview of the main results achieved during the re-design phase, relating it to the original objectives and to requirements defined in the suppliers’ description of work. Please, in terms of a gap analysis, refer to the

functional and technical specification as well as to previously submitted reports without repeating too much here.

3. Describe the innovative aspects of the work, including any new findings or techniques, having a focus on R&D aspects of your work done so far.

Please provide the PREFORMA consortium with a concise overview of the innovation aspects of your work, as expressed summarising the results of the design phase #1 review.

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

If applicable, explain the reasons for deviations or clarifications from what was agreed in the supplier's description of work, or for failing to have achieved critical objectives and the impact on the supplier's project. If applicable, propose corrective actions that will take place in the prototype phase #2 as such.

5. 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.

6. Interoperability between the conformance checkers.

In this sub-section, the supplier is expected to briefly describe the results of the work carried out to set up a framework to ensure interoperability between the conformance checkers developed inside and outside the PREFORMA project and the specifications of a common API.

7. Open Source approach.

In this sub-section, the supplier are asked to describe how they will address the relevant open source topics, the open source licensing, the way to address the open source communities, and the ideas in this respect for the project phases to come.

8. Standardisation efforts.

In this sub-section, the supplier shall, if applicable, describe how the supplier's project aims at contributing to the exploitation of existing standards relevant to the project aims and goals, or how the supplier consortium has thought about contributing to emerging standards. Maybe the supplier can describe how the consortium is going to address future changes on the existing standards taking into account that the near future will bring new archival standards.

9. Provision of data.

In this sub-section, the PREFORMA consortium asks the supplier to provide an overview on how the supplier consortium has been working / will work out the different sets of data needed to develop the respective module. This mainly considers the training data to be used internally but also the test data used by the PREFORMA consortium to test the modules and achievements of each of the suppliers working on the same file type. Eventually, demonstration data is needed to allow companies and organizations outside the PREFORMA consortium to spend their effort on developing their own modules but compare them with the PREFORMA modules by means of using the same correct and corrupt demonstration data sets.

10. Describe the aspects the supplier considers relevant for the upcoming prototyping phase 2 – in order to facilitate assessments of progress into next phase.

In this sub-section, the supplier is expected to shortly introduce the thought and plans of the supplier consortium on how to proceed with the development work in the next phase.

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

In this sub-section, the supplier is expected to add all other relevant information that did not fit the categories listed before.

4.2 SUPPLIERS' END OF DESIGN PHASE #2 REPORTS

The following paragraphs aim at introducing the particular expectations of the PREFORMA consortium as well as the responses and comments of the three suppliers, mainly to the aims and objectives questions. All other answers and comments can be found in the three end-of-design-phase-#2 reports that the suppliers provided the PREFORMA consortium members with, and all did so on time. The full reports can be found in annex 5.

4.2.1 PREFORMA Consortium

The PREFORMA consortium agreed to ask the questions listed in the previous section in order to get more information, for a better comparability, for a better evaluation of the description of work done by the suppliers. Due to the fact that the consortium decided to ask for three different media file types, the solutions need to be assessable in a way that a summary can be made on whether or not – as planned in the beginning of the project – the suppliers are on the right track with regard to PREFORMA, all contractual documents, the specifications, the DoW, and the additional work items like interoperability, community building, sustainability, and more.

As the reports had to be delivered by mid of March, the analysis of the reports and the summary of the responses and perhaps open items are drafts so far, and will more comprehensively be listed in the final version of the document. Same goes with regard to the individual reports to be analyzed below.

What was interesting nonetheless and should thus be mentioned in the beginning: all three supplier teams did not see a need to significantly changing and updating / upgrading their technical and functional specification documents. Of course there was a need to fine-tune, and the interoperability challenges brought up one or the other item of the respective specifications. But in general the suppliers had obviously done a very good job back in the design phase #1, and the evaluators had obviously selected the right teams for developing the checkers for the three media file types document (text), still image, and AV.

4.2.2 veraPDF

Being invited and obliged to submit the "end of design phase #2" report due mid of March 2016, the veraPDF team provided their documentation on time. During the re-design process, the veraPDFa consortium has done various steps of work, listed and explained in the following:

The veryPDF team has subjected their prototyping phase #1 design, implementation, and documentation to review by KEEPS Solutions. They moreover added various fixes and enhancements in line with the KEEPS review, improved the utility of the CLI within Linux pipelines, CLI batch processing and reporting. There was additionally the need to refactor the code base in order to break up the large library project and place incompatible licenses (Apache) into a discrete project. Last but not least, veraPDF created a prototype set of REST interfaces and related browser based GUI.

Even if obviously there was no reason to significantly alter the design in the light of implementation experience and feedback, veraPDFa nonetheless deviated from the planned schedule in a few areas.

The expected PDF/A2 and PDF/A3 corpus files will be released at the end of March (due end of Feb). This is due to extra effort required creating the XMP test corpus in the light of feedback from the archival community.

The release candidates of the PDF/A2 and PDF/A3 Validation Profiles will now be available by mid of May, and support ~80% of PDF/A2 and PDF/A3 requirements at the moment. These have been delayed by efforts spent for the initial version of the green field parser (in order to comply to licensing requirements) and for making improvements to the GUI and CLI applications in line with feedback from early adopters.

The first prototype of the policy checker will be incorporated into the end of March release. This has been deprioritized as veraPDFa had received, to date, no specific user requirements for policy checking.

4.2.3 EASY Innova

Being invited and obliged to submit the “end of design phase #2” report due mid of March 2016, the EASY Innova team provided their documentation on time. Apart from impressing results of their work, there are small deviations from the work plan that does not, as PREFORMA can see, the success of the development at the end of the prototyping phase #2.

As described in the “prototyping phase #1 report” (section 6) and earlier on in the technical specification of the DPF Manager, the first prototype did focus on the TIFF Conformance Checker. In the first development phase, the EASY INNOVA team provided a minimal shell implementation to produce TIFF checks. During the design phase #2, they have then redesigned the application from bottom to top. The supplier team members have studied the different available approaches about how to organize and structure the shell component taking into account the shell modularity itself, the process of decoupling the modules using event-driven communication and dependency injection.

During the design phase #2, EASY INNOVA has studied different ways to provide interoperability between the different conformance checkers. They thus designed a common report structure using preservation standards as PRIMES and METS, and they also developed a common command line interface API in order to create a first version of the conformance checker interoperability for the Open Source Workshop to be held in Stockholm in April 2016.

EASY INNOVA has evaluated the current TIFF library developed in the first prototype, and has designed a new structure using the new development patterns also used in the shell component. In the design phase #2, EASY INNOVA has also looked for improvements in their TIFF

test library, as the team wanted to create a collection of tiff files as complete as possible and promote it as a strategy to create a community around the TIFF file format.

During the design phase #2 as such, the DPF manager has been tested by the supplier's early adopter's community. The PREFORMA technical partner Packed has used the application to test 40.000 images from the Fundacio Antoni Tapies. Using the application automatic feedback feature, EASY INNOVA was able to have collected more than 8000 reports from different users and institutions like Aquafores, Oregon University, BJ Institutes, Massachusetts Institute of Technology, Antwerp Mode Museum, and University Library Dresden. Using the feedback from these tests, EASY has improved the DPF Manager and solved the bugs reported.

The creation of the new TI/A standard required the analysis of the current TIFF images stored by memory institutions. EASY INNOVA has managed to sign agreements with different memory institutions in order to get access to their TIFF master files to proceed to the analysis. This analysis is considered to be mandatory in order to create the first draft of the standard to be sent to the ISO working group

4.2.4 MediaArea

Being invited and obliged to submit the "end of design phase #2" report due mid of March 2016, the MediaArea team provided their documentation on time. A few deviations have been reported but they do not seem to threaten the success at the end of the development phases.

Through the cross-platform automated builds, MediaArea has become able to, in addition to having available the raw source code material for building the from scratch, provide daily builds for users to have the ability to download a ready-to-install copy of the software in all three shells, regardless of the user's preferred operating system platform. To support ease of installation and development, MediaConch is available through several package systems: Homebrew, Linuxbrew, and Debian. Forthcoming development includes integration into Fedora. Planning for the integration of MediaConch into Artefactual's Archivematica has begun between the two teams, and development will begin shortly.

While seeking common video muxing errors, MediaArea collected a significantly large data set (~100,000 records) of instances of Matroska-wrapped and/or FFV1encoded video files from archive.org and samples.ffmpeg.org, respectively. This database of samples, containing metadata and links to original files for download, will soon be made available in a small web application with faceted browse and search. This site will be available for all MediaConch users to be able to find and identify files that complement their testing needs. From this corpus we've been running a draft of the implementation checker on the entire dataset using MediaConch's configuration with curl so that the entire data set does not need to be local. Thus we are prepping to use the full output of the implementation in the test file faceted site to allow filtering of various errors and relation of them to other qualities such as the original muxer. The large test corpus provides real world examples of implementation errors. A small collection of synthetic files supplements this collection for instances where a real world example is not identified.

MediaConch has been establishing its presence in a variety of communities that have interest in the developing software by presenting or being represented at many different conferences. Please see below for examples and links to more information, slides, or video from these conferences. Finally, MediaArea has continued to build a community via the IETF (Internet Engineering Task Force) CELLAR (Codec Encoding for LossLess Archiving and Real-time transmis-

sion) working group, where work is being done to support the standardization of Matroska and FFV1. The working group's mailing list has been very active with members of the core Matroska community working on further developing the latest specification for standardization along with the MediaArea team.

4.3 SUMMARY

The PREFORMA CSA consortium partners could conclude that the three supplier teams were perfectly continuing to fulfill the needs of the project, in line with their own description of work, and get closer to each other in terms of the interoperability between the shells, the modules, the checkers, and the standardization efforts within and outside PREFORMA. In general, the PREFORMA partners got the impression that the three supplier teams are able and willing to continue on the road to the prototyping phase #2, the final formal evaluation at the end of the third project year, and the transfer of the developed software modules to the testing phase in the last project year.

5 THE SUPPLIERS' OPEN SOURCE WORKSHOP

The following paragraphs aim at introducing the expectations and intentions of the PREFORMA consortium to the so called Stockholm week and, with regard to the suppliers' part, to the Open Source Workshop to be held on April 7th.

Apart from presenting the results of the prototyping phase #1 as well as the design phase #2, the suppliers were expected to present their developed open source software solution in the instance of a demonstration, and in addition they were asked to provide the PREFORMA consortium as well as the audience of the workshop with the suppliers' joint approach to interoperability of the solutions, the checkers, the shells, and the APIs to develop.

5.1 REHEARSAL AND PREFORMA PRE-WORKSHOP

In order to prepare the PREFORMA consortium members and the suppliers alike for the Open Source Workshop presentations, PREFORMA had scheduled a rehearsal meeting in the afternoon of April 6th.

Two agenda items were foreseen: the reviewing, analyzing, and commenting process to the individual presentations of the three suppliers, and the discussion on the work results with regard to interoperability and API aspects provided by the suppliers. The results and the consequences / decisions of the afternoon session as well as their influence to the work in progress will shortly be presented in the following paragraphs.

More details about the procedure, the rehearsal as such, and the agreements for the following day can be found in Deliverable D3.4.

5.2 PREFORMA OPEN SOURCE WORKSHOP RESULTS

The PREFORMA Open Source Workshop was scheduled for April 7th with the aim to provide the three supplier teams with the opportunity to inform the PREFORMA consortium members, the project officer and the appointed reviewers about the current status of the development work as well as the results achieved in the interoperability strategy. The workshop was organised as an mixture between a conference and an exhibition, with more than 100 registered participants, so that both the community of open source software developers and the user communities (mainly interested memory institutions) got a unique opportunity to not only learn about the achievements but to also discuss about the progress, the results, the test data sets, and the requirements of the entire user community that goes beyond the PREFORMA memory institution partners and the user community addressed by the three suppliers. As the workshop was open to anyone interested in the topics, various representatives from different disciplines were expected to participate.

5.2.1 PREFORMA Expectations and Intentions

As there is a full deliverable which reports on the preparation and on the results of the workshop [D3.4], the present section will not enter into such details, linking to the content of D3.4 when necessary. It will focus instead on the lessons learned for the continuation of the work of the suppliers.

The workshop was organised by the PREFORMA project on 7 April 2016 in Stockholm and was hosted by Kungliga Biblioteket. The overall structure of the full day workshop was to devote the morning session to presentations and the afternoon session to interaction and discussion amongst workshop participants. The aim of the morning session was to convey an overview of the PREFORMA project and insights concerning key challenges for successful open source development as perceived by the two keynote speakers, Peter Bubestinger and Dr. Till Jaeger. The aim of the afternoon session was to report on development efforts undertaken by PREFORMA suppliers with highlights on the open source tools being developed, and to offer to the suppliers an opportunity for exhibiting their tools, thus stimulating interaction and dialogue between the suppliers themselves and the other participants.

5.2.2 PREFORMA Consortium

PREFORMA as such expected the workshop to become an essential part not only of the collaboration with the supplier teams but mainly for getting feedback from domain experts, users, lawyers, and other stakeholders invited to review the results achieved so far, and for providing feedback for the months to come. The PREFORMA CSA partners considered the workshop the concluding element of the first prototyping phase as well as for the re-design (design ‘2) phase, in order to finalize the informal evaluation as well as kicking off the preparation of the prototyping phase #2 lasting until the end of project year 3. Additionally, and as the prototyping phase #2 will be completed by the last formal review, the workshop presentations were expected to show the elements of the future road mapping for each supplier team, for the joint efforts in terms of interoperability and standardization, and the guidance of the suppliers by members of the PREFORMA CSA part. Last but not least, the workshop was expected to give an indication on how to prepare the experience workshop in Berlin at the end of project year 3.

As part of preparing and promoting the workshop, PREFORMA Communication Team designed and developed a dedicated website, periodical newsletters and press releases and a series of promotional materials which have been shipped to the partners for reaching out to potential attendees in different countries and different domains. PREFORMA representatives actively promoted the workshop also through social media channels and through already established networks.

5.2.3 veraPDF

Being invited and requested to actively attend the Open Source Workshop in Stockholm, the veraPDF team provided the following information (cfr. [D3.4]).

The veraPDF supplier team was represented by Joachim Jung, Boris Doubrov and Carl Wilson during the presentation entitled “veraPDF: definitive, open source PDF/A validation for digital preservationists”. Joachim Jung introduced the veraPDF consortium, which is jointly led by Open Preservation Foundation and PDF Association. The veraPDF software implements a conformance checker for the text file format PDF/A. The presenter explained that veraPDF consortium contributes to standardization efforts related to ISO 19005 variants through the PDF Association technical working group. Boris Doubrov addressed the challenge and complexity of PDF/A validation. This involves use of a suite of test files (such as the Isartor test suite) and establishing ground truth, a test corpus, and validation profiles. An example profile was shown and how resolution of ambiguities can be resolved. The presenter demonstrated the GUI for the con-

formance checker. Specifically, it was shown how a file can be tested and a report can be generated and inspected. Carl Wilson showed the CLI version of the software, and specifically a batch processing job where a series of pass and fail PDF (or XML) files are created. The web demo was also shown and described by the presenter together with the GitHub site for the different subprojects. The speaker also presented the installer and a progress report illustrating the status of different activities in the project. Ambitions concerning community building were expressed, and the outreach and next steps for the project were explicated.

For details of the presentation “veraPDF: definitive, open source PDF/A validation for digital preservationists” by VeraPDF, please see Deliverable D3.4 Appendix B.

5.2.4 EASY Innova

Being invited and requested to actively attend the Open Source Workshop in Stockholm, the EASY Innova team provided the following information (cfr. [D3.4]).

Easy-Innova was represented by Miquel Montaner (CTO), Xavier Tarrés Bonet, Peter Fornaro, and Josep Lluís de la Rosa during the presentation entitled “DPF Manager – The open source Community”. Miquel Montaner described the consortium and people involved in DPF Manager. The consortium consists of EasyInnova company (located in the Spanish town of Girona) which has a focus on the software development, the University of Basel in Switzerland which focuses on standardisation processes and aspects, and the University of Girona which has community building aspects as its main focus. Xavier Tarrés Bonet explained that DFP manager is conformance checker for different versions of the TIFF image format. The speaker stated that the software can be executed on multiple platforms in different ways, it exhibits a modular architecture, and uses a number of technologies (e.g. Java, Maven, Spring, Travis CI, JavaFx, and Jacpfx). It was also explained that DFP manager is an open source project released under licenses MPLv2+ and GPLv3+ which is provided at a company website, at the PREFORMA portal, and at GitHub. Peter Fornaro explained file format preservation and described important properties of file formats. The TIFF format and its use and importance for memory institutions was elaborated on. Further, the TI/A initiative was introduced as a TIFF format for archival. The involvement in TI/A standardisation was described, including the current status and timeline for the standardization process. Josep Lluís de la Rosa explained the challenge of community building. Specifically, it was reported that there are a number of registered early adopters of the DPF managers who have checked files and provided feedback. The presenter stated that there are TIFF experts involved (69 experts in 16 countries), and support of 61 memory institutions. Finally, the consortium announced their interest in further extending collaboration with new early adopters, TIFF experts, and software developers.

For details of the presentation “DPF Manager – The open source Community” by Easy-Innova, please see Deliverable D3.4 Appendix B.

5.2.5 MediaArea

Being invited and requested to actively attend the Open Source Workshop in Stockholm, the MediaArea team provided the following information (cfr. [D3.4]).

MediaArea was represented by Jérôme Martinez (CEO), during the presentation entitled “MediaConch – Implementation and policy checking on FFV1, Matroska, LPCM, and more”. Jérôme

explained that MediaConch is a conformance checker focused on video and audio content analysis, and is divided into an implementation checker and a policy checker. Examples of reports were shown containing different levels of detail, allowing for both high- and low level results from conformance- and policy checking. Further, the editor for creating user defined policies was presented. The presenter explained that MediaConch operates using different interfaces (graphical interface, web interface, command line, server, and library) and that output can be obtained in different formats (including XML, text, and HTML). It was also highlighted that the software is open source (released under licenses GPLv3+ and MPLv2+), relies on the MediaInfo metadata extraction tool, and uses various libraries (including Qt, sqlite, libevent, libxml2, and libxslt). Supported file formats are primarily Matroska, FFV1, and PCM (but other media formats supported by MediaInfo can be used with the policy checker). The presenter also stated that there is support for the VeraPDF plugin and DPF Manager plug-in, and that input to the checkers can be provided through local files, FTP/FTPS/SFTP, HTTP/HTTPS, and Amazon S3. It was explained that binaries are available on different platforms (including Windows, Mac, Linux, and potentially on embedded devices such as Raspberry Pi). Concerning standardisation efforts, the presenter announced involvement in the IETF workgroup CELLAR which aims to standardise Matroska, FFV1, and FLAC. It was also explained that the MediaConch project is co-led by the American archivist Dave Rice, and that the project is being presented at different events worldwide. The Matroska research corpus of video files was introduced, which is used to test real life files. The presenter also outlined the plans for further improvements of MediaConch, including improved GUI and functionality. It was explained that the future of the software beyond PREFORMA sponsorship depends on the community, its user requests, development efforts and support. Finally, different examples from use of the software in different scenarios were demonstrated (including online use, command line operation, and through use of plugins).

For details of the presentation “MediaConch – Implementation and policy checking on FFV1, Matroska, LPCM, and more” by MediaArea, please see Deliverable D3.4 Appendix B.

5.3 SUMMARY

With more than 100 participants representing various domains and sub-domains, the open source workshop can be considered a big success story in the lifetime of the PREFORMA project. Not only did the organizers get well-esteemed key note speakers – the presentations of the three supplier teams were of the same level of quality and opened the floor for many questions by the audience. Even if the maturity level of the three software packages slightly differed, PREFORMA as a whole got the impression that (a) all three supplier teams are on a very good way to achieve their goals, (b) the collaboration of the three teams in terms of interoperability of the solutions had significantly increased in terms of quality and quantity, and (c) the different approaches in terms of contributing to the standardization of the work formed a kind of network with all supplier teams benefitting from each other. The identified “type” of supplier team explained before (business people, hacker, and standardization experts) had an impact on the presentations as well as on the level of maturity but provided PREFORMA with the impression that the project itself is on a very good track, too.

6 LESSONS LEARNED FROM THE DESIGN PHASE #2

Based on the lessons learned during the design phase #1 about one year ago, the lessons learned during the prototyping phase #1, and the outcomes of the design phase #2 evaluation procedures, this section shortly presents the final lessons learned for the prototyping phase #2 and for the testing phase.

This section will contain information about the summarized results of the design phase #2, about the performance of communication and dissemination means, and about further aspects important for the progress and success of the PREFORMA project as such. Therefore, almost all of the aforementioned communication means and channels will be analyzed and evaluated internally, drawing conclusions for the progress of the project, especially for the prototyping phase #2 immediately following the design phase #2.

6.1 VIRTUAL MEETINGS

Similar to the previous phases, the means of virtual meetings between PREFORMA team members and the supplier teams turned out to be as efficient as the ones held in the design phase #1 as well as in the prototyping phase #1, and it is foreseen to continue the series of virtual meetings in the prototyping phase #2.

PREFORMA remains on the decision to give all suppliers at the same time the chance to raise questions, to come up with suggestions, and to discuss – even with the other supplier teams – about interoperability, about the shells, and about strategies to make sure the designed and developed modules and components were able to communicate with each other during and after the prototyping phase #1. It shall be repeatedly mentioned here that It had been announced to all suppliers that specifications, achievements, and results will be made publicly available after the award decision. So continuing this appropriate information chain procedure was one of the first steps PREFORMA as well as the three supplier teams decided to continue, and it turned out to be necessary to have established a set of a very efficient ways of communication.

In order to support the information exchange prior to the virtual meetings but also for submitting the meeting minutes, PREFORMA continuously used the mailing list for supplier communication. Any communication from a particular PREFORMA consortium member to the particular supplier could easily be followed and archived by all consortium members and all suppliers.

As an example, the minutes (shortened) of the virtual meeting between the PREFORMA consortium members and the suppliers are shown in the following:

PREFORMA	*Easy Innova*	*MediaArea*	*veraPDFa*
Antonella Fresa (Promoter) Bengt Neis (KB) Börje Justrell (RA)	Miquel Montaner Xavier Tarrés	Jérôme Martinez	Becky McGuinness Boris Doubrov Carl Wilson

Erwin Verbruggen (IBG) Magnus Geber (RA) Peter Pharow (FH) Sònia Oliveras (AjGi)			Duff Johnson Joachim Jung

Introduction and report of the visits to the suppliers

Börje Justrell gave an overall summary of the visit to the suppliers. The visits were a positive experience with each supplier displaying their own characteristics and approach to the PREFORMA project. The other members from PREFORMA who attended the meetings agreed with the assessment, and gave a short account of their impressions. Each supplier then shared their view on the meetings, which was considered a positive exercise.

Open Source Workshop; update on the program of the April 6 meeting

Bengt presented the planning of the Stockholm Workshop on April 6 at KB (Kungliga Biblioteket; the National Library of Sweden). The location for the first day, for preparations and internal meetings, will be at “garnisonen” (the Garrison).

The plan is to hold two separate, but concurrent, meetings between **09:00-12:00**. The PREFORMA consortium needs to have an internal meeting in preparation for the review. There will however be a room available for the Suppliers to meet, work and discuss, e.g. interoperability/API, and prepare for the workshop. Lunch will be held at **12:00**.

A joint meeting, with PREFORMA and the Suppliers, is scheduled at **13:00-18:00**, and divided into two parts. The focus on the first half of the meeting, **13:00-16:00**, will be on the presentations of the Suppliers along with identifying progress and gaps, work still to be done, according to the DoW per supplier; all suppliers are expected to be present and asked to comment, contribute. The second half, **16:00-18:00**, will focus on the interoperability/API aspects.

Requirements for presentations, demos and booths at Open Source Workshop (7 April)

The location for the second day, the actual workshop, will be at KB’s main building. Requirements for demos should be due three weeks in advance, and sent, together with the presentations to: bengt.neiss@kb.se.

Improvement in the communication PREFORMA-suppliers

In response to a request by the suppliers for an even better feedback process from PREFORMA, an e-mailing list for technical and interoperability issues has been set up. Benjamin gave a short description and purpose of the TWG (Technical Working Group) e-mailing list. The purpose is to stream-line the communication between PREFORMA and the suppliers.

While the PREFORMA consortium has limited resources, another attempt to improve the close working relationship with the suppliers, and thereby improve the feedback process, is to set up a liaison with each supplier and two PREFORMA members. Suppliers had no comments.

MediaArea would like to have more direct communication with individuals, like archivists, engaged in the project and feedback process, preferable using GitHub's tracker. veraPDF concurred and added that GitHub interaction is the best channel for making improvements.

It was suggested that PREFORMA should, within each respective member institution, promote engagement from individuals to use the CC (Conformance Checkers) and provide direct feedback to the suppliers; instead of PREFORMA gathering all incoming feedback and sending it to the Suppliers at a later time in a batch.

Status of the integration DPF Manager – Technical Space

Antonella Fresa described the Europeana Space project, and the integration of DPF Manager with Europeana Technical Space. Europeana is a repository of metadata that is aggregated from different cultural content repositories. See www.europeana-space.eu, and "MINT".

EasyInnova, represented by Xavier, and ETS (Europeana Technical Space), represented by Mario and Nasos, met on Skype the second of February 2016, to discuss what kind of integration they prefer and to plan the next steps. Since ETS uses Java, they preferred to use DPF Manager as a Java library. ETS also informed that they do not want to use the DPF manager policy checker nor the TIFF CC; ETS is only interested in extracting metadata.

They agreed that the reports from the different checks performed by ETS using DPF Manager will be sent automatically to the EasyInnova servers. Such reports are very useful for testing and validating that the DPF Manager works properly.

Both parties are currently very busy with "hackathons", thus they will begin work on integration at the beginning of March. Meanwhile, EasyInnova is preparing the specification of the API and documenting how the DPF Manager should be integrated with ETS. The ball is now in the ETS's court, thus maybe Antonella can push them to have some results/demo for the workshop.

Status of the work on interoperability

During the Suppliers meeting, discussions about interoperability/API resulted in some notes and conceptual ideas. The notes, together with other documents and an illustration, have been sent by e-mail to all Suppliers (20160211, subject: PREFORMA; Technical Working Group; PIF). The Suppliers' first draft of the, now called, PIF (PREFORMA Interoperability Framework), was made available on GitHub during the veraPDF meeting https://github.com/preforma/interop-sandbox

The draft has been forked; introducing a guiding principle for the, now called, PEA (PREFORMA Extension API) https://github.com/beyo-ra/interop-sandbox/blob/master/documentation-api.md.

There have not yet been any comments from the Suppliers, or any additional work on the interoperability/API parts. The Suppliers are currently concentrating on the end of redesign and accompanying release, and will focus on the interoperability/API as soon as possible.

PREFORMA glossary

In an effort to improve the communication with the Suppliers, and third parties in general, PREFORMA has introduced a glossary of terms and definitions used in the project. The glossary is a

continuous work in progress and will be updated and adjusted throughout the project's lifetime http://www.digitalmeetsculture.net/projects/preforma/virtuality/

The suppliers can consolidate the glossary, and request PREFORMA to make a specific definition, or a clarify a definition that is interpreted as unclear or ambiguous.

Next meeting

March 23th 2016, 14:30-16:30 CET.

6.2 VISITS TO THE SUPPLIERS' TEAMS

As explained in chapter 1, PREFORMA decided to visit the premises (in case this was possible) of the three supplier consortia. For various reasons, the visits turned out to be very interesting and promising, as both the project consortium and the suppliers learned a lot while discussing open aspects, the review results, the interoperability aspects, the community building, the standardization efforts, and many other aspects, as it can be seen from the agenda versions of the three events in annex 3.

The following sections contain a short summary of the results of the three visits, and relations to the progress made by the suppliers. As the visits had several purposes, among others the interoperability achievements and the negotiation of aspects found during the analysis of the documentation provided by the supplier so far, results might be repeated here but have the intention to better provide an overview of all the results achieved.

6.2.1 veraPDF

The feedback from PREFORMA to the agenda was raw comments from PREFORMA partner reviewers. They have the same template as the suppliers for guidance and compare progress between reports, but can otherwise comment freely.

The veraPDF team has a very good understanding of the technical issues, but the resources are difficult to navigate. The website and documentation needs to be improved to help users. The website is difficult to navigate. It is difficult to find links e.g. online demo. KEEPS has written a report that not only addresses technical issues, but also makes a number of suggestions to improve the structure and layout of the website. Becky and Joachim will review it over March before the April meeting.

The veraPDF team asked that we have a shorter feedback loop between PREFORMA and veraPDF. It would be useful if someone from PREFORMA could take ownership of the issues, especially for small things.

The PREFORMA members have proposed having one reviewer per supplier (ideally they would have two, one from a technical partner and one from a memory institution). They would have a look at progress each month so they are aware of what is going on so that someone is not starting from scratch after a long period.

6.2.2 EASY Innova

For the meeting with EASY INNOVA, the possible collaboration with the Europeana Space project[1] was seen as the main aspect for the meeting.

The aim of this cooperation is to test the integration of the open source conformance checkers developed in PREFORMA in the technical infrastructure developed in Europeana Space, which would allow to: provide users with additional search options based on the information extracted by the PREFORMA tools; ensure that the files stored in the Technical Space are in a format that can be accessed in the future; check whether a file conforms with specific policies that are set by Europeana Space content providers, e.g. in terms of copyright.

A representative from NTUA (National Technical University of Athens) participated in the meeting in Girona to present the Europeana Space project, including the development of the technical infrastructure, information about open content API, licensing right around the content, events business development and hackathons. In particular, he detailed the so called Technical Space[2], a web based application for the development of applications and services based on digital cultural content, and the embedded Media Checker.

The current library is used to extract the image metadata using different utilities as imagemagick, itext and ffmpeg.

In the discussion about the Integration of the DPF Manager and PREFORMA software in the Europeana Space Media Checker, all available use cases of the DPF Manager were analyzed. PREFORMA could take advantage of the use of java in both projects to use the DPF Manager as an API library.

The DPF manager will be used in Europeana Space to extract information about the metadata and not to define acceptance criteria for TIFF images submission.

EASY INNOVA then proposed the use of DPF Manager to generate an XML report of every TIFF image and store this XML report file. Europana Space developers could use the report information to show information in the website.

EASY INNOVA explained the possibility to receive feedback from all the checks done in the Europeana Space project in order to analyze the reports and search for errors, uncommon TIFF structures or tags, as we are doing now with the our early adopters community.

All participants agreed that, despite the integration with Europeana Space, EASY INNOVA should take into account the interoperability planning. The integration should be ready for the April Stockholm workshop.

[1] Europeana Space project: http://www.europeana-space.eu/

[2] WITH Technical Space: http://with.image.ntua.gr/assets/index.html

6.2.3 MediaArea

During the third visit, the PREFORMA team discussed with members of MediaArea.net several topics, among which: status of the conformance checkers; capabilities for software; evolving functionalities including interoperability and API; usability and user requirements, user groups, scenarios; testing for quality assurance and accuracy; code transparency, media files, and more; achieving reference implementation; awareness of what is still missing; status and progress in format specifications; reporting formats; scalability and optimization for large file sets; collaboration with other projects (such as for the integration of MediaConch with Archivematica) and with the open source community; status of standardization activities.

The main lesson learned was about the use of and feedback on the MediaConch software from the PREFORMA consortium's part. A problem was that evaluation/negotiation forms take time, and issues are often solved before reported. GitHub invites to a more conversational setting.

The PREFORMA consortium has on several occasions been urged to directly contact the Suppliers. The PREFORMA consortium is however a consortium, that is, it constitutes of different members and individuals, interested in and working on different questions with a limited time and resources. Engaging the Suppliers on GitHub is for some cumbersome, and e-mail is for others a preferable method of communication.

The evaluation/negotiation report is one way to focus the attention of the PREFORMA members to produce at least some form of feedback.

6.3 INTEROPERABILITY BETWEEN SUPPLIERS' WORK

The PREFORMA project has from the outset aspired to ensure that the three conformance checkers (CCs) can interact and work with each other, and other applications, systems, and in various environments.[3] The following text intends to give a brief reflective overview of the challenges met when working with interoperability in the PREFORMA project, the approaches to handle the problems, and the experiences gained.

The importance of interoperability had early on in the project been outlined in documents such as the *Challenge Brief* (v. 1.0).[4] In that document the architecture of the CCs and the relationship between them was described and illustrated.[5] There were however many details missing necessary for a technical implementation. It was therefore needed to clarify some concepts, and conceptualize new ones.

[3] See e.g. D2.1 *Overall Roadmap* (v. 2.5), sec. 3; D2.2 *Tender Specifications* (v. 2.1), sec. 4.2.1.

[4] Cf *Tender Specifications* (v. 2.1).

[5] See *Challenge Brief* (v. 1.1), sec. 4.4.

The three suppliers have been asked to start working on these aspects and prepare a proposal to be presented at the end of the re-design phase. PREFORMA assists to the extent needed for the Suppliers to complete their task,[6] and exploratory work was done on the subject.[7]

The focus therefore shifted from writing a specification to directly "implementing interoperability" using a pragmatic approach. The task proved somewhat more challenging than expected. In hindsight, the unique constellation of the PREFORMA project, a consortium contracting three independent suppliers, all with different competence and outlook on the project, may have been a decisive factor. The experience however should be valuable for coordinating future projects that are organized like PREFORMA.

At the abstract level there are the CCs. Conceptually they are supposed to have the same purpose and function, however, in practice they focus on three different file formats, which are used for three different purposes, and utilizing different technologies. At the concrete level there are the programming languages which form the implementation of the CCs. From an interoperability point of view a CC has to work with other CCs and programs written in other, existing or future, programming languages. The questions were therefore: *how* to communicate and exchange information (which protocols to use and defining an API); and *what* information and formats to exchange (defining the data objects and selecting an appropriate data exchange format)?

A flurry of activity took place over a period of time. E-mail correspondence and virtual meetings were complemented with physical meetings when, and where, possible.[8] The outcome of the activities was that PREFORMA provided a concrete *Use Case for Interoperability*,[9] and the Suppliers were given the task to draft a formal specification for the PREFORMA Interoperability Framework (later referred to as PIF).

One important event was the PREFORMA meeting in Pisa, which was attended by PREFORMA members who were involved in the interoperability work up until then, and the PREFORMA Project Management Team. Many conceptual issues were at that meeting clarified, which paved the way for a better communication with the Suppliers.[10] Another turning point came during the Suppliers' meetings, where PREFORMA had a chance to meet most of the Supplier members, or at least key persons representing them. Scheduled for a day, each meeting covered a wide range of topics, presenting an opportunity to discuss interoperability more fluently and interactively, and in a wider context.

The result from the meetings and all the work so far has been the establishment of a Technical Working Group (TWG), with one permanent representative from PREFORMA authorized to

---

[6] D8.3 *First Prototype Report* (v. 1.0) sec. 3.3, *Interoperability of different conformance checkers*.

[7] See e.g. the Legal opinion, sec. 12, *Code*, and sec. 14, *Interoperability*.

[8] For example, Suppliers meeting in Brussels 20150525 (D8.3 *First Prototype Report* (v. 1.0) sec. 2.2.2, see *First virtual meeting*).

[9] D8.3 *First Prototype Report* (v. 1.0) sec. 2.2.2, see unnumbered subsection *First virtual meeting*.

Use case for interoperability, 20150612.

[10] PREFORMA Plenary Meeting Minutes 20151008-09. The overview of interoperability/API, 20151016.

make decisions and lead the continued development. The TWG is drafting the PIF on GitHub,[11] and as of now conceptualized it as PIA (PREFORMA Integration API) and PEA (PREFORMA Extension API).

To summarize the progress: The conceptualization of the interoperability required insight in technical matters, where those with the concrete technical skill-set had three different outlooks and formats to work with, while those with the overall vision had a more abstract technical skill-set. In order to bridge that gap between all parties, designating ownership and establishing forums for communication, physical meetings, and writing a formal specification proved invaluable. The progress of ensuring the interoperability of the CCs has therefore been a progression of incremental steps that have been taken by different working groups.

6.4 END OF DESIGN PHASE #2 REPORTS

Summing up what has been said earlier in this report, the results of the reviews performed by up to seven experts per medial file type application were very promising. Therefore, the expectations to the three reports provided at the formal end of the design phase #2 were quite high, and all three suppliers provided high quality reports. Some of the aspects addressed in the reports (see chapter 4.2) became part of the preparatory work for the Stockholm week, the PREFORMA PMT meeting, the suppliers meeting, the rehearsal of the project and suppliers teams, and last but not least the Open Source Workshop.

6.5 SUPPLIERS OPEN SOURCE WORKSHOP

With more than 100 participants representing the important and interesting aspects from various domains and sub-domains all related to digital preservation and archiving technologies, rules, and policies, the Open Source Workshop held in Stockholm on April 7th, 2016, can be considered a real success story in the life time of the PREFORMA PCP project. The structure of key notes in the morning and supplier teams' presentation and Q&A parts in the afternoon appeared to be a perfect structure for allowing the teams to present, demonstrate, and explain, and allowed the expert audience to ask for clarification, progress, other file formats, and many other aspects of particular interest for the domain.

Not only did the Open Source Workshop organizers get well-esteemed key note speakers for the supplier teams to learn – the presentations of the three supplier teams, in the light of the key notes, appeared to be at the same level of quality, contained demonstration elements during or after the explanations, and thus opened the floor for many questions by the audience.

Even if the maturity level of the latest versions of the three software packages slightly differed, the PREFORMA CSA partners who participated in Stockholm got the impression that all three supplier teams managed to be on a very good way to achieve their overall goals. With about half a year to go, PREFORMA expects the suppliers to deliver on time and with the expected

[11] github.com/preforma/interop-sandbox (20160324).

quality. The workshop furthermore clearly showed that the collaboration of the three teams in terms of interoperability of the solutions had significantly increased in terms of quality and quantity. PREFORMA expects the interoperability tasks to be solved before the start of the testing phase. Last but not least, the different approaches in terms of actively contributing to the standardization of the work allowed the three supplier teams to form a kind of internal network with all supplier teams benefitting from the experience of each other.

Eventually, the identified three “types” of the supplier teams explained before (the business people, the virtual hacker team, and the network and standardization experts) had an impact on the presentations as well as on the level of maturity but provided PREFORMA with the impression that the project itself is on a very good track, too.

6.6 SUMMARY

PREFORMA achieved a lot of lessons learned in the context of the design phase #2 performed from November 2015 until April 2016, and completed by the Stockholm PREFORMA week.

A rather efficient mechanism, already successfully performed during the design phase #1 as well as during the prototyping phase #1, was the establishment of frequent virtual meetings with the supplier teams and selected members of the PREFORMA CSA consortium. Usually run once a month, the virtual meetings provided both the supplier teams and the PREFORMA consortium members with a unique opportunity to raise important or hot topics and questions, and to ask for progress reports in an informal way, adding information to the more formal reports requested less frequently. The virtual meeting mechanism will therefore be kept going on during the prototyping phase #2, and perhaps, with a different focus, during the testing phase.

Real meetings were foreseen to complement the virtual meetings. Real and face2face meetings were always scheduled and performed in the context of meetings that were planned anyway, e.g. the Stockholm week. These meetings were usually considered to negotiate the development work results, to rate and value the status of interoperability and standardization activities, and to allow the supplier teams for meeting each other even independently from the presence of the PREFORMA CSA consortium members. As both parties considered this mechanism an important means of collaboration, PREFORMA will keep running these physical meetings, e.g., in the context of the Berlin experience workshop week.

The visits of PREFORMA CSA team members to the supplier teams were something that PREFORMA did not have in mind when planning the activities for the design phase #2. Both the suppliers and the PREFORMA CSA members learned a lot about each other from these visits to the three selected supplier teams, their members, their premises, and their way of working. It was a good opportunity to work together, check progress, exchange ideas, and plan the next steps.

Last but not least, the Stockholm Workshop formally concluded the design phase #2. Having received all relevant documents, software releases, test and training data files before running the workshop, the presentations, demonstrations and discussions which took place between the supplier teams and the audience provided a lot of additional information that complemented the informal review results, providing a very good picture of what have been achieved do far by the suppliers.

7 CONCLUSION

PREFORMA got the impression that the three selected supplier teams are perfectly able and willing to perform the task PREFORMA had applied to them. During the course of the design phase #2, the collaboration between the suppliers in terms of interoperability and standardization has significantly increased, as demonstrated during the Open Source Workshop in Stockholm.

The prototyping phase #2 will mainly base their procedure, their collaboration, their communication as well as the presentation of reports, results, documents and software on the findings of the design phase #2. A lot has been said about the procedure of the design phase #2 as well as on different results and lessons learned that can easily be found in the previous chapters.

Intermediate software releases and documentation will help both PREFORMA and the suppliers to even improve quality and quantity of the results. PREFORMA will support the supplier teams by appointing technical and domain experts to significantly increase the timeliness and quality of the evaluation in a way that the suppliers can rely on timely feedback to each and every single document or release. This, as a lesson learned from the past phases, will support both PREFORMA and the suppliers in achieving the overall goals of the project as such.

ANNEX 1: ASSIGNMENT OF REVIEWERS

The following list reflects the allocation and assignment of evaluation experts to media file types and suppliers for the informal “design phase #2” review activities:

Media format	Supplier	Technical experts	Domain experts	External experts	Open source experts
Text	VeraPDFa Consortium	FRAUNHOFER PACKED RA (Benjamin)	RA (Magnus) EVKM LGMA	Jozo IvanoviĆ (National Archives of Croatia)	University of Skövde
Image	Easy Innova - IMAGE	FRAUNHOFER PACKED RA (Benjamin)	SPK KB KIK-IRPA	Jan Dalsten Sørensen (National Archives of Denmark)	University of Skövde
Av	MediaArea.net	FRAUNHOFER PACKED RA (Benjamin)	S&V AJGI GFC	Peter Bubestinger, Hermann Lewetz (Österreichische Mediathek)	University of Skövde

In terms of assignment of suppliers and reviewers, PREFORMA suggests to keep the same formal assignments that were used for the design phase #1, i.e.

- RA, EVKM, and LGMA have evaluated VeraPDF project as domain experts.
- SPK, KB, and KIK-IRPA have evaluated Easy Innova’s DPF as domain experts.
- S&V, AJGI, and GFC have evaluated MediaArea’s MediaConch as domain experts.
- FRAUNHOFER and PACKED will have a look at all the three reports as the PREFORMA project’s technical experts.
- RA (Benjamin Yousefi) reviewed all three sets of documentation mainly from an interoperability viewpoint.
- HS had a look at all documents from an open source viewpoint.
- Jozo Ivanović have evaluated the veraPDF project as external expert.
- Jan Dalsten Sørensen have evaluated Easy Innova project as external expert.
- Peter Bubestinger and Hermann Lewetz have evaluated MediaArea as external expert.

ANNEX 2: FEEDBACK FORM ON FINAL RELEASE - OCT 2015

The software releases are available in the Open Source Portal (http://www.preforma-project.eu/open-source-portal.html). The accompanying reports (both the final one and the intermediate one) are available in the project's repository under the /Tender/First prototyping phase/Evaluation. Still in the repository (/Tender/First design phase/Evaluation folder) it is possible to find also the outcomes of the first design phase (specification documents provided by the suppliers and evaluation report provided by us). We recommend taking them into account when evaluating the present releases, to check whether the progress is in line with what was agreed at the end of the first design phase.

1. General comments

<If you cannot find an appropriate category below>

2. Result from examination

<Questions concerning how the Conformance Checker validates, e.g., how did file x y z validate, good, poorly, bad, etc.>

3. The Conformance Checker

<Questions concerning the Conformance Checker itself, e.g., the setup, the usage, identified bugs, interface issues, etc.)

4. Final report

<Comments in relation to the subjects in the Final Report>

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

<Suppliers shall provide the PREFORMA consortium with a concise overview of the release developed so far, and of the functionalities that are available at the time of this report. Furthermore, suppliers shall highlight which is the progress compared to the last intermediate release (July 2015) and how are they addressing the comments received from the PREFORMA consortium.>

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.>

Dissemination and community building

< Suppliers shall provide the PREFORMA consortium with the list of dissemination activities that they have undertaken to promote their open source project. Furthermore, suppliers shall describe any potential long-term collaborations/partnerships entered into.>

Open Source approach

< Suppliers shall provide the PREFORMA consortium with a description of how they addressed the relevant open source topics, best practices, and licensing and how did they progress in setting up an open source community around the developed tools.>

Standardisation efforts

< Suppliers shall provide the PREFORMA consortium with a description of how they are actively contributing to the standardization process in their domain, by means of providing feedback on the existing standards contributing as well as the way on how to support emerging standards.>

Gap analysis and next steps

< Suppliers shall provide the PREFORMA consortium with a description of the status of the work compared to what was planned in the functional and technical specification that they provided at the end of design phase 1. Suppliers shall highlight critically what it is still missing in the current release, and which are their plans to overcome the gaps.>

ANNEX 3: SUPPLIER VISIT AGENDA VERSIONS

Agenda for PREFORMA visit to veraPDF in Brussels (January 28th)

Location: Dual Lab office, Clos du Parnasse 12C, 1050 Brussels, Belgium

PREFORMA attendees: Börje Justrell, Benjamin Yousefi, Peter Pharow

veraPDF consortium attendees: Joachim Jung, Carl Wilson, Becky McGuinness (OPF), Boris Doubrov, Artem Kostyukovich (Dual Lab); (Duff Johnson from the PDF Association will possibly be available to join remotely for part of the day)

Presentation of the attendees of the visit and role call (1230 – 1330);

PREFORMA delegation and visit topics to address (1230 – 1300);

Staff of veraPDF consortium (1300 – 1330);

Lunch Break (1330 – 1430);

Discussion on PREFORMA issues (PREFORMA, veraPDF) (1430 – 1900);

- Capabilities for software;
- Evolving functionalities including interoperability and API;
- Usability and user requirements, user groups, scenarios;
- Testing for quality assurance and accuracy;
- Code transparency;
- Achieving reference Implementation;
- Awareness of what is still missing;
- Process of community building / development;
- Demo of the software
- Standardization

PDF/A Technical Working Group meeting: will coincide with this face-to-face meeting, so there possibility to drop in on meeting to gain an insight of how it works.

Dinner with all attendees of PREFORMA, veraPDFa, and MediaArea (2000 – 2300)

Agenda for PREFORMA visit to EASY INNOVA in Girona (January 20th and 21st)

Location: Easy Innova premises, Girona (location see separate document)

PREFORMA attendees: Börje Justrell, Benjamin Yousefi, Antonella Fresa, Peter Pharow, Sònia Oliveras i Artau

Europeana Space attendees: Marios Phiniketos

Easy Innova consortium attendees: Josep Lluís de la Rosa, Miquel Montaner, Xavi Tarrés, Robert Salló, Víctor Muñoz, Lukas Rosenthaler (UniBas), Peter Fornaro (UniBas)

Welcome coffee (9:00-9:10)

Presentation of the attendees of the visit (09:15 – 10:00);

PREFORMA delegation and visit topics to address (09:20 – 09:40);

Staff of EasyInnova consortium including Basel University (09:40 – 10:00);

Integration of PREFORMA image solution into Europeana Space (discussion led by Marios, 10:00 – 12:00);

Visiting the premises of Easy Innova and a quick tour of the Scientific & Technological Park (12:00 – 13:00);

Lunch Break (13:00 – 14:00);

Discussion on PREFORMA issues (PREFORMA, Easy Innova) (14:00 – 18:00);

- Capabilities for software;
- Evolving functionalities including interoperability and API;
- Usability and user requirements, user groups, scenarios;
- Testing for quality assurance and accuracy;
- Transparency of code, media files, and more;
- Achieving reference Implementation;
- Awareness of what is still missing;
- Process of community building;
- Status of standardization activities.

Touristic Visit to the Girona Old Town (19:00-21:00)

Dinner with all attendees (21:00 – 23:00)

Visit to Girona City Archive (managed by Sonia, January 21st, 09:00 – 11:00);

Agenda for PREFORMA visit to MediaArea in Brussels (January 29th)

Location: Meeting will take place at Hotel Brussels, "Louise" meeting room, Avenue Louise 315, 1050 Brussels Belgium, +32 264 02 415

PREFORMA attendees: Börje Justrell, Benjamin Yousefi, Peter Pharow, Peter Bubestinger, Erwin Verbruggen

MediaArea consortium attendees:

Presentation of the attendees of the visit and role call (1000 – 1100);

PREFORMA delegation and visit topics to address (1000 – 1030);

Staff of MediaArea consortium (1030 – 1100);

Discussion on PREFORMA issues (PREFORMA, MediaArea) (1100 – 1600);

- Capabilities for software;
- Evolving functionalities including interoperability and API;
- Usability and user requirements, user groups, scenarios;
- Testing for quality assurance and accuracy;
- Code transparency;
- Achieving reference Implementation;
- Awareness of what is still missing.
- Status and progress in format specifications, relationship to implementation checker
- Reporting formats
- Scalability and optimization for large file sets
- Integration of MediaConch with Archivematica
- Collaboration with other projects (VeraPDF, DPFManager, and others)
- Collaboration with the open source community (Debian, Ubuntu, Fedora, EPEL)
- Build environments and living distros having their own build environment (i.e. Linux distros with update by Internet)

Lunch Break (in between, e.g. 1300 – 1400);

ANNEX 4: NEGOTIATION REPORTS TO SUPPLIERS

Feedback on the final release - Oct 2015 - veraPDF

In this document we have compiled the comments that we received so far by the members of the PREFORMA Evaluation Committee on the final release of the first prototyping phase (end of October 2015), including those on the text in the Final report.

The comments are normally taken in to the document more or less in the shape they were sent in, e.g. without filtering and editing, and relate both to specific and more general issues. Sometimes they might reflect different opinions but overall we hope that they provide a useful input for the next period.

The aim is that these comments shall serve as a base for further discussions.

1. General comments

veraPDF has created a platform for the open source project based on the GitHub platform. The different tools GitHub provides for open source projects fills most of the different needs of communication within an open source project. However, the community is not tools and technology but people cooperating to reach common goals. In that respect the activity of the community is very low and it is important that the work to achieve this is increased. One of the main challenges for the project is to have a thriving community so that the software will live on when the Preforma project ends.

2. Result from examination

The Conformance Checker (CC) runs okay in my estimation. I used it on my regular work PC, downloaded the files from the Internet, and ran a few tests using files from the Bavaria and Isartor collections.

Validation of PDF/A-1b compliant files against PDFA-1b profile is mainly good. Validation of noncompliant files ends with an execution exception in processing (the report is not accessible via GUI).
Validation of PDF/A-3 files against PDFA-3b profile hasn't been successful.
Reports do not declare the profile correctly whenever an non-PDFA-1b profile is selected.

The latest release should be at http://verapdf.org/software/ but there at least the windows version does not work at all. So what we tested we don't know if it is the latest version.

I have downloaded – from http://downloads.verapdf.org/rel/[12] - and tested the 06.46 version of the Vera Pdf, dated 4 nov, and checked conformance of PDF-files by using the Bavaria test suite as well as the Isartor test suite. The Isartor test suite is a collection of synthetic test files, whereas the Bavaria test suite is a collection of both synthetic and real world test files.
While testing the Quick Start Guide has been used. The interface at page 5 look different from what I get, maybe because of different version(Windows vs Mac).
When starting two error messages occurs in the DOS-window.
Checks have included running the software by checking test files defined as being in conformance with the ISO-standard, as well as those that are not. The software provides accurate results. Messages from the software include “PDF File is not compliant with Validation Profile requirements” and the opposite “PDF File is compliant with Validation Profile requirements”. When testing I received a machine readable report by clicking on the “View XML” option, and tested the option to save the XML-file in XML-format, which worked out just fine. The software provides the end user with four profiles to choose from: PDF/A 1 A, PDF/A 1 b, PDF/A 2 b and PDF/A 3 b. Checks of policy requirements as a functionality is not yet available.
The Metadata Fixer is a component which allows for simple fixes of the metadata embedded In the file, making the compliant with the standard specification. I´ve done a test of this, using a file with “fail” status from the “vera pdf corpus” included in the download of the software):

Validation Profile:	PDF/A-1B validation profile
Compliance status:	**Failed**
Statistics	
Processing time:	00:00:00.105
Total rules in profile:	95
Passed Checks:	404
Failed Checks:	1
Metadata Fixes Status:	Success
Completed Metadata Fixes:	2

However, for some test files, no metadata fixing occurs and no error message is given. When metadata fixing is reported new files are actually created and they are different looking at the file size. But the new file don’t seem to be fix in a correct way as the do not validate without error.
In the report is stated that the fixer should be 70% ready and the "The implementation is believed to be functionally complete".

[12] The link at: http://downloads.verapdf.org/rel/verapdf-installer.zip did not work at (2015-11-30, 2015-12-01)

All PDF/A files validated OK. Given report was good, both in HTML and XML. Regular PDF files got validation errors, as expected. User interface is nice and easy to understand, even thought it looks like regular java tool. I would like to have more nicely designed UI for the future but this is not mandatory.

3. The Conformance Checker

Interface issues: I have only been able to find the GUI-version of the 06.46 software, whereas the Final report outline a number of other types of interfaces (Command Line Interface, Application Programming Interface, API, Representational State Transfer Interface, and of course Web Browser which is planned for a later release). Explanation provided by Vera PDF in Proto-typing Phase 1 Final Report regarding the CLI and API: “We´ve not been happy with the implementation”.

Setup and Usage: The GUI Interface is easy to download, but is rather rudimentary and robust in its overall features and design, which may have some negative effects with regard to usage and usability.

+1 the installation process. Looks and reacts à la Java.... But it works great in all other respects.
+1 for portable installation.
+1 small size
+1 CLI finally released (in the dev version)
-1 for GUI problem (see below)
+1 documentation of PDF/A in the result report (It seems to be a work in progress! Some assertions have it)
<ruleId specification="ISO_19005_1" clause="6.5.3" testNumber="4"/><message>For all annotation dictionaries containing an AP key, the appearance dictionary that it defines as its value shall contain only the N key. If an annotation dictionary’s Subtype key has a value of Widget and its FT key has a value of Btn, the value of the N key shall be an appearance subdictionary; otherwise the value of the N key shall be an appearance stream.</message>…)

Neither the public release 0.6.46 nor the development release 0.7.45 GUI is working on Ubuntu 14.14 Ubuntu x64 with minimal install, with a basic windows manager (openjdk-7-jre).

Version 0.7.45 however has a CLI. Initializing the CLI:

```
log4j:ERROR Could not find value for key log4j.appender.file
log4j:ERROR Could not instantiate appender named "file".
Version: 0.7.45
Built: Tue Dec 01 23:48:53 UTC 2015
Developed and released by the veraPDF Consortium.
Funded by the PREFORMA project.
Released under the GNU General Public License v3.


Usage: veraPDF [options] FILES
Options:
```

```
-x, --extract
Extract and report PDF features.
Default: false
-f, --flavour
Choose built in Validation Profile flavour, e.g. 1b. Alternatively supply
0 to turn off PDF/A validation.
Default: 1b
Possible Values: [0, 1a, 1b, 2a, 2b, 2u, 3a, 3b, 3u]
--format
Choose output format:
Default: xml
Possible Values: [xml, mrr, html]
-h, --help
Shows this message and exits.
Default: false
-l, --list
List built in Validation Profiles.
Default: false
-p, --profile
Load a Validation Profile from given path and exit if loading fails. This
overrides any choice or default implied by the -f / --flavour option.
--success, --passed
Logs successful validation checks.
Default: false
--version
Version information.
Default: false
```

“Released under the GNU General Public License v3.” should be “Released under the GNU General Public License v3 or later and Mozilla Public License v2 or later.”
What is the format “mrr”?
ODT refers to a file converted from ODT to PDF/A using LibreOffice 4.0.
DOC refers to a file converted from DOC to PDF/A using Microsoft Office 2003 with Adobe Acrobat Professorial X.
DOCX refers to a file converted from DOCX to PDF/A using Microsoft Office 2010.

Feature report

PURPOSE extract as much contextual and structural metadata as possible without attempting to determine conformance to any specific version or level of PDF/A.

OUTCOME expected contextual metadata reported; some structural metadata reported as outline and xobjects, but paragraphs are noticeable missing in the result for ODT and DOCX, but are reported in the result for DOC.

CONCLUSION some issues given the random limited test, but great potential; is being developed as outlined in the functional/technical report.

/verapdf -x -f 0 --format html|xml|mrr ./testfile/<file>
COMMENT *--format* Mrr defaults to XML? It was possible to switch to HTML. **OUTPUT ERRORS** log4j:ERROR Could not find value for key log4j.appender.file log4j:ERROR Could not instantiate appender named "file". **Errors when running pdf converted from ODT** *(These errors did not show up when testing DOC or DOCX)*
./verapdf -x -f 0 --format xml "./testfile/metadata.LO4.0.internal.pdf ./verapdf -x -f 0 --format xml "./testfile/hierarkisk struktur.LO4.0.internal.pdf"
252 errors regarding 00:38:15,864 WARN main PDTrueTypeFont:codeToGID:516 - Can't map code 0 in font BAAAAA+Arial-BoldMT **Followed by a couple of hunded more similar errors for various variations of TimeNewRoman:** 00:38:16,016 WARN main PDTrueTypeFont:codeToGID:516 - Can't map code 252 in font CAAAAA+**TimesNewRomanPS-BoldMT** 00:38:16,023 WARN main PDTrueTypeFont:codeToGID:516 - Can't map code 28 in font DAAAAA+**TimesNewRomanPS-ItalicMT** 00:38:17,410 WARN main PDTrueTypeFont:codeToGID:516 - Can't map code 201 in font EAAAAA+**TimesNewRomanPSMT**

PDF/A1-b Release Candidate

PURPOSE testing the basic conformance level for PDF/A-1 released on October 31st 2015.

OUTCOME 1 fatal error, 1 minor error (warnings), and the rest were evaluated as expected.

CONCLUSION lack of test files expected to conform to PDF/A-1b where the content is not raster graphics. Uncertain how many of the files will be considered valid in another validator. A more comprehensive test is needed. Furthermore, given the small set of test files, although created from a variety of source files, a fatal error was surprising.

/verapdf -x -f 1b ./testfile/<file>				
Random selected files				
i	**file**	**Source**	**exp**	**out**
1	dokument.lo4.odg.jpg.arp10.ocr-	Adobe Acrobat Pro X	valid	valid

	exact.a1b.pdf	(PDF/A-1b file containing a jpg image.)		
2	radio.a1b.pdf	Adobe Acrobat Pro X (PDF/A-1b file containing a radio button form.)	N/A	fail
3	ep-rå-förordning-910-2014_20140828_eidas.download.pdf	Generated by an unknown PDF creator in Mozilla Firefox	fail	fail
4	PDFA.PDFC.1b.pdf	PDFCreator Version 1.7.2 GPL Ghostscript 9.10; modified using iTextSharp 5.2.1 (c) 1T3XT BVBA	fail	fail
5	RA-IT UPPHANDLING PC 2015 version 1.6.pdf	Generated by CutePDF Writer PScript5.dll Version 5.2.2 GPL Ghostscript 9.16	N/A	fail
6	FH141020151_1.pdf	Unknown source	fail	fail
7	pv_2011_1converted.AR10.a1b.pdf	PDFTron ERROR no output except errors	N/A	Error
8	__TMP__.pdf	pdfescape.com	N/A	valid

OUTPUT ERRORS
log4j:ERROR Could not find value for key log4j.appender.file
log4j:ERROR Could not instantiate appender named "file".

File 3 raised the following warnings
10:18:17,049 WARN main PBoxPDStructElem:getChildrenFromArray:67 - Children type of Structure Tree Root or Structure Element in array must be 'COSDictionary' but got: COSArray
10:18:17,051 WARN main PBoxPDStructElem:getChildrenFromArray:67 - Children type of Structure Tree Root or Structure Element in array must be 'COSDictionary' but got: COSArray
10:18:17,051 WARN main PBoxPDStructElem:getChildrenFromArray:67 - Children type of Structure Tree Root or Structure Element in array must be 'COSDictionary' but got: COSArray
10:18:17,051 WARN main PBoxPDStructElem:getChildrenFromArray:67 - Children type of Structure Tree Root or Structure Element in array must be 'COSDictionary' but got: COSArray
10:18:17,051 WARN main PBoxPDStructElem:getChildrenFromArray:67 - Children type of Structure Tree Root or Structure Element in array must be 'COSDictionary' but got: COSArray
File 3 raised numerous exceptions like
01:59:08,379 ERROR main PBoxPDGroup:getColorSpace:57 - Problems with color space obtaining on group. Expected a name or array but got: null
java.io.IOException: Expected a name or array but got: null

at org.apache.pdfbox.pdmodel.graphics.color.PDColorSpace.create(PDColorSpace.java:189)
at org.apache.pdfbox.pdmodel.graphics.color.PDColorSpace.create(PDColorSpace.java:63)
at org.apache.pdfbox.pdmodel.graphics.color.PDColorSpace.create(PDColorSpace.java:48)
at org.apache.pdfbox.pdmodel.graphics.form.PDGroup.getColorSpace(PDGroup.java:73)
at org.verapdf.model.impl.pb.pd.PBoxPDGroup.getColorSpace(PBoxPDGroup.java:46)
at org.verapdf.model.impl.pb.pd.PBoxPDGroup.getLinkedObjects(PBoxPDGroup.java:39)
at org.verapdf.pdfa.validators.BaseValidator.addAllLinkedObjects(BaseValidator.java:194)
at org.verapdf.pdfa.validators.BaseValidator.checkNext(BaseValidator.java:136)
at org.verapdf.pdfa.validators.BaseValidator.validate(BaseValidator.java:87)
at org.verapdf.pdfa.validators.BaseValidator.validate(BaseValidator.java:371)
at org.verapdf.cli.VeraPdfCliProcessor.processStream(VeraPdfCliProcessor.java:101)
at org.verapdf.cli.VeraPdfCliProcessor.processPath(VeraPdfCliProcessor.java:85)
at org.verapdf.cli.VeraPdfCliProcessor.processPaths(VeraPdfCliProcessor.java:72)
at org.verapdf.cli.VeraPdfCli.main(VeraPdfCli.java:66)
File 7 raised numerous exceptions like
02:25:12,333 ERROR main PBoxPDMetadata:getXMPPackage:80 - Problems with parsing metadata. No type defined for {http://ns.adobe.com/xap/1.0/mm/}OriginalDocumentID
org.apache.xmpbox.xml.XmpParsingException: No type defined for {http://ns.adobe.com/xap/1.0/mm/}OriginalDocumentID
at org.apache.xmpbox.xml.DomXmpParser.createProperty(DomXmpParser.java:303)
at org.apache.xmpbox.xml.DomXmpParser.parseChildrenAsProperties(DomXmpParser.java:285)
at org.apache.xmpbox.xml.DomXmpParser.parseDescriptionRoot(DomXmpParser.java:209)
at org.apache.xmpbox.xml.DomXmpParser.parse(DomXmpParser.java:173)
at org.verapdf.model.impl.pb.pd.PBoxPDMetadata.getXMPPackage(PBoxPDMetadata.java:74)
at org.verapdf.model.impl.pb.pd.PBoxPDMetadata.getLinkedObjects(PBoxPDMetadata.java:59)
at org.verapdf.pdfa.validators.BaseValidator.addAllLinkedObjects(BaseValidator.java:194)
at org.verapdf.pdfa.validators.BaseValidator.checkNext(BaseValidator.java:136)
at org.verapdf.pdfa.validators.BaseValidator.validate(BaseValidator.java:87)
at org.verapdf.pdfa.validators.BaseValidator.validate(BaseValidator.java:371)
at org.verapdf.cli.VeraPdfCliProcessor.processStream(VeraPdfCliProcessor.java:101)
at org.verapdf.cli.VeraPdfCliProcessor.processPath(VeraPdfCliProcessor.java:85)
at org.verapdf.cli.VeraPdfCliProcessor.processPaths(VeraPdfCliProcessor.java:72)
at org.verapdf.cli.VeraPdfCli.main(VeraPdfCli.java:66)

Current Text at top:
‘Please specify input PDF, Validation Profile and press ‘Validate’’.

This wording could be more user friendly, suggest:
‘Please select the PDF file for checking’: Locate this text above that section.

Please select the required validation profile. Locate this text above that section.

Then press Validate’.

Choose Profile section:

It would be more user friendly if there was a drop down arrow here where a user could choose the validation profile from a list of profile options rather than having to having to scroll through various folders.

Choose PDF:
I tried a number of Vera's own test file and the system worked as it should either giving me the 'PDF file is compliant/not with validation profile requirements as appropriate.

I tried some of our PDF files and got the message 'Execution exception in processing' each time.

Generate Reports:

- The reports open and save correctly in both xml and html version.
- As a non technical user, I didn't find the detailed information provided on the failed checks very useful, examples below. Can the terminology be more user friendly to assist non technical users in addressing the problems identified in the failed checks.

Rule	Status
ID: 6-1-2-t01	
The file header shall begin at byte zero and shall consist of "%PDF-1.n" followed by a single EOL marker, where 'n' is a single digit number between 0 (30h) and 7 (37h)	Failed
1 occurrences	
root	
ID: 6-1-3-t01	
The file trailer dictionary shall contain the ID keyword whose value shall be File Identifiers as defined in ISO 32000-1:2008, 14.4	Failed
1 occurrences	

Settings Panel
These additional settings are useful.

Installation and setup are easy.
There are errors in processing noncompliant files (execution exception in processing, error in creating features collection).
Validation reports are without identification of the file validated.
No batch processing: manual selecting of every single file for checking could affect the acceptance of the tools very seriously.
Information available via menu "About" doesn't say much about the Conformance Checker.

It is good that you can select version specific profiles or choose profiles for specific PDF sections. I kind of thought that it also automatically selects correct PDF/A version for me but it will not do that. I need to know what version PDF/A file I have. Problem is that sometime you do not know specific version. Of course you can always use some extra tool for identification but it

would be nice, if there is an option, where validator chooses correct profile for me, based on input file PDF metadata information.

4. Final report

At Github Vera Pdf, some release notes are published. The notes outline how the software development has progressed thus far. First, features of each version are described and pointed too. Then, the infrastructure is brought up. Third some development new regarding test corpus is listed and brought to the forefront. The version 0.6 is characterized by stable implementation of the PDF model for PDF/A-1b.

The vera consortium seems to have addressed the comments made by PREFORMA in a constructive fashion, according to the document “Reponse to 2015/09 Examination Observations”. The theme of the current evaluation of the progress of the vera pdf software seems to highlight how the consortium has responded to the D4.3. The discussion has been summed up in a memo by Benjamin “Scope of the discussion” which outlines some of the outstanding issues. The main issue highlighted points to the much needed commitment by suppliers to reach a particular goal, “(that) a user can download all the necessary dependencies to a USB or burn it on a CD, and use it to build and run the Conformance Checker on the minimal installed OS with no Internet connection”.

Overall the functionality in the product which they developed so far and described, works as intended. It works pretty much as described in previous documentations, minus the PDF part, which I thought is also included. Basic functionality is there but there is also room for the improvements.

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

The question of how frequent releases of stable versions are supposed to be has been discussed in the project. In the D4.3 a request is made that stable versions of software should be provided on a monthly basis. In the roadmap provided by Vera, this request has been re-interpreted into a series of five milestones, echoing the main delivery points of the projects.[13] If this is acceptable, I would suggest that each new release is launched for instance by a webinar, where new features are explained and so on.

A manual is now provided on-line: “Vera PDF Desktop Graphical User Interface. Quick Start Guide. Version 0.6 30/10/2015”. The Guide is supposed to explain the basic concepts of vera PDF performance. A glossary is included, which is useful. The main controls of the graphic user interface are explained in the manual, which comes in handy. Once a validation of files has been made, the user can choose to either view or save a human readable report or a machine

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

readable report. I am not sure of the pedagogic value of the manual however - I downloaded the software first, read the manual afterwards (!), but the one seems to mirror the other.

According to the vera pdf github page, there is currently lots of development work carried out in various projects[14] - a sign that there is a strong drive towards progress within the consortium. Under each project at the git hub page, you can follow when the latest updates were done as well as get information on what the work set out to do. This activity mirrors at least in part the request made in the D4.3 of "frequent releases".[15]

The issue of interoperability has been brought up, and some conclusions have been reached by discussions that Benjamin has chaired. Issues brought up have included how build/compile and runtime dependencies for conformance checkers for the various operative systems – an issue which is highlighted in the D4.3 where a request is made that "source code is available for download in a single file for several different deployment platforms" (cf ch.2.4 in the D4.3). Some modifications of this of request have been agreed by PREFORMA in the discussions following the Intermediate report.

Testing

The files used for testing the VeraPDF comes from a set created by the unit Development and eGovernment within Riksarkivet for the purposes of evaluating the PDF/A-1a support in programs that create PDF/A.
The programs, or the source programs, were: LibreOffice 4.0, the Adobe Acrobat Professional X (10.1.7) extension for Microsoft Office 2003, and Microsoft Office 2010.
The file formats, or source files or formats, were ODT, DOC and DOCX.
Each file in the set contains certain functionalities or group of functionalities created within the source program, such as document metadata, annotations, embedding of resources (e.g. images, movie, sound), logical structure (e.g. styles, language settings for document and specific text segments, description for images), and so on. For each set file there were three version converted to PDF/A: PDF/A using LibreOffice; DOC using Acrobat X, DOCX using Microsoft Office 2010.
The set contains a total of 61 PDF -files; 17 doc; 15 docx, 17 odt. (Some source files were converted using different settings.)

The release was tested using veraPDF corpus, files from Isartor test suite, local PDF and PDF/A-1 files, and some files generated, for this testing, by an e-invoice service producing PDF/A-3 files (ZUGFeRD format).
VeraPDF corpus and Isartor test suite contain files for testing conformance to PDF/A-1b (PDF/A-2b section of the VeraPDF corpus contain one file only for file header testing).
PDF/A-3 files are generated for testing the PDFA-3b profile in the release.

[14] https://github.com/veraPDF

[15] PREFORMA Deliverable D4.3, p.5.

Dissemination and community building

The Vera PDF website forms the central hub of the community building taking place[16]. You can join the community for instance by subscribing to an email list. Two particular communities are pointed out as being of pivotal importance to the advancement of the vera pdf software: a) the memory institutions which are invited to provide amongst other things policy requirements, in order for vera pdf to formulate policy profiles to be used by the software in the near future (one of the key components of the PREFORMA software has to do with checking for compliance with individual policy requirements/statements by memory institutions); b) commercial software developers who produces software that create, process and view PDF-files. The Prototyping Phase 1 Final Report outlines several other ways in which vera pdf is communicating and disseminating information (twitter account, linked in, presentations, conferences and so on). Much help in the furtherance of the software is provided by the networks associated with the PDF/A Association and the Digital Preservation Coalition.

In the Prototyping Phase 1 Final Report, vera pdf states that it has made “substantial progress to win the involvement” of various commercial stakeholders, including Adobe Systems, but does not say in how, in what way, and even less which are the items that may of strategic importance as the work is technically solid enough to enable these stakeholders to create an interest in the development.

The issue of building an open source community is a major strategy in PREFORMA – outlined in the Call Documents of the Tender (ITT, Challenge Brief, Framework Agreement; the early Deliverables (D2.1, D2.2 and D.3.1); and also mentioned in the D4.3 (ch 1, and 2) where successful open source projects are being described as “highly collaborative” and “community based”. Therefore, I would personally like to see examples of flows of input and output not just between the organisations within the vera consortium – but also between vera and commercial as well as non-commercial software developers.

Looks good, maybe more social media channels?

Open Source approach

Please find below feedback from the Skövde partner concerning the “Prototyping Phase 1 – Final Report” from VeraPDF, which is focused on issues related to open source. For convenience, when we below write “the report from VeraPDF” we refer to the “Prototyping Phase 1 – Final Report” from VeraPDF. In light of our previous feedback (concerning the “Prototyping Phase 1 – Intermediate Report” from VeraPDF that we received in August 2015) which had references to version 2.0 of deliverable D4.3, let us first clarify that all references to D4.3 in this feedback refers to version 1.0 of this deliverable (i.e. the version of D4.3 that was made available to all suppliers in December 2014 after acceptance in the PREFORMA internal review). As agreed in later discussions between PERFORMA partners and suppliers, even if

[16] http://verapdf.org/community/

there are no substantial differences between versions 1.0 and 2.0 of D4.3 (and formally it was version 2.0 that is the outcome of the PREFORMA review), this feedback refers to requirements expressed in version 1.0 of D4.3. Our feedback is based on PREFORMA requirements which all are necessary pre-requisites for a successful outcome of the PREFORMA project.

Having reviewed the content of the “Prototyping Phase 1 – Final Report” from VeraPDF we make a number of observations. Overall, we find that the supplier has made significant progress. However, we identify 8 issues in this feedback which all must be addressed. For several of these identified issues it is now urgent that the suppliers adhere to the requirements if the project should have a chance to successfully address the PREFORMA R&D challenge.

Issue #1 – Provision of source code

PREFORMA requires that a supplier provides the complete source code (i.e. a single zip-file containing all necessary files for creating a running instance of the source code) under two specific licenses (“MPL v2.0 or later” and “GPL v3 or later”) on the open source portal (i.e. http://www.preforma-project.eu/open-source-portal.html). There must be one zip-file containing ‘the complete source code necessary for each deployment platform’ (as required in D4.3).

Observation: The report from VeraPDF includes direct links (on page 11) to a set of zip-files but all links are broken. Further, all links seem to refer to the development platform (i.e. the GitHub repository) instead of the open source portal as required in D4.3. For example, the filename for one of the links included in the report suggests that the complete source code (under “GPLv3 or later”) is included in the file: https://github.com/veraPDF/veraPDF-library/releases/download/v0.6.26/gpl-source-0.6.zip. Similarly, the filename for another link included in the report suggests that the complete source code (under “MPLv2 or later”) is included in the file: https://github.com/veraPDF/veraPDF-library/releases/download/v0.6.26/mpl-source-0.6.zip.
However, we note (as indicated by the VeraPDF supplier via later email communication with PREFORMA) that zip-files containing source code actually are available on the development platform via the direct links provided by the supplier in email communication (e.g. https://github.com/veraPDF/veraPDF-library/releases/download/v0.6.31/gpl-source-0.6.zip). However, if source code is only made available via the development platform (GitHub) it does not fulfil the PREFORMA requirements for how source code shall be provided.

Action #1 for the supplier: The supplier must provide the complete source code (as required in D4.3) on the open source portal. It is now very urgent that the supplier fulfils this requirement. When will the supplier provide the source code as required in version 1.0 of D4.3?

Comment 1.1: There are several technical and legal reasons for why PREFORMA has expressed this requirement. In short, it is an absolute requirement that the supplier distributes the complete source code (i.e. the supplier must make an ‘external distribution’) before the software can be used and redistributed by anyone who represents other organisations (i.e. before a supplier has made an external distribution of the source code to another organisation, such as the open source portal, the software cannot be used by anyone without significant risks).

Comment 1.2: The minutes from the meeting with PREFORMA partners and representatives for all three suppliers (held 28 October 2015) clearly states that “source code and executables shall be provided on the Open Source portal”. The importance of this requirement has been stressed several times since and it was communicated to all six suppliers in December 2014 via version 1.0 of deliverable D4.3 (and also in the review comments related to the release July 2015). In light of this, we are very puzzled by the fact that the source code (and executables) have not been provided on the Open Source portal as required in version 1.0 of D4.3.

Comment 1.3: For a number of reasons (including control, copyright and patent related reasons) it is standard practice in all European public sector framework agreements that the supplier of open source software must provide the software to the customer (i.e. through an external distribution of the software) in order to minimise risks for the customers. These reasons have also informed PREFORMA requirements concerning these aspects and it is therefore critical that the supplier provides the software on the open source portal (since the portal is controlled by a different organisation, in this case the PREFORMA partner Promoter, which is an organisation that is external to the supplier).

Comment 1.4: In light of that source code has not yet been provided as required (i.e. the supplier has not yet provided the source code to the open source portal as required in version 1.0 of D4.3) would like to clarify a potential misconception raised earlier in the document “veraPDF consortium: Response to 2015/09 Examination Observations” which has earlier been provided to the PREFORMA consortium. In this document, it is mentioned: “Distributing binaries, as opposed to distributing source code, can be a legally tricky exercise. Has the PREFORMA consortium considered the legal implications of distribution of the build environment?” Potential legal issues are one important reason for why PREFORMA require that the supplier provides the software (source code, build environment and executables) as required in version 1.0 of D4.3. We note that it can even be impossible to use the software if the supplier lacks all the necessary rights for providing the software as required in version 1.0 of D4.3 and it is consequently not surprising that it is standard practice in public sector procurement to require that the supplier provides the software (i.e. when the supplier provides the source code and binaries it is required that the supplier has all the necessary rights to do so). For the PREFORMA consortium (and any other user of software which has been distributed under “GPLv3 or later” and “MPLv2 or later” from PREFORMA), it is unrealistic if all users must first conduct a legal analysis before using the software. Precisely for these reasons, it is now critical that the supplier provides the code (i.e. an ‘external distribution’ must take place and it is the supplier that must do this). Please note that it must be possible for any individual to distribute (and redistribute in a cascade) the software (source code and executables) under “MPLv2 or later” and “GPLv3 or later”. For PREFORMA it is critical that the supplier has obtained all necessary rights so that any individual (without any restriction) can distribute (and redistribute in a cascade) the software (the complete source code and executables) under “MPLv2 or later” and “GPLv3 or later”. After the supplier has provided the software on the open source portal it will be possible and meaningful to conduct a more comprehensive analysis of these (rather complex) issues.

Comment 1.5: From our initial analysis of source code available on the development platform Github (https://github.com/veraPDF/veraPDF-library/releases/download/v0.6.31/gpl-source-0.6.zip, https://github.com/veraPDF/veraPDF-library/releases/download/v0.6.31/mpl-source-0.6.zip) we have significant concerns for how the software is licensed. From the approach taken by the supplier, it seems that fundamental licensing requirements in PREFORMA will not be fulfilled when software will be provided by the supplier (assuming that code currently available in the zip files on the development platform controlled by the supplier will be provided by the supplier on the open source portal).

Comment 1.6: Related to the requirements for provision of source code, the PREFORMA requirements also state that the supplier shall make an external distribution of "all digital assets" related to "all open source projects" at the end of the project, please see section 1 of version 1.0 of D4.3 for further details.

Issue #2 – Provision of 'road-map' on the development platform

PREFORMA requires that a supplier provides an up-to-date road-map for the different versions of the software. Please see section 2.3 in version 1.0 of D4.3: "There shall be an up-to-date roadmap with detailed milestones for different (development version, stable version, and deployed (LTS) version) releases."

Observation: The report from VeraPDF lacks detail concerning the requirement for an 'up-to-date roadmap' and it is unclear to what extent the supplier has addressed (and plans to address) this requirement.

Action #2 for the supplier: Please clarify how (and if) this requirement has been addressed. Please also clarify when an up-to-date roadmap will be prominently exposed (to potential external contributors) on the development platform.

Comment 2.1: Provision of an up-to-date roadmap for the project which targets external (potential and active) external contributors (i.e. a roadmap which does not address PREFORMA partners) is an important enabler for promotion of external contributions (i.e. the web page http://verapdf.org/roadmap/ is positive, but its content is tightly linked to PREFORMA, whereas it should be targeted at external contributors (and non-PREFORMA individuals).

Issue #3 – Time-based provision of 'stable' releases

PREFORMA requires that the supplier provides, on a monthly basis, releases which have been exposed to a certain level of QA. Please see section 1 in version 1.0 of D4.3: "Stable versions (provided on a monthly basis) have been exposed to a certain level of QA in the development process."

Observation: The report from VeraPDF lacks detail concerning this requirement and it is unclear to us to what extent such 'stable versions' have been exposed to a certain level of QA.

Action #3 for the supplier: Please clarify how (and if) this requirement has been addressed. Please also clarify when this requirement will be fulfilled.

Comment 3.1: Even if the supplier has not provided 'stable versions' on the open source platform as required (please see issue #1 concerning 'provision of software') we are unsure from the content of the report to what extent this requirement has been (and will be) addressed.

Issue #4 – Identical software under both "GPLv3 or later" and "MPLv2 or later"

PREFORMA requires (as expressed in section 2.3 of version 1.0 of D4.3) that 'The "MPLv2 or later" version and the "GPLv3 or later" version of the software that are developed, maintained, and distributed shall always be identical.'

Observation: As the source code of the software has not yet been provided as required (please see issue #1 above) it is unclear to us if this requirement will be fulfilled (when the supplier will adhere to the requirements expressed in issue #1). We note that source code is provided on the development platform.

Action #4 for the supplier: As the source code of the software has not yet been provided as required (in D4.3) it remains to be seen if the supplier needs to take further action concerning this issue.

Comment 4.1: We note that the supplier currently maintains software under several different open source licenses on the development platform (which is fine). However, it remains to be seen precisely which source code will be provided that fulfils the requirements expressed in D4.3 (including those related to issue #1 and issue #4). Only after source code has been provided as required it is possible (and meaningful) to assess issue #1 and issue #4.

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

PREFORMA requires that an executable shall be provided for each platform. As detailed in section 2.5 of version 1.0 of D4.3 executables (and corresponding source code) shall always be provided on the open source platform "for several different platforms (at least for: MS Windows 7, Mac OSX, common Linux distributions including Ubuntu, Fedora, Debian, and Suse). For each platform specific executable there shall always be an up-to-date corresponding source code that can be downloaded as a single file."

Observation: The report from VeraPDF lacks detail (the report does not contain exact links to specific zip files containing executables for each platform) concerning this requirement and it is unclear to us when the supplier will fulfil this requirement.

Action #5 for the supplier: The supplier needs to provide an executable of the source code for each deployment platform as required (in version 1.0 of D4.3) on the open source portal. Please clarify when this requirement will be fulfilled.

Comment 5.1: The minutes from the meeting with PREFORMA partners and representatives for all three suppliers (held 28 October 2015) clearly states that “source code and executables shall be provided on the Open Source portal”. The importance of this requirement has been stressed several times since it was initially communicated to all six suppliers in December 2014 via version 1.0 of deliverable D4.3 (including in the review of the July release). In light of this, we are very puzzled by the fact that the executables (and source code) have not been provided on the Open Source portal as required in version 1.0 of D4.3.

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

PREFORMA requires that the software can be used via standard web browsers. Please see section 2.5 of version 1.0 of D4.3 for details: “First, using any standard web browser any individual, both working for a memory institution and other interested individuals, can use the open source software which is provided for use at the open source project website for checking conformance for a specific file format.” In this requirement, a ‘standard web browser’ refers to “the most recent stable release of each of the following web browsers: Chrome, Firefox, Internet Explorer, and Opera.”

Observation: The report from VeraPDF lacks detail concerning this requirement and it is unclear to us when the supplier will fulfil this requirement.

Action #6 for the supplier: The supplier needs to provide the software for use via a standard web browser as required (in version 1.0 of D4.3). Please clarify when this requirement will be fulfilled.

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

PREFORMA requires that the supplier provides detailed documentation concerning interpretation of the technical specification of each file format used. Please see section 2.1 in version 1.0 of D4.3 for details: “Each open source project is expected to evolve detailed documentation concerning precisely how technical specifications of file formats are being interpreted and implemented in software. Such details is critical feedback for organisations maintaining technical specifications of file formats and from open provision of details concerning precise interpretations of different parts of a technical specification has been (and should be) interpreted will constitute a very valuable
resource for the broader open source and standards communities. With evolving precision in such open publication (via issue trackers and other means in each Open Source project) there will be an ongoing process for scrutiny of interpretations that eventually promotes improved quality of both how technical specifications should be clarified and how technical specifications should be interpreted and implemented in software.”

Observation: The report from VeraPDF lacks detail concerning this requirement and it is unclear to us when the supplier will fulfil this requirement.

Action #7 for the supplier: The supplier needs to address this requirement as required (in version 1.0 of D4.3). Please clarify when this requirement will be fulfilled.

Comment 7.1: This requirement is a novel aspect of the PREFORMA R&D project and this requirement it critical for successfully addressing the R&D challenge in PREFORMA.

Comment 7.2: Please note that this requirement addresses two aspects. First, it addresses complete and consistent interpretation of the technical specification of each file format (as specified). This seeks to contribute to an improved technical specification of each file format (thereby contributing to improved quality in standardisation). Second, it addresses complete and consistent interpretation of the technical specification when implemented in software. This seeks to contribute to an improved congruence between the software implementation of a specific file format and its technical specification (thereby contributing to improved quality in faithful software implementation of file formats).

Issue #8 – Provision of software which can be redistributed in a cascade

PREFORMA requires that the supplier provides all code (i.e. all source code; tool chain for building executables; and executables etc.) under open source licenses (on the open source portal) and that all code can be distributed and redistributed by any individual. Please see section 2.3 in version 1.0 of D4.3 for further details:
'All software developed and maintained in each open source project will be provided under two specific open source licenses (www.opensource.org), namely: both Mozilla Public License "MPL v2.0 or later" and under GNU General Public license 3.0 "GPLv3 or later". This implies that all source code from the open source project that is necessary for creating an executable can be distributed (in a cascade) under these specific licenses to anyone. For example, a user at the memory institution Riksarkivet can download the complete source code for all software maintained in the open source project and redistribute the downloaded complete source code (under these specific licenses) to another organisation wishing to use the software.'

Observation: The report from VeraPDF lacks detail concerning this requirement and it is unclear to us when the supplier will fulfil this requirement. Concerning the provision of the open source-licensed tool chain for building executables it has been agreed (between suppliers and the PREFORMA partners) that provision of this part can be delayed until early 2016 (at the very latest, it shall be delivered before the PREFORMA open source workshop to be held in Stockholm in April 2016). Further, it is critical that the supplier has all the necessary rights for provision of source code and executables under the two specific licenses (under "MPL v2.0 or later" and "GPL v3 or later") and provision of tool chain (under OSI-approved licenses) for building executables (under "MPL v2.0 or later" and "GPL v3 or later") as required in version 1.0 of D4.3.

Action #8 for the supplier: The supplier needs to address this requirement as required (in version 1.0 of D4.3). Please clarify when the source code, tool chain for building executables, and executables on the open source portal will be provided as required in version 1.0 of D4.3.

Comment 8.1: This requirement is a pre-requisite for gaining full control of developed software for memory institutions and as such it is critical for successfully addressing the R&D challenge in PREFORMA. It is not sufficient to only host software on the development platform and other sites which PREFORMA partners do not control. When suppliers provide the software on the open source portal (according to the PREFORMA requirements in version 1.0 of D4.3) it follows that PREFORMA partners will have some control of developed software, and legal risks for any organisation wishing to use software from the PREFORMA project are also reduced.

Comment 8.2: In light of previous discussions, we would like to stress that PREFORMA requirements for provision of software on the open source platform (i.e. when the supplier makes an external distribution of the complete software as required in version 1.0 of D4.3) will fulfil some basic requirements for provisions of software (e.g. minimises legal risks for any user of the software, ensures some control for PREFORMA partners and external contributors, and promotes longevity of developed solutions). However, this should not be confused with efforts undertaken by the supplier for developing long-term sustainable open source communities related to developed software (on GitHub and elsewhere).

Comment 8.3: We note that strict adherence to licensing requirements (i.e. any individual must be able to redistribute the software (the complete source code and executables) under “MPLv2 or later” and “GPLv3 or later”) is a necessary, but not sufficient, pre-requisite for any planned activities related to integration of software. Before the supplier provides software (the complete source code and executables) under “MPLv2 or later” and “GPLv3 or later” on the open source platform as required in version 1.0 of D4.3 all efforts related to integration are potentially wasted. In case the supplier would fail to adhere to these licensing requirements, it would make it impossible to use and integrate developed solutions as planned in PREFORMA.

Comment 8.4: Strict adherence to licensing requirements (i.e. any individual must be able to redistribute the software (the complete source code and executables) under “MPLv2 or later” and “GPLv3 or later”) is a necessary, but not sufficient, pre-requisite for anticipated development of a long-term sustainable business ecosystem related to developed software. Before the supplier provides software (the complete source code and executables) under “MPLv2 or later” and “GPLv3 or later” on the open source platform as required in version 1.0 of D4.3 all efforts from the other two suppliers (and all external contributors) related to the software would potentially be wasted.
After the supplier has provided the software on the open source platform (i.e. which imply that an external distribution of the software has taken place) as required in version 1.0 of D4.3 it is a strong sign to the other two suppliers (and any other individual who is a potential external contributor) that the supplier is convinced that the supplier has all the necessary rights to distribute the software. In case the supplier would fail to adhere to the requirements in version 1.0 of D4.3 it would significantly reduce business opportunities for the two other suppliers (and any other current and potential participant in the broader community related to the software). For this reason, it is critical that the supplier urgently provides the software on the open source platform.

Comment 8.5: In addition to fulfilment of requirements for provision of software on the open source portal (as detailed in version 1.0 of D4.3), the supplier is also expected to provide and

promote PREFORMA software via other channels. For example, appropriate means for promotion of the broader communities related to PREFORMA software include provision of 'Live-CDs' (or via USB:s) containing the complete source code and executables which allow anyone to use and scrutinise developed software without the need for installation.

Comment 8.6: For promotion of the broader business ecosystem and development communities, when executables (under "MPLv2 or later" and "GPLv3 or later") of software are made available (irrespectively of how) it is important to always make the corresponding complete source code available (under "MPLv2 or later" and "GPLv3 or later"). It is essential to always clearly indicate open source licenses used. Plans for future releases (as indicated in an up-to-date roadmap for the project) should also indicate future (planned) availability of executables and source code. For example, the web page provided via a site controlled by the supplier (http://verapdf.org/software/) includes buttons with links to executables for different platforms (i.e. for Windows http://downloads.verapdf.org/rel/verapdf-installer.zip; for Mac http://downloads.verapdf.org/rel/verapdf-installer.zip; and for Linux http://downloads.verapdf.org/rel/verapdf-installer.zip). All these links are broken. Please correct the links. Further, this web page (http://verapdf.org/software/) currently lacks information about licenses. The page should clearly indicate that software is made available under "MPLv2 or later" and "GPLv3 or later". Further, the web page should also include links to the complete corresponding source code (made available under "MPLv2 or later" and "GPLv3 or later") from which the executable has been developed.

Comment from Riksarkivet: we have noted that the veraPDF consortium does provide a roadmap with milestones for different releases[17] - a supply of information which mirrors a request made in the D4.3. There is a great need to handle the issues of project related open source approaches constructively, with an eye to what is actually being developed, while still maintaining adherence to the valuable insights and standard setting ambitions of the D4.3. The way in which Benjamin on behalf of PREFORMA handled the issues is a good example (documented in an e-mail from a meeting 2015-10-26). Carl Wilson, veraPDF, offered to take the lead in developing a practical solution.

Standardisation efforts

In the Prototyping Phase 1 Final Report, the vera pdf consortium comments that few developments have been made since July. A meeting in Basel – which is brought up in the Intermediate as well as the Final Report of the first prototyping phase – were the technical working group of the consortium (TWG) will meet in person in order with the ISO WG to facilitate cooperation is highlighted by vera as having strategic importance. Since the involvement in and capacity to engage directly with the ISO was one of strong points of the bid of the consortium last year, I think it would be of some importance to bring this up further. Sure, since the start of Phase 2,

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

the vera pdf consortium has connected with the relevant ISO working groups, and some outcomes are mentioned in the Final Report, including decisions taken as a results of this particular activity, but I am not sure whether this has been discussed at all at supplier level in the project. Too soon perhaps?

They keep working towards it. It is very slow and long process and they already made some nice efforts and keep doing that in 2016.

Gap analysis and next steps

Development: (+)

Testing: (+ +): A minor issue is that the link from the website did not work, meaning that I had to find the latest version at ***http://downloads.verapdf.org/rel***.

Progress: (+): Steady, with a workable new stable version of the software at hand by late October. If the roadmap of releases – five of them, rather than monthly – is accepted by the project management team, I would like to see that new releases are presented in webinars or something similar to it, in order for new features to be highlighted and explained further.

Open Source Approach: the issues raised by Skövde need to be addressed.

Community Building: (-) more should be done perhaps to facilitate an open source community around the project, since this is outlined as a major strategy of PREFORMA.

Standardisation Efforts (-), perhaps unnecessary remark of mine, since the issues related to this have not been at the project agenda as of late.

They did not provide any new timeline or work plan for future, means that they stick with the current plan that they already provided in previous reports. Development status report is nice and informative that they added end of the report.

Feedback on the final release - Oct 2015 - DPF Manager

In this document we have compiled the comments that we received so far by the members of the PREFORMA Evaluation Committee on the final release of the first prototyping phase (end of October 2015), including those on the text in the Final report.

The comments are normally taken in to the document more or less in the shape they were sent in, e.g. without filtering and editing, and relate both to specific and more general issues. Sometimes they might reflect different opinions but overall we hope that they provide a useful input for the next period.

The aim is that these comments shall serve as a base for further discussions.

1. General comments

EasyInnova has created a platform for the open source project based on the GitHub platform. The different tools GitHub provides for open source projects fills most of the different needs of communication within an open source project. However, the community is not tools and technology but people cooperating to reach common goals. In that respect the activity of the community is very low and it is important that the work to achieve this is increased. One of the main challenges for the project is to have a thriving community so that the software will live on when the Preforma project ends.

The downloaded version (1.2.1) of the software was not working as expected since some basic functions did not work. We therefore decide to check the software repository and build the software on our own. Since this is also part of the requirements we think reporting on this is valuable for both the evaluation as well as the software developers. On Nov. 22nd and updated version (1.2.2) was made available for download. It may have addressed some issues that were raised directly in the repository due to this report.

2. Result from examination

The test files consist of different TIFF files from our own digitization projects. The oldest files were created around the year 2000 and the newest ones in 2015. The files differ in size from 30 MB – 2 GB. The test environment has been a Windows 7 Enterprise SP1 64-bit, 8 GB RAM and the software version tested has been DPF manager version 1.1.1. The tests have been done on individual files using the GUI application.
The software seems to perform quite well and appears to be stable when working with smaller files. However, when validating the 2 GB TIFF file the software crashed with no explanation. The gui still worked but the validation process ended with no result.

3. The Conformance Checker

Error-message in DOS-window when starting.
When testing the Fixer I can't see it works, can't find any new tiff files created.

The User manual does not mention IMPORT and DELETE buttons (still easy to understand when testing).
The new version contains many new good features.

DOCUMENTATION:

- The user documentation need some improvement, not all aspects are covered (see also issue)
- Positive: User and Developer Forum
- The user manual is not complete. While it gives install and run guidance it does not provide e.g. exemplary explanation of what to find inside a report (just how to create one)

SETUP:

- Installer provided (tested for Windows)
- Easy installation and running
- Some basic functionality did not work as expected

BUILD:

- No release tags in github available – release cannot be retrieved in source code which is quite a problem
- Main README.md in github should give some pointers on how to build
- Positive: Website http://dpfmanager.org/community.html provide a build environment and a guide on how to build
- Maven needs to be installed in order to build the system
- Build itself is straight forward with normal developer knowledge (even for a non Java user)
- Dependencies are resolved correctly.

ISSUES:

During the evaluation phase some issues were raised for both usability and stability (i.e. bugs) The evaluator decided to directly report those issue in the github. As a positive remark, the issues have been addressed immediately during the evaluation phase and most of them have already be resolved.
https://github.com/EasyinnovaSL/DPFManager/issues/37
https://github.com/EasyinnovaSL/DPFManager/issues/38
https://github.com/EasyinnovaSL/DPFManager/issues/39
https://github.com/EasyinnovaSL/DPFManager/issues/40
https://github.com/EasyinnovaSL/DPFManager/issues/41
https://github.com/EasyinnovaSL/DPFManager/issues/42
https://github.com/EasyinnovaSL/DPFManager/issues/43
https://github.com/EasyinnovaSL/DPFManager/issues/44
https://github.com/EasyinnovaSL/DPFManager/issues/45
https://github.com/EasyinnovaSL/DPFManager/issues/46
https://github.com/EasyinnovaSL/DPFManager/issues/48
https://github.com/EasyinnovaSL/DPFManager/issues/49
https://github.com/EasyinnovaSL/DPFManager/issues/50

There's a name clash with: http://www.dpfmanager.co.uk/index.html

Tested version is 1.2.1 (both on Mac & Windows)
Installation: easy
Testing environment Windows

- Windows 7 (standard PC)

Testing environment Mac

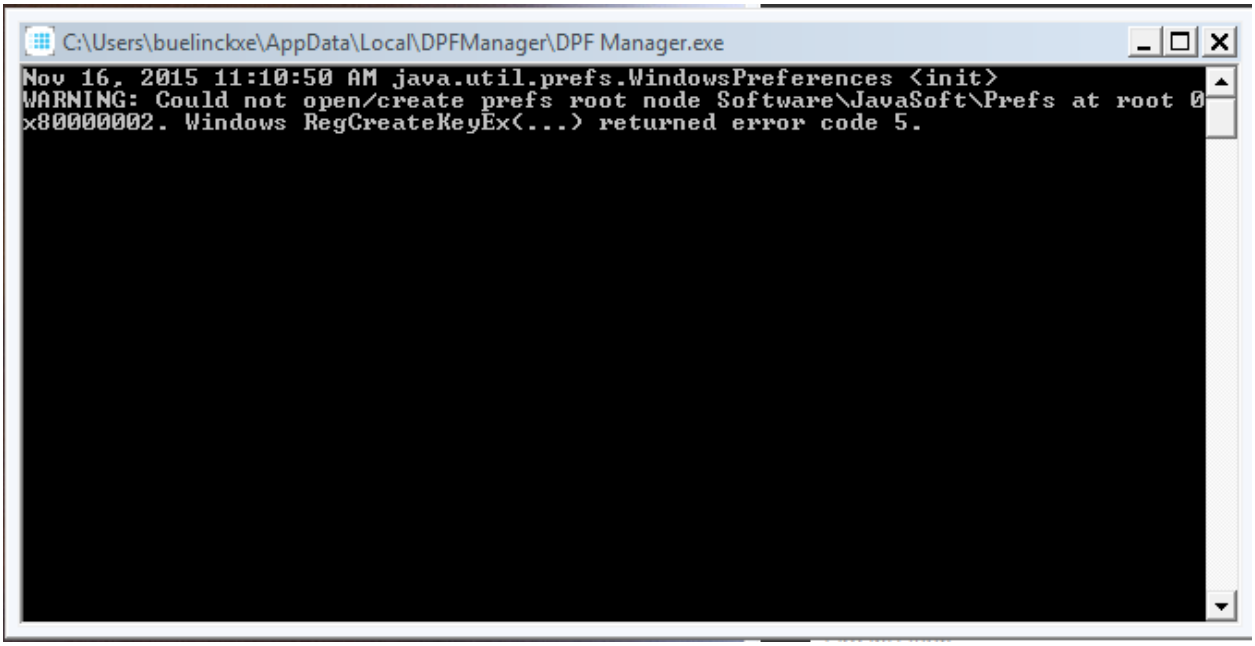
- MacBook Pro (Retina, 15-inch, Late 2013)
- Memory 16 GB 1600 MHz DDR3
- OS X Yosemite 10.10.5

Setup

- OS X installation very simple (as always): just drag-and-drop
- Would have been nice to include sample basic configuration files

Usage

- Apart from the interface issues (see below for really bad design) the graphical UI seems fairly well designed and easy to issue
- "About" has no information on which version of DPF Manager in installed
- Error code opening program (Windows) [see screenshot]

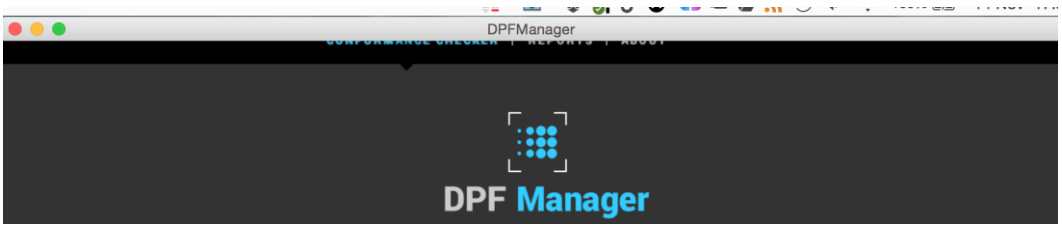


Windows version has "Baseline HTML.dpf" config file.
Macintosh version has no config files
There should be an indication that the program is processing something, now it's just a white screen.

- Making up your own set of criteria to test against needs of course to be expanded but is from a user's view okay

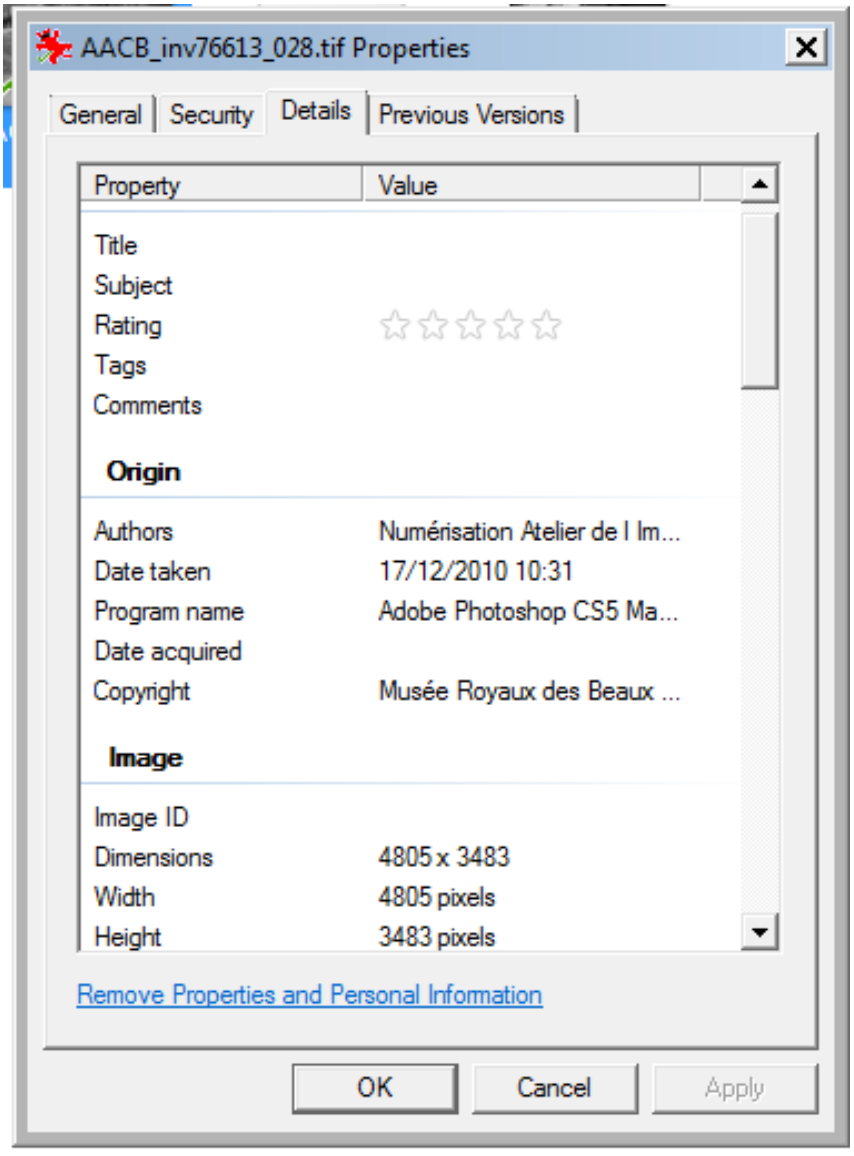
Mac version Interface issues

- **Clickable region is up to 1 cm lower than the indicated place to click, which make sit hard to guess where to click [see screenshot]**



- **Top menu barely (or almost not) visible**
- **Using the keyboard shortcut to hide (cmd-H on Mac) brings up the About-window**
- **Choosing the command to hide from the menu also brings up the About-window**
- **When in final report (HTML- impossible to return to main menu (top menu choices didn't work anymore) (linked to poor interface design on Mac ?)**

Accents problem when showing copyright info: [see screenshots]



Tags

Tag Id	Tag Name	Value
256	ImageWidth	4805
257	ImageLength	3483
258	BitsPerSample	16
259	Compression	1
262	PhotometricInterpretation	1
271	Make	Canon
272	Model	Canon EOS 5D Mark II
274	Orientation	1
277	SamplesPerPixel	1
282	XResolution	2400000/10000
283	YResolution	2400000/10000
296	ResolutionUnit	2
305	Software	Adobe Photoshop CS5 Macintosh
306	DateTime	2010:12:21 19:37:56
315	Artist	Numrisation Atelier de l Imagier
33432	Copyright	Muse Royaux des Beaux Arts de Be lgique - Archives de l'Art contempor ain en Belgique

The UI is fairly easy to use. The idea of giving the user opportunity to indicate/ criteria is good

The following should be considered during further development.

- It should be possible to find information about the version of the software in GUI and/or as a parameter to the CLI. Most helpful when performing evaluation.
  It should be possible to log (or choose to log) the validation process in the GUI as well as in the CLI.
  It should be possible to choose and test more than a single file at a time from the GUI which I think is possible in later versions.
  The software should work with large TIFF files. The process as of today seems to consume quite a lot of RAM during validation and should, if possible, be optimized and require less resources (ie. allocation of RAM during validation).

Environment:
Test was conducted on a Sony-Vaio with Windows 10 pro
Prozessor: Intel i7-4500U @ 1.8 GHz 2.39 GHZ
RAM: 8 GB
A 64-Bit-Version of Windows was used
Free on Harddisk: More than 100 GB

General:

- Cannot change language. Everything is in english
- The dpf window cannot be enlarged (nor resized)
- Help buttons are not working

How the test was run:

Action1

- Download of windows-executable from http://dpfmanager.org/community.html#portables

Action2

- Download of http://dpfmanager.org/Downloads/Manual%20Offline.pdf (Instruction for installation)

Action3

- Reading of Instruction for installation

Observation: Many typos, like “The first step is to install de Java version 8 in the computer computer (if it is not already installed).“

Observation: Instruction says: “1st step is to install Java JDK8 if it is not installed.”

a) How can I find out, if this version of Java is installed? No information given.
b) If not already installed I shall run a provided “.exe”-file. This file is not provided
c) Then I shall install a “maven” folder. This equally is not provided.

Re-reading the first paragraph of instructions I see that the installation requires the “zip-file of the Build”.

d) What is a build?
e) Where is the installation-instruction for the Windows-executable?

Action4

- Unzipped the windows executable

Action5

- Run the windows.exe – file

A desktop icon appears

Action6

- Clicked on desktop icon

A command-window (???) opens. Above it a second window appears with the DPF-Manager icon

Observation: This window is really confusing me. It should not be there. Don’t understand why it is there! From here on I call this the “ugly window” while I call the other the “dpf-window”

Action7

- Closer look at dpf-window

Observation: This window says “conformance checker”. And it says “Files” with an “i” behind.

Neither a click on the “i” nor a mouseover or click on “Files” made anything happen.

Observation: Below this there is a slot which says “Select a file” with a “select”-Button behind.

Action8
- Tested it by clicking on Select
Observation: Seems I can only upload/check one file at a time
Observation: Below the slot/select, there is a line with “Configuration” and again an “i”.
Neither a click on the “i” nor a mouseover or click on “Configuration” made anything happen.

Action9
- Continue the closer look at dpf-window
Observation: Below “confirmation”-line there is field which says “Baseline HTML.dpf”
Observation: Don’t see what this is good for. Explanation needed. I guess this is the set of rules ???

Action10
- Continue the closer look at dpf-window
Observation: Next line has “Import”-Button: What can I import? Files? Rules? Reports? No explanation given
Observation: Same line has “New”-Button: What does this mean? No explanation given
Observation: Same line has “Delete”-Button: Do I delete my files? What is meant with this? No explanation given
At the button there is a “Check files”-Button. How can I check file**s** when I only can upload one file at a time?

Action11
- Tried to get a fullsize DPF-Manager-Window by clicking the square in head-line
Observation: Did not work! Window jumped to the top-left corner of the screen but no resizing

Action12
- Moved mouse around
Obervation: Mouse is not moving smooth anymore!

Action13
- Selected a jpg-File with a click on the select-button:
Observation: The ugly window showed a lot of unintelligible messages.
After this I could not select another file (none at all) in the DPF-Manager Conformance Checker window.

Action14
- Clicked on “Reports” (top-line)
Got no report.

Action15
- Clicked on “About”
Observation: Instead of help or a contact – I learned the “DPF Manager is the most advanced TIFF conformance checker for digital preservation”

Action16
- Checked the ugly window again
One of the messages in the ugly window says: javafx.fxml.FXMLLoader$ValueElemet process Value.
Warning: Loading FXML document with javaFX API of version 8.0.40 by JavaFX runtime of version 8.0.25
Does this mean I have to downgrade my java ?
Why does the executable not check if a working version of java is installed?

Test the "Build"

Action1: Downloaded the windows-build-Version (quite big)
Action2: Unzipped the folder
Action3: Found "How to build"-File and read it
Action4: Found JDK-8u60. Installed it
Action5: Copied – according to instructions – Maven-Folder to the windows-desktop
Action6: Changed – according to instructions the line "localRepository" in maven/settings.xml
Action7: Found in the "How to build"-File the line: "Add "C:/Path/to/maven/bin" to the PATH environment variable. **How to do that?**
Action8: Went on with reading the "How to build"-file. I should type –cd "C:/Path/to/DPFManager" where?
Action9: Should type "-mvn install"? Where?
Action10: Windows 10 Pro is new for me. After a while I found the command line window of windows and run it. The paths I typed into the window were autocompleted by windows with backslashes instead of slashes. Anyway. Stopped action because I got the impression I am going to lose control over the process AND especially because I did not "Add … to the PATH environment variable".

[After installing the JDK-8u60 from the "build"-zip the already installed windows-executable is tested again]

1) Loaded a .jpg-File using the "Select a file"-Slot.
2) Selected "Baseline HTML.dpf"
3) Report appeared, called: "Multifile report" (I uploaded only one file)
4) Report says: "0 files processed" (it does not say why the file was not "processed")
5) Copied an image-folder with some jpg-Files in it (no tiff-file) to the "Select a file" [written in singular] slot. There was no information that I can copy whole folder into the slot – just tried it. Folder contained only jpeg-Files. The ugly windows gives messages in English: File does not exist or is not a TIFF. The Report still says: 0 files processed. This I guess is wrong because the files were checked for being TIFF which means they were processed!
6) Copied an image-folder with 3 jpg and 1 Tiff-File. Report says: "1 files" [plural] "processed"
   1 conforming to policy checker (I did not set up a policy checker)
   1 conforming to Baseline Profile

7) Took the same folder and reformatted (with Adobe Photoshop CS2) one of the ".jpg"-Files into ".TIF". Repeated last operation (folder to slot). Report says: 1 files (plural) processed … (rest as before). This time an image appeared at the bottom of the report.
8) Took the same folder and renamed the tif-File (the original) into jpg: Report says: 0 files processed, i.e. the jpg was not detected as being a tif !!!
9) Took the same folder and renamed the tif-File into ".TIFF" (two ff at the end): Report says: 1 files processed
10) Took the same folder and copied the tif-File to a second version with lzw-compression: Report says 2 files processed. 2 conform to policy checker and 2 conform to Baseline Profile. Below the report I find both images.
11) By accident I clicked on the second of the files (the lzw-Tiff) and got a "single file report" (the software did not inform me about this possibility). In this report I find some information, but the compression is not mentioned in words but only as "5" (which corresponds to LZW but I had to find that out myself)
12) Took the same folder. Made a copy of the LZW-File saving it with Packbits compression (so it is compressed two times). Report says: Passed. (all checks, i.e. policy, baseline profile). Compression is given as 32773. So the first compression (LZW below Packbits) is not detected or not reported.
13) Took the same folder. Took one of the TIF-Images (no compression) and resized it to 2222 dpi with 28.000x18.000px resulting in a file with 1.94 Gigabyte. The whole folder now contains 6 Tif-Files plus 6 jpg-Files. "Processing" went on and on – the ugly window said "processing file" (with the name of the big file behind it) and it had a curser blinking. Above the dpf-window there was a layer simply saying: "Processing". After 20 Minutes I stopped the process because I got no Report.
14) Took the same folder – deleted the last file. Took one of the TIF-Images (no compression) and resized it to 20 dpi with 20000x29999px resulting in a file with 2.23 Gigabyte. The "processing"-layer stayed for longer than 5 minutes. Interrupted the test (Folder contained again 6 tif and 6 jpg-Files). Got no report from the programme.
15) Took same folder – deleted last file. Took one the TIF-Images (no compression) and resized it to 20dpi with 20x20px, resulting in a file with 24,5 KB. Report says, this file is ok. Get a very ugly looking (enlarged) image in the "Multifile Report".
16) Deleted last file. Took a jpg image and put it into a tif-image (layer). Resized everything to 2000x1333px with 72dpi. Saved the whole file with layers (file-size: 17.3 MB). Put the whole folder (6+6 images) to the select-a-file-slot. No report after 10 min of waiting. Test interrupted
17) Took only the multilayered file (i.e. not the folder) to the test: 10 minutes running without reporting! Test interrupted
18) Deleted the file multilayer-file
19) Next Level. Clicked on "New" and answered all questions with yes (checked all checkboxes). Took one of the tif-files (previously reported as ok) and run the system: After 20min it was still processing. Test interrupted.

4. Final report

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

The development of the software is well under way and seems to be on track. Features have been implemented in a good way and in during an adequate time span. However, the release process lacks a proper changelog where you can follow the development between different releases. It would also be good to have the different development stories (backlog) published in a roadmap through GitHub. As far as we can see the only available information today is reported bugs.

Testing

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

The collection of test files (the suppliers collection) is sufficient for the time being. The goal for the improvement of the image collection will probably be an important collection for many software developers and fit well into the goals for the Preforma project.

Dissemination and community building

Several instruments have been set-up in order to support community building. These are:

- A blog (blog entry frequency should be increased)
- Tweets (good frequency, 198 Tweets, 81 follower on Dec.12th 2015)
- Developer area (forum, IRC, docs and github tools)
- Contribution guidelines prepared (not really there)

Overall, the measures taken are good. The most important thing is to keep them updated and synced with the progress within and especially **beyond** the PREFORMA project. Nothing is worse than an outdated wesite/blog/doc/HowTo!

As an instrument of gaining more attraction a live webinar could be done.

Open Source approach

Versioning and better in-repository documentation needs to be provided.
A detailed contributing guide needs to be set-up. It's mentioned "See the contributing guide for more information and for an introduction into the development workflow. […]" but there is no link.

Please find below feedback from the Skövde partner concerning the "Prototyping Phase 1 – Final Report" from Easy Innova, which is focused on issues related to open source. For convenience, when we below write "the report from Easy Innova" we refer to the "Prototyping Phase 1 – Final Report" from Easy Innova. In light of our previous feedback (concerning the "Prototyping Phase 1 – Intermediate Report" from Easy Innova that we received in August 2015) which had references to version 2.0 of deliverable D4.3, let us first clarify that all references to D4.3 in this feedback refers to version 1.0 of this deliverable (i.e. the version of D4.3 that was made available to all suppliers in December 2014 after acceptance in the

PREFORMA internal review). As agreed in later discussions between PERFORMA partners and suppliers, even if there are no substantial differences between versions 1.0 and 2.0 of D4.3 (and formally it was version 2.0 that is the outcome of the PREFORMA review), this feedback refers to requirements expressed in version 1.0 of D4.3. Our feedback is based on PREFORMA requirements which all are necessary pre-requisites for a successful outcome of the PREFORMA project.

Having reviewed the content of the “Prototyping Phase 1 – Final Report” from Easy Innova we make a number of observations. Overall, we find that the supplier has made significant progress. However, we identify 8 issues in this feedback which all must be addressed. For several of these identified issues it is now urgent that the suppliers adhere to the requirements if the project should have a chance to successfully address the PREFORMA R&D challenge.

Issue #1 – Provision of source code

PREFORMA requires that a supplier provides the complete source code (i.e. a single zip-file containing all necessary files for creating a running instance of the source code) under two specific licenses (“MPL v2.0 or later” and “GPL v3 or later”) on the open source portal (i.e. http://www.preforma-project.eu/open-source-portal.html). There must be one zip-file containing ‘the complete source code necessary for each deployment platform’ (as required in D4.3).

Observation: The report from Easy Innova lacks direct links to a set of zip-files. Further, all links from the report refer to the development platform (i.e. the GitHub repository) instead of the open source portal as required in D4.3. If source code is only made available via the development platform (GitHub) it does not fulfil the PREFORMA requirements for how source code shall be provided.

Action #1 for the supplier: The supplier must provide the complete source code (as required in D4.3) on the open source portal. It is now very urgent that the supplier fulfils this requirement. When will the supplier provide the source code as required in version 1.0 of D4.3?

Comment 1.1: There are several technical and legal reasons for why PREFORMA has expressed this requirement. In short, it is an absolute requirement that the supplier distributes the complete source code (i.e. the supplier must make an ‘external distribution’) before the software can be used and redistributed by anyone which represents other organisations (i.e. before a supplier has made an external distribution of the source code to another organisation, such as the open source portal, the software cannot be used by anyone without significant risks).

Comment 1.2: The minutes from the meeting with PREFORMA partners and representatives for all three suppliers (held 28 October 2015) clearly states that “source code and executables shall be provided on the Open Source portal”. The importance of this requirement has been stressed several times since and it was communicated to all six suppliers in December 2014 via version 1.0 of deliverable D4.3 (and in the review comments related to the release July 2015). In light of this, we are very puzzled by the fact that the source code (and executables) have not been provided on the Open Source portal as required in version 1.0 of D4.3.

Comment 1.3: For a number of reasons (including control, copyright and patent related reasons) it is standard practice in all European public sector framework agreements that the supplier of open source software must provide the software to the customer (i.e. through an external distribution of the software) in order to minimise risks for the customers. These reasons have also informed PREFORMA requirements concerning these aspects and it is therefore critical that the supplier provides the software on the open source portal (since the portal is controlled by a different organisation, in this case the PREFORMA partner Promoter, which is an organisation that is external to the supplier).

Comment 1.4: From our initial analysis of source code available on the development platform and the site controlled by the supplier (http://dpfmanager.org/community.html#source_code) we have significant concerns for how the software to be provided on the open source portal may be licensed. From the approach taken by the supplier, it seems that fundamental licensing requirements in PREFORMA will not be fulfilled when software will be provided by the supplier (assuming that code currently available in the zip files (e.g. http://www.easyinnova.com/dpfmanager/Downloads/Releases/Release-1.2/Windows/src01-2015-10-28.zip) controlled by the supplier will be provided by the supplier on the open source portal). Potential legal issues are one important reason for why PREFORMA require that the supplier provides the software (source code, build environment and executables) as required in version 1.0 of D4.3. We note that it can even be impossible to use the software if the supplier lacks all the necessary rights for providing the software as required in version 1.0 of D4.3 and it is consequently not surprising that it is standard practice in public sector procurement to require that the supplier provides the software (i.e. when the supplier provides the source code and binaries it is required that the supplier has all the necessary rights to do so). For the PREFORMA consortium (and any other user of software which has been distributed under “GPLv3 or later” and “MPLv2 or later” from PREFORMA), it is unrealistic if all users must first conduct a legal analysis before using the software. Precisely for these reasons, it is now critical that the supplier provides the code (i.e. an ‘external distribution’ must take place and it is the supplier that must do this). Please note that it must be possible for any individual to distribute (and redistribute in a cascade) the software (source code and executables) under “MPLv2 or later” and “GPLv3 or later”. For PREFORMA it is critical that the supplier has obtained all necessary rights so that any individual (without any restriction) can distribute (and redistribute in a cascade) the software (the complete source code and executables) under “MPLv2 or later” and “GPLv3 or later”. After the supplier has provided the software on the open source portal it will be possible and meaningful to conduct a more comprehensive analysis of these (rather complex) issues.

Comment 1.5: Related to the requirements for provision of source code, the PREFORMA requirements also state that the supplier shall make an external distribution of “all digital assets” related to “all open source projects” at the end of the project, please see section 1 of version 1.0 of D4.3 for further details.

Issue #2 – Provision of ‘road-map’ on the development platform

PREFORMA requires that a supplier provides an up-to-date road-map for the different versions of the software. Please see section 2.3 in version 1.0 of D4.3: “There shall be an up-to-date roadmap with detailed milestones for different (development version, stable version, and deployed (LTS) version) releases.”

Observation: The report from Easy Innova lacks detail concerning the requirement for an ‘up-to-date roadmap’ and it is unclear to what extent the supplier has addressed (and plans to address) this requirement.

Action #2 for the supplier: Please clarify how (and if) this requirement has been addressed. Please also clarify when an up-to-date roadmap will be prominently exposed (to potential external contributors) on the development platform.

Comment 2.1: Provision of an up-to-date roadmap for the project which targets external (potential and active) external contributors (i.e. a roadmap which does not address PREFORMA partners) is an important enabler for promotion of external contributions.

Issue #3 – Time-based provision of ‘stable’ releases

PREFORMA requires that the supplier provides, on a monthly basis, releases which have been exposed to a certain level of QA. Please see section 1 in version 1.0 of D4.3: “Stable versions (provided on a monthly basis) have been exposed to a certain level of QA in the development process.”

Observation: The report from Easy Innova lacks detail concerning this requirement and it is unclear to us to what extent such ‘stable versions’ have been exposed to a certain level of QA.

Action #3 for the supplier: Please clarify how (and if) this requirement has been addressed. Please also clarify when this requirement will be fulfilled.

Comment 3.1: Even if the supplier has not provided ‘stable versions’ on the open source platform as required (please see issue #1 concerning ‘provision of software’) we are unsure from the content of the report to what extent this requirement has been (and will be) addressed.

Issue #4 – Identical software under both “GPLv3 or later” and “MPLv2 or later”

PREFORMA requires (as expressed in section 2.3 of version 1.0 of D4.3) that ‘The “MPLv2 or later” version and the “GPLv3 or later” version of the software that are developed, maintained, and distributed shall always be identical.’

Observation: As the source code of the software has not yet been provided as required (please see issue #1 above) it is unclear to us if this requirement will be fulfilled (when the supplier will adhere to the requirements expressed in issue #1). We note that source code is provided (on a site controlled by the supplier): http://dpfmanager.org/community.html#source_code

Action #4 for the supplier: As the source code of the software has not yet been provided as required (in D4.3) it remains to be seen if the supplier needs to take further action concerning this issue.

Comment 4.1: We note that the supplier currently maintains software under several different open source licenses on the development platform (which is fine). However, it remains to be seen precisely which source code will be provided that fulfils the requirements expressed in D4.3 (including those related to issue #1 and issue #4). Only after source code has been provided as required it is possible (and meaningful) to assess issue #1 and issue #4.

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

PREFORMA requires that an executable shall be provided for each platform. As detailed in section 2.5 of version 1.0 of D4.3 executables (and corresponding source code) shall always be provided on the open source platform "for several different platforms (at least for: MS Windows 7, Mac OSX, common Linux distributions including Ubuntu, Fedora, Debian, and Suse). For each platform specific executable there shall always be an up-to-date corresponding source code that can be downloaded as a single file."

Observation: The report from Easy Innova lacks detail (the report does not contain exact links to specific zip files containing executables for each platform) concerning this requirement and it is unclear to us when the supplier will fulfil this requirement.

Action #5 for the supplier: The supplier needs to provide an executable of the source code for each deployment platform as required (in version 1.0 of D4.3) on the open source portal. Please clarify when this requirement will be fulfilled.

Comment 5.1: The minutes from the meeting with PREFORMA partners and representatives for all three suppliers (held 28 October 2015) clearly states that "source code and executables shall be provided on the Open Source portal". The importance of this requirement has been stressed several times since it was initially communicated to all six suppliers in December 2014 via version 1.0 of deliverable D4.3 (including in the review of the July release). In light of this, we are very puzzled by the fact that the executables (and source code) have not been provided on the Open Source portal as required in version 1.0 of D4.3.

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

PREFORMA requires that the software can be used via standard web browsers. Please see section 2.5 of version 1.0 of D4.3 for details: "First, using any standard web browser any individual, both working for a memory institution and other interested individuals, can use the open source software which is provided for use at the open source project website for checking conformance for a specific file format." In this requirement, a 'standard web browser' refers to "the most recent stable release of each of the following web browsers: Chrome, Firefox, Internet Explorer, and Opera."

Observation: The report from Easy Innova lacks detail concerning this requirement and it is unclear to us when the supplier will fulfil this requirement.

Action #6 for the supplier: The supplier needs to provide the software for use via a standard web browser as required (in version 1.0 of D4.3). Please clarify when this requirement will be fulfilled.

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

PREFORMA requires that the supplier provides detailed documentation concerning interpretation of the technical specification of each file format used. Please see section 2.1 in version 1.0 of D4.3 for details: “Each open source project is expected to evolve detailed documentation concerning precisely how technical specifications of file formats are being interpreted and implemented in software. Such details is critical feedback for organisations maintaining technical specifications of file formats and from open provision of details concerning precise interpretations of different parts of a technical specification has been (and should be) interpreted will constitute a very valuable
resource for the broader open source and standards communities. With evolving precision in such open publication (via issue trackers and other means in each Open Source project) there will be an ongoing process for scrutiny of interpretations that eventually promotes improved quality of both how technical specifications should be clarified and how technical specifications should be interpreted and implemented in software.”

Observation: The report from Easy Innova lacks detail concerning this requirement and it is unclear to us when the supplier will fulfil this requirement.

Action #7 for the supplier: The supplier needs to address this requirement as required (in version 1.0 of D4.3). Please clarify when this requirement will be fulfilled.

Comment 7.1: This requirement is a novel aspect of the PREFORMA R&D project and this requirement it critical for successfully addressing the R&D challenge in PREFORMA.

Comment 7.2: Please note that this requirement addresses two aspects. First, it addresses complete and consistent interpretation of the technical specification of each file format (as specified). This seeks to contribute to an improved technical specification of each file format (thereby contributing to improved quality in standardisation). Second, it addresses complete and consistent interpretation of the technical specification when implemented in software. This seeks to contribute to an improved congruence between the software implementation of a specific file format and its technical specification (thereby contributing to improved quality in faithful software implementation of file formats).

Issue #8 – Provision of software which can be redistributed in a cascade

PREFORMA requires that the supplier provides all code (i.e. all source code; tool chain for building executables; and executables etc.) under open source licenses (on the open source

portal) and that all code can be distributed and redistributed by any individual. Please see section 2.3 in version 1.0 of D4.3 for further details:
'All software developed and maintained in each open source project will be provided under two specific open source licenses (www.opensource.org), namely: both Mozilla Public License "MPL v2.0 or later" and under GNU General Public license 3.0 "GPLv3 or later". This implies that all source code from the open source project that is necessary for creating an executable can be distributed (in a cascade) under these specific licenses to anyone. For example, a user at the memory institution Riksarkivet can download the complete source code for all software maintained in the open source project and redistribute the downloaded complete source code (under these specific licenses) to another organisation wishing to use the software.'

Observation: The report from Easy Innova lacks detail concerning this requirement and it is unclear to us when the supplier will fulfil this requirement. Further, it is critical that the supplier has all the necessary rights for provision of source code and executables under the two specific licenses (under "MPL v2.0 or later" and "GPL v3 or later") and provision of a tool chain (under OSI-approved licenses) for building executables (under "MPL v2.0 or later" and "GPL v3 or later") as required in version 1.0 of D4.3.

Action #8 for the supplier: The supplier needs to address this requirement as required (in version 1.0 of D4.3). Please clarify when the source code, tool chain for building executables, and executables on the open source portal will be provided as required in version 1.0 of D4.3.

Comment 8.1: This requirement is a pre-requisite for gaining full control of developed software for memory institutions and as such it is critical for successfully addressing the R&D challenge in PREFORMA. It is not sufficient to only host software on the development platform and other sites which PREFORMA partners do not control. When suppliers provide the software on the open source portal (according to the PREFORMA requirements in version 1.0 of D4.3) it follows that PREFORMA partners will have some control of developed software, and legal risks for any organisation wishing to use software from the PREFORMA project are also reduced.

Comment 8.2: In light of previous discussions, we would like to stress that PREFORMA requirements for provision of software on the open source platform (i.e. when the supplier makes an external distribution of the complete software as required in version 1.0 of D4.3) will fulfil some basic requirements for provisions of software (e.g. minimises legal risks for any user of the software, ensures some control for PREFORMA partners and external contributors, and promotes longevity of developed solutions). However, this should not be confused with efforts undertaken by the supplier for developing long-term sustainable open source communities related to developed software (on GitHub and elsewhere).

Comment 8.3: We note that strict adherence to licensing requirements (i.e. any individual must be able to redistribute the software (the complete source code and executables) under "MPLv2 or later" and "GPLv3 or later") is a necessary, but not sufficient, pre-requisite for any planned activities related to integration of software. Before the supplier provides software (the complete source code and executables) under "MPLv2 or later" and "GPLv3 or later" on the open source platform as required in version 1.0 of D4.3 all efforts related to integration are potentially

wasted. In case the supplier would fail to adhere to these licensing requirements, it would make it impossible to use and integrate developed solutions as planned in PREFORMA.

Comment 8.4: Strict adherence to licensing requirements (i.e. any individual must be able to redistribute the software (the complete source code and executables) under “MPLv2 or later” and “GPLv3 or later”) is a necessary, but not sufficient, pre-requisite for anticipated development of a long-term sustainable business ecosystem related to developed software. Before the supplier provides software (the complete source code and executables) under “MPLv2 or later” and “GPLv3 or later” on the open source platform as required in version 1.0 of D4.3 all efforts from the other two suppliers (and all external contributors) related to the software would potentially be wasted.
After the supplier has provided the software on the open source platform (i.e. which imply that an external distribution of the software has taken place) as required in version 1.0 of D4.3 it is a strong sign to the other two suppliers (and any other individual who is a potential external contributor) that the supplier is convinced that the supplier has all the necessary rights to distribute the software. In case the supplier would fail to adhere to the requirements in version 1.0 of D4.3 it would significantly reduce business opportunities for the two other suppliers (and any other current and potential participant in the broader community related to the software). For this reason, it is critical that the supplier urgently provides the software on the open source platform.

Comment 8.5: In addition to fulfilment of requirements for provision of software on the open source portal (as detailed in version 1.0 of D4.3), the supplier is also expected to provide and promote PREFORMA software via other channels. For example, appropriate means for promotion of the broader communities related to PREFORMA software include provision of ‘Live-CDs’ (or via USB:s) containing the complete source code and executables which allow anyone to use and scrutinise developed software without the need for installation.

Comment 8.6: For promotion of the broader business ecosystem and development communities, when executables (under “MPLv2 or later” and “GPLv3 or later”) of software are made available (irrespectively of how) it is important to always make the corresponding complete source code available (under “MPLv2 or later” and “GPLv3 or later”). It is essential to always clearly indicate open source licenses used. Plans for future releases (as indicated in an up-to-date roadmap for the project) should also indicate future (planned) availability of executables and source code. For example, the web page provided via a site controlled by the supplier (http://dpfmanager.org/) which include links to executables for different platforms (http://dpfmanager.org/index.html#download) currently lacks information about licenses. The page should clearly indicate that software is made available under “MPLv2 or later” and “GPLv3 or later”. Further, this web page should also include links to the complete corresponding source code (made available under “MPLv2 or later” and “GPLv3 or later”) from which the executable has been developed.

Standardisation efforts

The standardization effort seems to be on track and developing according to expectations.

Gap analysis and next steps

Major gap of this first prototype were bugs and usability issues rendering the software not usable for an in depth-evaluation. These issues need to (and already have been) address asap. Since DPF Manager has been created from scratch some functionality (e.g. the policy definition) make a rather infant impression and need significant improvements both in terms of usability and stability. In terms of usability we suggest an intense feedback look with the later users of the tool.

Feedback on the final release - Oct 2015 - MediaConch

In this document we have compiled the comments that we received so far by the members of the PREFORMA Evaluation Committee on the final release of the first prototyping phase (end of October 2015), including those on the text in the Final report.

The comments are normally taken in to the document more or less in the shape they were sent in, e.g. without filtering and editing, and relate both to specific and more general issues. Sometimes they might reflect different opinions but overall we hope that they provide a useful input for the next period.

The aim is that these comments shall serve as a base for further discussions.

1. General comments

MediaArea has created a platform for the open source project based on the GitHub platform. The different tools GitHub provides for open source projects fills most of the different needs of communication within an open source project. However, the community is not tools and technology but people cooperating to reach common goals. In that respect the activity of the community is very low and it is important that the work to achieve this is increased. One of the main challenges for the project is to have a thriving community so that the software will live on when the Preforma project ends.

Throughout the First Design and the First Prototype phase, MediaArea has proven to be a communicative, critical and responsive partner in the project. I have personally gotten to know the team as a driven and dedicated group of people that see in the project the opportunity to contribute to the wider audiovisual preservation community.

The group has shown to be a stimulating and critical conversation partner that aims to improve project outcomes by voicing inefficiencies and uncertainties about common supplier topics, such as licensing and the shared API. With the **Supplier Response to Feedback on the Intermediate release – July 2015** the MediaArea team has responded adequately and in-depth to comments from the Preforma evaluation team. Throughout conference presentations to technical experts in the audiovisual archiving community and thanks to open discussions on message boards and mailing lists, the group's networking and standardisation efforts are getting increasing attention.

Instructions how to use the software can be found at
https://mediaarea.net/MediaConch/documentation/HowToUse.html
Instructions how to install I have not found at all.

2. Result from examination

I tried the offered test file first. The output was exactly as described.
Then I tried one of the AVI files (with FFV1 v1 codec) stored in our storage. It worked the same.

Even if the AVI container was not recommended, MediaConch can deal with it.

3. The Conformance Checker

Installation, tested on UI (Mac), CLI (both regular and via Homebrew), works well. The help section is extensive and useful.

DOCUMENTATION:

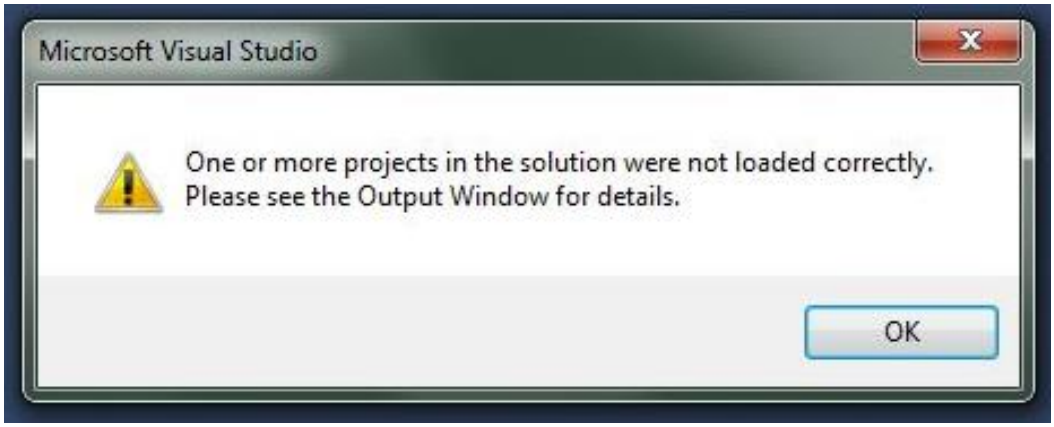
- The website https://mediaarea.net/MediaConch/ covers most of the important information
- Positive: Getting Started, How To Use and Software Demo Areas

SETUP:

- Installer provided (tested for Windows)
- Easy installation and running

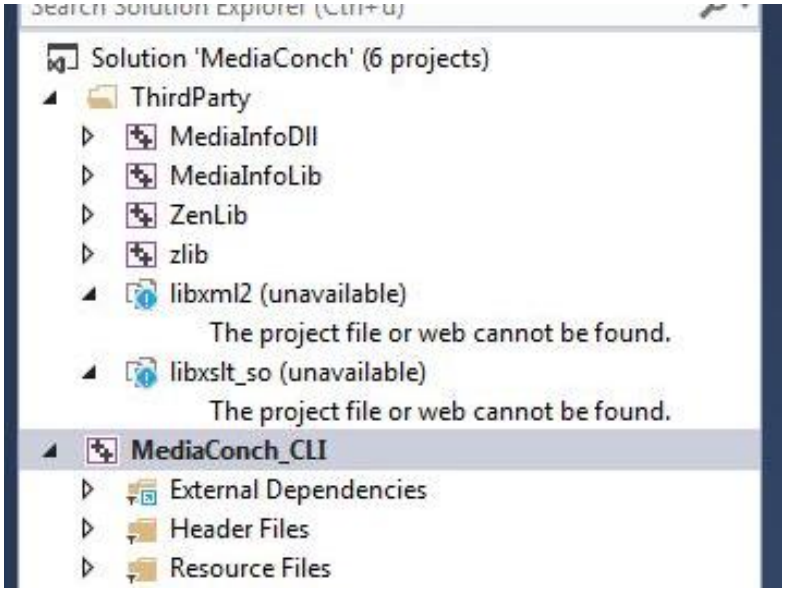
BUILD:

- On Windows:
  - Main README.md in github should give more information on how to build the project on Windows – currently it is just one line: "You need to install Microsoft Visual Studio 2013"
  - Opening the MediaConch/Project/MSVC2013/MediaConch.sln warnings accursed:

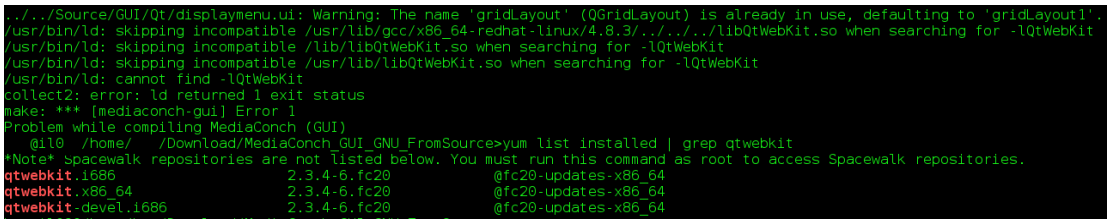


One or more projects in the solution were not loaded correctly.

  - Libxml2 and libxslt_so seems to be missing



  - GUI and CLI Versions could not be built
- On Fedora 20:
  - Main README.md in github lists all dependencies under Linux and the commands how to install them with *yum install*
  - Positive: In the main folder is a sh-Script which should do all install steps
  - GUI could not be built: on the list of dependencies is qt4 and qtwebkit listed, which I have also installed
  - But I addressed the following problem:



Existing Library QtWebKit 4.8.3 was skipped -> Error

  - CLI version built without problems

I downloaded the Windows version and the installation was fine, clear and without problems. Conformance checker works fine. I tried with 2 different files.

- There isn't the "Help" tab
- There isn't any reference to the PREFORMA project.
- There isn't any link to the GitHub page
- To use the service you need to register. The registration is easy and mandatory fields are minimal. Registration and validation user is easy, quick and it works perfect.
- With a 250 MB AV file it says that is too large. There aren't any specification about which sizes are accepting.

4. Final report

In the final three months of the prototyping phase, the MediaConch team has refined defined conformance checking XML schemas for file reporting; added new functionalities and features to its GUI, CLI, and web- based UI; and, improved on its documentation, demonstration, and test file corpus for research and development.

As I was at the IASA and the FIAT/IFTA, I can confirm the success of the discussions around the presentations of MediaCronch.

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

Progress is good, with the approval of the CELLAR working group being most impressive. The group works on clarifying the ways in which information is presented to users. It would be useful for future reports to include conference presentations and feedback on how they were received by the respective audiences - in the case of MediaInfo for example it's important to note the IASA Technical Committee meeting, the wrongly perceived lack of involvement of which was source for some contestation.

Bug reports are addressed and fixed in the recent version.

Testing

MediaArea is building a varied sample of test files and is dedicated to finding out the various ways in which they can malperform.

Dissemination and community building

Several instruments have been set-up in order to support community building. These are:

- A blog
- Tweets (13 Tweets, 19 follower on Dec. 2th 2015)

Overall, the measures taken are good. The most important thing is to keep them updated and synced with the progress within and especially **beyond** the PREFORMA project. Nothing is worse than an outdated wesite/blog/doc/HowTo!

As an instrument of gaining more attraction a live webinar could be done.

Community building efforts are good and happen both with the open source developer and the audiovisual archiving communities. I look very much forward to the white paper as it will be a useful tool to show archivists how the work done can impact their choices and workflows.

Open Source approach

Please find below feedback from the Skövde partner concerning the "Prototyping Phase 1 – Final Report" from MediaConch, which is focused on issues related to open source. For convenience, when we below write "the report from MediaConch" we refer to the "Prototyping

Phase 1 – Final Report" from MediaConch. In light of our previous feedback (concerning the "Prototyping Phase 1 – Intermediate Report" from MediaConch that we received in August 2015) which had references to version 2.0 of deliverable D4.3, let us first clarify that all references to D4.3 in this feedback refers to version 1.0 of this deliverable (i.e. the version of D4.3 that was made available to all suppliers in December 2014 after acceptance in the PREFORMA internal review). As agreed in later discussions between PERFORMA partners and suppliers, even if there are no substantial differences between versions 1.0 and 2.0 of D4.3 (and formally it was version 2.0 that is the outcome of the PREFORMA review), this feedback refers to requirements expressed in version 1.0 of D4.3. Our feedback is based on PREFORMA requirements which all are necessary pre-requisites for a successful outcome of the PREFORMA project.

Having reviewed the content of the "Prototyping Phase 1 – Final Report" from MediaConch we make a number of observations. Overall, we find that the supplier has made significant progress. However, we identify 8 issues in this feedback which all must be addressed. For several of these identified issues it is now urgent that the suppliers adhere to the requirements if the project should have a chance to successfully address the PREFORMA R&D challenge.

Issue #1 – Provision of source code

PREFORMA requires that a supplier provides the complete source code (i.e. a single zip-file containing all necessary files for creating a running instance of the source code) under two specific licenses ("MPL v2.0 or later" and "GPL v3 or later") on the open source portal (i.e. http://www.preforma-project.eu/open-source-portal.html). There must be one zip-file containing 'the complete source code necessary for each deployment platform' (as required in D4.3).

Observation: The report from MediaConch lacks direct links to a set of zip-files. Further, all links from the report refer to the development platform (i.e. the GitHub repository) instead of the open source portal as required in D4.3. If source code is only made available via the development platform (GitHub) it does not fulfil the PREFORMA requirements for how source code shall be provided.

Action #1 for the supplier: The supplier must provide the complete source code (as required in D4.3) on the open source portal. It is now very urgent that the supplier fulfils this requirement. When will the supplier provide the source code as required in version 1.0 of D4.3?

Comment 1.1: There are several technical and legal reasons for why PREFORMA has expressed this requirement. In short, it is an absolute requirement that the supplier distributes the complete source code (i.e. the supplier must make an 'external distribution') before the software can be used and redistributed by anyone which represents other organisations (i.e. before a supplier has made an external distribution of the source code to another organisation, such as the open source portal, the software cannot be used by anyone without significant risks).

Comment 1.2: The minutes from the meeting with PREFORMA partners and representatives for all three suppliers (held 28 October 2015) clearly states that "source code and executables shall

be provided on the Open Source portal". The importance of this requirement has been stressed several times since and it was communicated to all six suppliers in December 2014 via version 1.0 of deliverable D4.3 (and also in the review comments related to the release July 2015). In light of this, we are very puzzled by the fact that the source code (and executables) have not been provided on the Open Source portal as required in version 1.0 of D4.3.

Comment 1.3: For a number of reasons (including control, copyright and patent related reasons) it is standard practice in all European public sector framework agreements that the supplier of open source software must provide the software to the customer (i.e. through an external distribution of the software) in order to minimise risks for the customers. These reasons have also informed PREFORMA requirements concerning these aspects and it is therefore critical that the supplier provides the software on the open source portal (since the portal is controlled by a different organisation, in this case the PREFORMA partner Promoter, which is an organisation that is external to the supplier).

Comment 1.4: From our initial analysis of source code available on the development platform and the site controlled by the supplier (https://mediaarea.net/MediaConch/downloads/source.html) we have some concerns for how the software to be provided on the open source portal may be licensed. From the approach taken by the supplier, it seems that fundamental licensing requirements in PREFORMA will not be fulfilled when software will be provided by the supplier (assuming that code currently available in the zip files (e.g. https://mediaarea.net/download/binary/mediaconch/15.10/MediaConch_CLI_15.10_GNU_FromSource.tar.gz) controlled by the supplier will be provided by the supplier on the open source portal). Potential legal issues are one important reason for why PREFORMA require that the supplier provides the software (source code, build environment and executables) as required in version 1.0 of D4.3. We note that it can even be impossible to use the software if the supplier lacks all the necessary rights for providing the software as required in version 1.0 of D4.3 and it is consequently not surprising that it is standard practice in public sector procurement to require that the supplier provides the software (i.e. when the supplier provides the source code and binaries it is required that the supplier has all the necessary rights to do so). For the PREFORMA consortium (and any other user of software which has been distributed under "GPLv3 or later" and "MPLv2 or later" from PREFORMA), it is unrealistic if all users must first conduct a legal analysis before using the software. Precisely for these reasons, it is now critical that the supplier provides the code (i.e. an 'external distribution' must take place and it is the supplier that must do this). Please note that it must be possible for any individual to distribute (and redistribute in a cascade) the software (source code and executables) under "MPLv2 or later" and "GPLv3 or later". Please note that it must be possible for any individual to distribute (and redistribute in a cascade) the software (source code and executables) under "MPLv2 or later" and "GPLv3 or later". For PREFORMA it is critical that the supplier has obtained all necessary rights so that any individual (without any restriction) can distribute (and redistribute in a cascade) the software (the complete source code and executables) under "MPLv2 or later" and "GPLv3 or later". After the supplier has provided the software on the open source portal it will be possible and meaningful to conduct a more comprehensive analysis of these (rather complex) issues.

Comment 1.5: Related to the requirements for provision of source code, the PREFORMA requirements also state that the supplier shall make an external distribution of “all digital assets” related to “all open source projects” at the end of the project, please see section 1 of version 1.0 of D4.3 for further details.

Issue #2 – Provision of ‘road-map’ on the development platform

PREFORMA requires that a supplier provides an up-to-date road-map for the different versions of the software. Please see section 2.3 in version 1.0 of D4.3: “There shall be an up-to-date roadmap with detailed milestones for different (development version, stable version, and deployed (LTS) version) releases.”

Observation: The report from MediaConch lacks detail concerning the requirement for an ‘up-to-date roadmap’ and it is unclear to what extent the supplier has addressed (and plans to address) this requirement.

Action #2 for the supplier: Please clarify how (and if) this requirement has been addressed. Please also clarify when an up-to-date roadmap will be prominently exposed (to potential external contributors) on the development platform.

Comment 2.1: Provision of an up-to-date roadmap for the project which targets external (potential and active) external contributors (i.e. a roadmap which does not address PREFORMA partners) is an important enabler for promotion of external contributions.

Issue #3 – Time-based provision of ‘stable’ releases

PREFORMA requires that the supplier provides, on a monthly basis, releases which have been exposed to a certain level of QA. Please see section 1 in version 1.0 of D4.3: “Stable versions (provided on a monthly basis) have been exposed to a certain level of QA in the development process.”

Observation: The report from MediaConch lacks detail concerning this requirement and it is unclear to us to what extent such ‘stable versions’ have been exposed to a certain level of QA.

Action #3 for the supplier: Please clarify how (and if) this requirement has been addressed. Please also clarify when this requirement will be fulfilled.

Comment 3.1: Even if the supplier has not provided ‘stable versions’ on the open source platform as required (please see issue #1 concerning ‘provision of software’) we are unsure from the content of the report to what extent this requirement has been (and will be) addressed.

Issue #4 – Identical software under both “GPLv3 or later” and “MPLv2 or later”

PREFORMA requires (as expressed in section 2.3 of version 1.0 of D4.3) that ‘The “MPLv2 or later” version and the “GPLv3 or later” version of the software that are developed, maintained, and distributed shall always be identical.’

Observation: As the source code of the software has not yet been provided as required (please see issue #1 above) it is unclear to us if this requirement will be fulfilled (when the supplier will adhere to the requirements expressed in issue #1). We note that source code is provided on the development platform.

Action #4 for the supplier: As the source code of the software has not yet been provided as required (in D4.3) it remains to be seen if the supplier needs to take further action concerning this issue.

Comment 4.1: We note that the supplier currently maintains software under several different open source licenses on the development platform (which is fine). However, it remains to be seen precisely which source code will be provided that fulfils the requirements expressed in D4.3 (including those related to issue #1 and issue #4). Only after source code has been provided as required it is possible (and meaningful) to assess issue #1 and issue #4.

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

PREFORMA requires that an executable shall be provided for each platform. As detailed in section 2.5 of version 1.0 of D4.3 executables (and corresponding source code) shall always be provided on the open source platform "for several different platforms (at least for: MS Windows 7, Mac OSX, common Linux distributions including Ubuntu, Fedora, Debian, and Suse). For each platform specific executable there shall always be an up-to-date corresponding source code that can be downloaded as a single file."

Observation: The report from MediaConch lacks detail (the report does not contain exact links to specific zip files containing executables for each platform) concerning this requirement and it is unclear to us when the supplier will fulfil this requirement.

Action #5 for the supplier: The supplier needs to provide an executable of the source code for each deployment platform as required (in version 1.0 of D4.3) on the open source portal. Please clarify when this requirement will be fulfilled.

Comment 5.1: The minutes from the meeting with PREFORMA partners and representatives for all three suppliers (held 28 October 2015) clearly states that "source code and executables shall be provided on the Open Source portal". The importance of this requirement has been stressed several times since it was initially communicated to all six suppliers in December 2014 via version 1.0 of deliverable D4.3 (including in the review of the July release). In light of this, we are very puzzled by the fact that the executables (and source code) have not been provided on the Open Source portal as required in version 1.0 of D4.3.

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

PREFORMA requires that the software can be used via standard web browsers. Please see section 2.5 of version 1.0 of D4.3 for details: "First, using any standard web browser any individual, both working for a memory institution and other interested individuals, can use the

open source software which is provided for use at the open source project website for checking conformance for a specific file format." In this requirement, a 'standard web browser' refers to "the most recent stable release of each of the following web browsers: Chrome, Firefox, Internet Explorer, and Opera."

Observation: The report from MediaConch lacks detail concerning this requirement and it is unclear to us when the supplier will fulfil this requirement.

Action #6 for the supplier: The supplier needs to provide the software for use via a standard web browser as required (in version 1.0 of D4.3). Please clarify when this requirement will be fulfilled.

Comment 6.1: The website "MediaConchOnline" (https://mediaarea.net/MediaConchOnline/) contains a button for "Login" (https://mediaarea.net/MediaConchOnline/login). However, from the information in the report and on the web page it is unclear what is provided on this closed webpage. Further, even if the software can be used after registration it would not fulfil the requirement as it is required that any user should be able to use the software using a standard web browser without this (and any other form of) restriction.

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

PREFORMA requires that the supplier provides detailed documentation concerning interpretation of the technical specification of each file format used. Please see section 2.1 in version 1.0 of D4.3 for details: "Each open source project is expected to evolve detailed documentation concerning precisely how technical specifications of file formats are being interpreted and implemented in software. Such details is critical feedback for organisations maintaining technical specifications of file formats and from open provision of details concerning precise interpretations of different parts of a technical specification has been (and should be) interpreted will constitute a very valuable
resource for the broader open source and standards communities. With evolving precision in such open publication (via issue trackers and other means in each Open Source project) there will be an ongoing process for scrutiny of interpretations that eventually promotes improved quality of both how technical specifications should be clarified and how technical specifications should be interpreted and implemented in software."

Observation: The report from MediaConch lacks detail concerning this requirement and it is unclear to us when the supplier will fulfil this requirement. However, we note from the report that there has been significant activity related to standardisation of the formats addressed which is very positive.

Action #7 for the supplier: The supplier needs to address this requirement as required (in version 1.0 of D4.3). Please clarify when this requirement will be fulfilled.

Comment 7.1: This requirement is a novel aspect of the PREFORMA R&D project and this requirement it critical for successfully addressing the R&D challenge in PREFORMA.

Comment 7.2: Please note that this requirement addresses two aspects. First, it addresses complete and consistent interpretation of the technical specification of each file format (as specified). This seeks to contribute to an improved technical specification of each file format (thereby contributing to improved quality in standardisation). Second, it addresses complete and consistent interpretation of the technical specification when implemented in software. This seeks to contribute to an improved congruence between the software implementation of a specific file format and its technical specification (thereby contributing to improved quality in faithful software implementation of file formats).

Issue #8 – Provision of software which can be redistributed in a cascade

PREFORMA requires that the supplier provides all code (i.e. all source code; tool chain for building executables; and executables etc.) under open source licenses (on the open source portal) and that all code can be distributed and redistributed by any individual. Please see section 2.3 in version 1.0 of D4.3 for further details:
'All software developed and maintained in each open source project will be provided under two specific open source licenses (www.opensource.org), namely: both Mozilla Public License "MPL v2.0 or later" and under GNU General Public license 3.0 "GPLv3 or later". This implies that all source code from the open source project that is necessary for creating an executable can be distributed (in a cascade) under these specific licenses to anyone. For example, a user at the memory institution Riksarkivet can download the complete source code for all software maintained in the open source project and redistribute the downloaded complete source code (under these specific licenses) to another organisation wishing to use the software.'

Observation: The report from MediaConch lacks detail concerning this requirement and it is unclear to us when the supplier will fulfil this requirement. Further, it is critical that the supplier has all the necessary rights for provision of source code and executables under the two specific licenses (under "MPL v2.0 or later" and "GPL v3 or later") and provision of a tool chain (under OSI-approved licenses) for building executables (under "MPL v2.0 or later" and "GPL v3 or later") as required in version 1.0 of D4.3.

Action #8 for the supplier: The supplier needs to address this requirement as required (in version 1.0 of D4.3). Please clarify when the source code, tool chain for building executables, and executables on the open source portal will be provided as required in version 1.0 of D4.3.

Comment 8.1: This requirement is a pre-requisite for gaining full control of developed software for memory institutions and as such it is critical for successfully addressing the R&D challenge in PREFORMA. It is not sufficient to only host software on the development platform and other sites which PREFORMA partners do not control. When suppliers provide the software on the open source portal (according to the PREFORMA requirements in version 1.0 of D4.3) it follows that PREFORMA partners will have some control of developed software, and legal risks for any organisation wishing to use software from the PREFORMA project are also reduced.

Comment 8.2: In light of previous discussions, we would like to stress that PREFORMA requirements for provision of software on the open source platform (i.e. when the supplier

makes an external distribution of the complete software as required in version 1.0 of D4.3) will fulfil some basic requirements for provisions of software (e.g. minimises legal risks for any user of the software, ensures some control for PREFORMA partners and external contributors, and promotes longevity of developed solutions). However, this should not be confused with efforts undertaken by the supplier for developing long-term sustainable open source communities related to developed software (on GitHub and elsewhere).

Comment 8.3: We note that strict adherence to licensing requirements (i.e. any individual must be able to redistribute the software (the complete source code and executables) under “MPLv2 or later” and “GPLv3 or later”) is a necessary, but not sufficient, pre-requisite for any planned activities related to integration of software. Before the supplier provides software (the complete source code and executables) under “MPLv2 or later” and “GPLv3 or later” on the open source platform as required in version 1.0 of D4.3 all efforts related to integration are potentially wasted. In case the supplier would fail to adhere to these licensing requirements, it would make it impossible to use and integrate developed solutions as planned in PREFORMA.

Comment 8.4: Strict adherence to licensing requirements (i.e. any individual must be able to redistribute the software (the complete source code and executables) under “MPLv2 or later” and “GPLv3 or later”) is a necessary, but not sufficient, pre-requisite for anticipated development of a long-term sustainable business ecosystem related to developed software. Before the supplier provides software (the complete source code and executables) under “MPLv2 or later” and “GPLv3 or later” on the open source platform as required in version 1.0 of D4.3 all efforts from the other two suppliers (and all external contributors) related to the software would potentially be wasted. After the supplier has provided the software on the open source platform (i.e. which imply that an external distribution of the software has taken place) as required in version 1.0 of D4.3 it is a strong sign to the other two suppliers (and any other individual who is a potential external contributor) that the supplier is convinced that the supplier has all the necessary rights to distribute the software. In case the supplier would fail to adhere to the requirements in version 1.0 of D4.3 it would significantly reduce business opportunities for the two other suppliers (and any other current and potential participant in the broader community related to the software). For this reason, it is critical that the supplier urgently provides the software on the open source platform.

Comment 8.5: In addition to fulfilment of requirements for provision of software on the open source portal (as detailed in version 1.0 of D4.3), the supplier is also expected to provide and promote PREFORMA software via other channels. For example, appropriate means for promotion of the broader communities related to PREFORMA software include provision of ‘Live-CDs’ (or via USB:s) containing the complete source code and executables which allow anyone to use and scrutinise developed software without the need for installation.

Comment 8.6: For promotion of the broader business ecosystem and development communities, when executables (under “MPLv2 or later” and “GPLv3 or later”) of software are made available (irrespectively of how) it is important to always make the corresponding complete source code available (under “MPLv2 or later” and “GPLv3 or later”). It is essential to always clearly indicate open source licenses used. Plans for future releases (as indicated in an up-to-date roadmap for the project) should also indicate future (planned) availability of

executables and source code. For example, the web page provided via a site controlled by the supplier (mediaarea.net) which include links to executables for different platforms (https://mediaarea.net/MediaConch/download.html) currently lacks information about licenses. The page should clearly indicate that software is made available under “MPLv2 or later” and “GPLv3 or later”. Similarly, each page related to each specific platform (e.g. https://mediaarea.net/MediaConch/downloads/windows.html) should also clearly indicate that software is made available under “MPLv2 or later” and “GPLv3 or later”. Further, each such page (e.g. https://mediaarea.net/MediaConch/downloads/windows.html) should also include links to the complete corresponding source code (made available under “MPLv2 or later” and “GPLv3 or later”) from which the executable has been developed.

Standardisation efforts

The group’s approach seems to be fruitful and feasible within the time given in the project - the approval of the working group (after the final report was handed in) is an excellent sign.

Gap analysis and next steps

Clear view of the work ahead.

Major gap of this first prototype were bugs.

ANNEX 5: SUPPLIERS' END OF DESIGN PHASE #2 REPORTS

The following pages contain the full end of phase reports filled in by the suppliers to report the work done during the re-design phase.