DELIVERABLE

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

D8.8 - Monitoring of the Open Source Project implementation

Revision: version 1.0 – final (first release)

Authors:

Björn Lundell (University of Skövde)
Jonas Gamalielsson (University of Skövde)

Contributors:

Björn Lundell (University of Skövde)
Jonas Gamalielsson (University of Skövde)

Reviewers:

Börje Justrell, Magnus Geber, Benjamin Yousefim Per Elfner (Riksarkivet)
Bengt Neiss (Kungliga biblioteket)
Bert Lemmens (Packed)

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	Organisation	Description
0.1	2015-08-31	B. Lundell & J. Gamalielsson	HS	Initial draft
0.2	2015-11-30	B. Lundell & J. Gamalielsson	HS	Initial full draft
0.3	2016-02-08	B. Lundell & J. Gamalielsson	HS	Full draft after internal review
0.4	2016-02-16	B. Lundell & J. Gamalielsson	HS	Full draft after internal review and terminology clarification
1.0	2016-02-16	C.Prandoni	PROMOTER	Final version for submission

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This deliverable contains original unpublished work except where clearly indicated otherwise.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

Deliverable D8.8 reports on monitoring of the Open Source Project implementations. Based on development efforts for each supplier, this deliverable provides feedback on their use of: an open work practice for development; frequent open releases; and promotion activities aiming towards a sustainable community. In particular, it focuses on establishing sustainable communities, together with an assessment of how this is succeeded. The deliverable presents an evaluation of how each open source project implementation adheres to requirements expressed in deliverable D4.3. In so doing, the deliverable provides an evaluation of the extent to which best practices from community driven open source projects have been adopted with adherence to full transparency for all digital assets. Specifically, the evaluation considers software and associated digital assets provided via links to developed and provided resources (including source code, executables, and test files) and tools (including software configuration management system, mailing lists, and build environment) used in each open source project. An important outcome from this evaluation is a report on adherence to requirements (as specified in D4.3) and an assessment of how contracted organisations have managed to establish thriving and long-term sustainable open source communities of relevance for memory institutions and other stakeholder groups.

1 INTRODUCTION

PREFORMA (PREservation FORMAts for culture information/e-archives) is a Pre Commercial Procurement (PCP) project financially supported by the European Commission under its FP7-ICT Programme. The objective of deliverable D8.8 is to report on monitoring of the Open Source Project implementations. This is an initial version of deliverable D8.8 which reports on what suppliers have achieved at project month 24. Deliverable D8.8 will be updated at month 36.

The initial version of deliverable D8.8 complements the main deliverable from the prototyping phase (deliverable D8.3) and the updated version of deliverable D8.8 complements the main deliverable from the second prototyping phase (deliverable D8.5). Deliverable D8.8 is supplementing deliverable D8.3 on how the suppliers have adhered to utilising effective open source work practices in their first three prototypes. Since the three suppliers provided their respective "intermediate report"[1] and their respective "final report"[2], an overarching observation is that they have gradually accounted for feedback from PREFORMA review comments[3] and have gradually adjusted their work in order for the PREFORMA project to successfully achieve its goals and fulfill the PREFORMA R&D challenge.

Based on development efforts for each supplier undertaken in relation to each open source project, this deliverable provides feedback on adherence to requirements in deliverable D4.3 and an evaluation of how each open source project use: an open work practice for development; frequent open releases; and promotion activities aiming towards a sustainable community. In so doing, this deliverable reports on observations concerning achievements made so far.

The rest of this document is organised as follows. Section 2 elaborates on adherence to requirements for provision of open source projects, section 3 presents an initial evaluation of sustainability of open source projects and associated communities, and section 4 concludes the content of the deliverable.

[1] From the "PREFORMA Prototyping Phase 1" after completion of their work in July 2015

[2] From the "PREFORMA Prototyping Phase 1" after completion of their work in October 2015

[3] This includes the feedback from the Skövde partner concerning the "PREFORMA Prototyping Phase 1 – Intermediate Report" provided in August 2015 (and provided as part of the "Feedback on the intermediate release" from the PREFORMA consortium) and the feedback from the Skövde partner concerning the "PREFORMA Prototyping Phase 1 – Final Report" provided in November 2015 (and provided as part of the "Feedback on the final release" from the PREFORMA consortium)

2 ADHERENCE TO REQUIREMENTS FOR PROVISION OF OPEN SOURCE PROJECTS

This section elaborates on establishment of long-term sustainable open source projects and highlights important aspects concerning what needs to achieved in order to establish thriving and long-term sustainable open source communities.

To achieve long-term sustainable open source communities of relevance for memory institutions and other stakeholder groups there is a need for contracted organisations to successfully manage and conduct a number of fundamental activities. To this end, the section presents necessary activities for successfully addressing the PREFORMA R&D challenge, and in so doing elaborates fulfillment of requirements for provision of open source projects (as specified in deliverable D4.3). Successful establishment of long-term sustainable open source communities also presupposes adherence to business and user needs, whilst at the same time adhering to community norms, values, and established work practices in the broader open source communities.

2.1 ON DOCUMENTATION

Long-term sustainability of an open source project is promoted through effective communication of long-term vision of goals and plan for how the project will evolve over time. Such longevity is supported through a number of means, including: provision of roadmaps (and other documentation and information) tailored for different stakeholder groups; and documentation of source code and associated digital assets. For additional information on these means, please see deliverable D4.3.

For **provision of roadmaps** and associated information from each open source project it is essential to address: potential code contributors (external and independent of the PREFORMA project) with relevant information in order to attract interest and code contributions from external contributors. Similarly, information from each open source project also needs to address potential external users (and contents of roadmaps need to be tailored accordingly). Relevant information includes information concerning conditions for active participation. Similarly, relevant information also includes conditions for use and distribution of software. Further, conditions for involvement and handling of (and potential transfer of) copyright, trademarks, patents, and other aspects which impact on the extent to which it is possible to attract contributions for open source projects need to be addressed. Governance issues and organisation of long-term management for each open source project needs to be properly addressed (e.g. foundations[4] and other forms of organisational entities should be considered). The scope for roadmaps need to cover strategic plans and releases planned for software from each open source project at least for the time period until December 2020.

[4] For example, the Document Foundation (initially established for the LibreOffice open source project, https://www.documentfoundation.org/) may constitute a relevant source for information and inspiration.

For provision of roadmaps and other documentation it is essential to address: stakeholders contributing to improved quality of the file format which is implemented in the open source project. Such information needs to address different stakeholder groups related to each file format, including: participants in the working group governed by the organisation maintaining the file format; representatives for other organisations interested in the precise interpretation of the file format and how it has (and should) be implemented in software; and other individuals interested in technical, strategic, and policy aspects of how the file format has been (and should be) interpreted and implemented in software. Relevant information includes information concerning processes for how interpretations (and misinterpretations) of file formats and associated implementation in software can be made transparent. The important mission of achieving clarity and improved quality of file formats requires 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. Such ongoing processes need to be transparent and inclusive for all relevant stakeholders in an open collaboration hosted on open collaboration platforms (e.g. GitHub) as further elaborated in section 2.1 in deliverable D4.3.

For provision of roadmaps and other documentation aimed at other organisations and suppliers (including potential business partners) it is essential to address business, service, and support offerings in order to promote a sustainable business related to each open source project. To promote open collaboration, opportunities for collaboration with local partners[5] should ideally be highlighted. Opportunities for use of, and services related to, integrated software should be highlighted. For example, software provided by the contracted PREFORMA supplier A (from the open source project it is contracted with) which has been integrated with software developed in open source projects provided by one (or both) of the other contracted PREFORMA suppliers B and C constitutes a potentially very valuable business offering for each supplier, as well as the broader community. With this approach, a supplier focusing on one media type may benefit from software developed by the two other suppliers focusing on the two other media types. Thereby each supplier can provide services and business offerings related to three open source projects (each one focused on a specific media type, i.e. text, image, and A/V) even if they primarily focus on one (for which they are contracted by PREFORMA). Further, information from each open source project should be exposed in a way which ideally attracts a broader business ecosystem.

For provision of **documentation of source code and associated digital assets** it is essential to adopt community values and norms as well as established practices amongst professional organisations providing open source software to customers (something which necessitates fulfilment of minimum basic requirements in established practices in framework agreements for

[5] Successful collaboration related to sustainable open source projects often consists of a variety of business partners which collaborate. In many cases it is essential to have local knowledge of needs in specific domains and countries, whilst at the same time having access to specialised know-how often provided by internationally recognised partners which collaborate in vibrant open business ecosystems. For memory institutions it may be beneficial to develop good relations with local business partners that genuinely understand the domain in which they operate.

public sector procurement of open source software). For details on documentation of source code (with suggestions for informative references) see further section 2.1 in deliverable D4.3.

2.2 ON USE OF DEVELOPMENT PLATFORM AND TOOLS

Long-term sustainability of an open source project and associated communities is promoted through use of an open collaboration platform (such as GitHub) and use of open source tools with associated work practices. Such longevity is promoted by establishment of long-term sustainable communities by adoption of best practices from open source development which adheres to community norms and values. For additional information on these practices, please see deliverable D4.3.

An important overarching principle for development and use of the development platform is that all information provided on the platform for each open source project is self-contained with strict adherence to established licenses to aid clarity concerning conditions for participation and involvement in the project. From previous research it is well known that unclear conditions for participation in, and use of software from, open source projects may cause significant tension in communities and consequently inhibit opportunities for collaboration and integration with other projects.

2.3 ON PROVISION OF SOURCE CODE

Long-term sustainability of an open source project and associated communities is promoted through provision of source code under clear licensing and IPR conditions. To aid longevity of software developed and maintained (on the open platform) in each open source project the option “or later” is required for the two specific copyleft licenses used (“MPL v2.0 or later” and “GPLv3 or later”) for all software which is to be distributed to (and used by) memory institutions and the broader community. It is required that all software can be distributed in a cascade under these two specific copyleft licenses. On a regular basis, software shall be distributed[6] to and provided on the Open Source Portal[7] (OSP), which is a site controlled by the PREFORMA consortium. For further information concerning provision of source code and associated digital assets, please see deliverable D4.3.

[6] The issue of when software can be considered to have been distributed is a complex one which has received researchers attention, see e.g.: https://fosdem.org/2016/schedule/event/triggering_copyleft/. However, it is clear that distribution has occurred when suppliers provide software under the PREFORMA licenses (i.e. “GPLv3 or later” and “MPLv2 or later”) on the OSP as required in PREFORMA. As distribution triggers copyleft obligations it is important that suppliers provide (monthly) stable releases of the software on the OSP in order to maximise availability long-term and minimise legal risks for users.

[7] Each supplier has a dedicated web page for download of software on the OSP. For veraPDF the software is provided via http://www.preforma-project.eu/verapdf-download.html, for Easy Innova the software is provided via http://www.preforma-project.eu/dpfmanager-download.html, and for MediaArea the software is provided via http://www.preforma-project.eu/mediaconch-download.html.

2.4 ON PROVISION OF BUILD ENVIRONMENT

Long-term sustainability of an open source project is promoted through provision of build environment and its source code. The build environment (i.e. the specific tool chain used for creation of a running instance of the open source code) must be provided under an open source license, i.e. a license approved by the Open Source Initiative (www.opensource.org). For further information concerning provision of open source tools creation of an executable (i.e. a running instance of the open source code) for each deployment platform, please see deliverable D4.3.

2.5 ON PROVISION OF EXECUTABLES

Long-term sustainability of an open source project is promoted through provision of executables. There shall always be executables for several different deployment platforms (for details, please see D4.3). For each platform specific executable there shall always be an up-to-date corresponding open source code that can be downloaded as a single file from the OSP and the open collaboration platform. To promote longevity of software it is essential to provide effective instructions for how to create the executable from the source code. For further information concerning provision of executables, please see deliverable D4.3.

2.6 ON OPEN SOURCE IMPLEMENTATION OF STANDARDS

Long-term sustainability of a technical specification of a file format is promoted through implementation in open source software for which there is transparent information on how the specification of the file format has been interpreted. Such transparent information includes both the source code itself and associated documentation of precisely how different features in the technical specification of the file format has been implemented in software. Longevity of files is promoted when the file format used for each file is implemented in open source software which is provided under the two specific copyleft licenses used (“MPL v2.0 or later” and “GPLv3 or later”). Such provision of open source software promotes quality of the technical specification of the file format and may significantly support standardisation processes. Establishment of an open source community for interpretation of synthetic test files (including files perceived to be “correct” and files perceived to be “incorrect” by the specific individual or organisation creating a specific file) promotes quality of the file format and supports individuals implementing the file format in software. As the set of synthetic test files available on the open collaboration platform increases (with associated interpretations and comments on each interpretation) evolves, this open collaboration supports consensus on how the file format should be interpreted and thereby contributes to improved standardisation. For further information concerning implementation of file formats in open source software, please see deliverable D4.3.

2.7 ON ACHIEVING SUSTAINABLE OPEN SOURCE PROJECTS

Long-term sustainability of a vibrant business ecosystem presupposes sustainable open source projects and associated communities. There are a number of business models used by companies involved in open source projects and fundamental to most is adherence to and appreciation of values and norms established in open collaborations. There are a number of

factors which impact on establishment of successful and long-term sustainable open source communities. It has been noted that establishment of long-term sustainable communities is a challenge and some even consider it as an art[8]. For example, the extent to which an open source project successfully manages to attract and maintain contributions from talented contributors has shown to be an important aspect[9]. Previous research has shown mixed success for different open source projects concerning establishment of vibrant communities[10]. Similarly, another important aspect is collaboration between communities for a file format standard and communities for its implementation in open source software[11]. Further, an open source project needs to recognise and be adaptive to that there may be a number of different motivations for external contributors[12].

Use and reuse of software from different open source projects need to recognise and adhere to licensing and other (technical, legal, and cultural) conditions. For long-term sustainable open source projects (including software from PREFORMA) it is critical to adhere to all such conditions in order to successfully achieve intended goals. This includes strict adherence to licensing requirements when software is to be integrated with software from other projects. For example, if supplier A in PREFORMA wishes to integrate software from supplier B and C it is critical that all software strictly adheres to the same licensing requirements in order to allow for integration, distribution, and redistribution of integrated software.

For further information concerning achieving sustainable open source projects, please see deliverable D4.3.

[8] See for example: Bacon, J. (2009) The Art of Community: Building the New Age of Participation, O'Reilly, ISBN: 978-0-596-15671-8.

[9] See for example research results from the evolution of the LibreOffice project: Gamalielsson, J. and Lundell, B. (2014) Sustainability of Open Source software communities beyond a fork: How and why has the LibreOffice project evolved?, The Journal of Systems and Software, Vol. 89(1), pp. 128-145. http://dx.doi.org/10.1016/j.jss.2013.11.1077.

[10] See for example: Teixeira, J., Robles, G. and González-Barahona, J. M. (2015) Lessons learned from applying social network analysis on an industrial Free/Libre/Open Source Software ecosystem, Journal of Internet Services and Applications, Vol. 6(14), pp. 1-27. http://dx.doi.org/10.1186/s13174-015-0028-2

[11] See for example research results on a widely adopted file format and its implementation in open source software: Gamalielsson, J., Lundell, B., Feist, J., Gustavsson, T. and Landqvist, F. (2015) On organisational influences in software standards and their open source implementations, Information and Software Technology, Vol. 67, pp. 30-43. http://dx.doi.org/10.1016/j.infsof.2015.06.006

[12] See for example: Bonaccorsi, A. and Rossi, C. (2006) Comparing Motivations of Individual Programmers and Firms to Take Part in the Open Source Movement: From Community to Business, Knowledge, Technology & Policy, Winter 2006, Vol. 18(4), pp. 40-64.

3 EVALUATION OF SUSTAINABILITY OF OPEN SOURCE PROJECTS AND ASSOCIATED COMMUNITIES

This section presents an initial assessment of how contracted organisations have managed to establish thriving and long-term sustainable open source communities of relevance for memory institutions and other stakeholder groups.

For each subsection we provide an overarching observation followed by an initial assessment of each open source project. The assessment is based on important aspects concerning what needs to achieved in order to establish thriving and long-term sustainable open source communities (as raised in section 2 and detailed in deliverable D4.3). Our assessment of software related to each open source project has been conducted by the end of November 2015 (and subsequent analysis[13] will be conducted once software has been provided as required in D4.3).

We refer (below) to the three open source projects as follows: “veraPDF” refers to the open source project implementing **text**; “DPF Manager” refers to the open source project implementing **image**; and “MediaConch” refers to the open source project implementing **A/V**. Further, when referring to the supplier behind each open source project we refer to: “veraPDF consortium” when referring to the “veraPDF” open source project; “Easy Innova” when referring to the “DPF Manager” open source project; and “MediaArea” when referring to the “MediaConch” open source project.

3.1 ASSESSING DOCUMENTATION

Concerning an up-to-date road-map for the different versions of the software which includes detailed milestones for different (development version, stable version, and deployed (LTS) version) releases, we make the following observations.

From our assessment of the work conducted by the **veraPDF consortium** concerning this aspect we make the following observations. First, we note that the veraPDF consortium provides a roadmap (http://verapdf.org/roadmap/). However, the content of the roadmap currently provided[14] is primarily targeted at the PREFORMA consortium[15] instead of targeting external potential contributors. For example, there is currently no roadmap with information focused on external contributors from different stakeholder groups beyond PREFORMA and there is no information concerning release plans for the time period until 2020. Such information may be fundamental to any potential contributor and collaborator interested in longevity of

[13] In addition to this we have conducted some analysis on source code available on development platforms, but in order to conduct a relevant analysis there is a need for suppliers to first provide the software as required in D4.3.

[14] as observed 8 February 2016.

[15] In acknowledging that a detailed roadmap for the PREFORMA consortium may also have significant value, such a roadmap may also serve a worthwhile purpose.

software. Second, at time for the review[16], we note that source code has been provided on the OSP[17].

From our assessment of the work conducted by **Easy Innova** concerning this aspect we make the following observations. First, we note that Easy Innova does not provide a roadmap with information focused on external contributors from different stakeholder groups beyond PREFORMA and there is no information concerning release plans for the time period until 2020. Such information may be fundamental to any potential contributor and collaborator interested in longevity of software. Second, at time for the review[18], we note that source code has been provided on the OSP[19].

From our assessment of the work conducted by **MediaArea** concerning this aspect we make the following observations. First, we note that MediaArea does not provide a roadmap with information focused on external contributors from different stakeholder groups beyond PREFORMA and there is no information concerning release plans for the time period until 2020. Such information may be fundamental to any potential contributor and collaborator interested in longevity of software. Second, at time for the review[20], we note that source code has been provided on the OSP[21].

3.2 ASSESSING USE OF DEVELOPMENT PLATFORM AND TOOLS

Concerning use of an open collaboration platform (such as GitHub) and use of open source tools with associated work practices, we make the following observations.

From our assessment of the work conducted by the **veraPDF consortium** concerning this aspect we make the following observations. First, we note that the open collaboration platform (GitHub) is actively used by the veraPDF consortium. Second, a number of components and software (under several different open source licenses) are maintained by the consortium on the platform. Third, the OSP has been used for provision of software from the open source project maintained on GitHub. However, we have been unable to compile source code provided by the supplier on the OSP by use of open source tools.

From our assessment of the work conducted by **Easy Innova** concerning this aspect we make the following observations. First, we note that the open collaboration platform (GitHub) is actively used by Easy Innova. Second, a number of components and software (under several different open source licenses) are maintained by the consortium on the platform. Third, the OSP has been used for provision of software from the open source project maintained on

[16] 8 February 2016.

[17] http://www.preforma-project.eu/verapdf-download.html

[18] 8 February 2016.

[19] http://www.preforma-project.eu/dpfmanager-download.html

[20] 8 February 2016.

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

GitHub. However, we have been unable to easily[22] compile source code provided by the supplier on the OSP by use of an open source tool chain[23].

From our assessment of the work conducted by **MediaArea** concerning this aspect we make the following observations. First, we note that the open collaboration platform (GitHub) is actively used by MediaArea. Second, a number of components and software (under several different open source licenses) are maintained by the consortium on the platform. Third, the OSP has been used for provision of software from the open source project maintained on GitHub. We have been able to compile source code provided by the supplier on the OSP. It should be noted that the supplier has so far[24] not fulfilled the PREFORMA requirement concerning provision of an open source tool chain on the OSP for compilation of source code provided on the OSP.

3.3 ASSESSING PROVISION OF SOURCE CODE

The Table below shows an overview of how each supplier has provided monthly releases[25] of source code on the OSP. Each row shows different releases expected to be provided by each supplier in order to fulfill the PREFORMA requirement for time based (monthly) stable releases[26]. The two rows representing months during which the PREFORMA consortium has reviewed the work[27] conducted by the suppliers is highlighted in bold face. Each cell in the table identifies each release with associated date for when it was released.

[22] It is expected that it shall be possible to use an open source tool chain for compiling the source code provided on the OSP by use of a simple command (e.g. “make all”) via a script provided by the supplier.

[23] It should be noted that the supplier currently provides proprietary tools for compilation of source code, something which does not fulfil PREFORMA requirements.

[24] In noting that requirements for provision of an open source tool chain is included in deliverable D4.3, we acknowledge that suppliers will fulfil these requirements no later than early 2016 (before the open source workshop) as agreed with PREFORMA partners.

[25] As observed 8 February 2016.

[26] For details on PREFORMA requirements concerning provision of time based (monthly) stable releases, see deliverable D4.3.

[27] It should be noted that PREFORMA partners considered reports (“PREFORMA Prototyping Phase 1 – Intermediate Report” provided to report on achievements made until July 2015 and “PREFORMA Prototyping Phase 1 – Final Report” provided to report on achievements made until October 2015) from each supplier and all achievements made so far (including software made available by suppliers on GitHub and their own websites) at time for each review.

Month \ Supplier	veraPDF consortium[28]	Easy Innova[29]	MediaArea[30]
December 2015	Not provided	Not provided	31 December 2015[31]
November 2015	0.8.5 (11 Dec. 2015)	1.2.3 (10 Dec. 2015)	30 November 2015
October 2015	**0.6.46**[32] (4 Nov. 2015)	**1.2**[33] (28 Oct. 2015)	**31 October 2015**[34]
September 2015	Not provided	1.1.1 (02 Oct. 2015)	39 September 2015
August 2015	0.4.11 (16 Sep. 2015)	1.1 (29 Sep. 2015)	31 August 2015
July 2015	**0.2.9** (16 Jul. 2015)	**1.0** (31 Jul. 2015)	**31 July 2015**

Concerning provision of source code under clear licensing and IPR conditions, we make the following observations.

From our assessment of the work conducted by the **veraPDF consortium** concerning this aspect we make the following observations. First, we note that source code (for stable releases) has been provided on the OSP. From our initial analysis of source code provided on the OSP we note that the source code has not yet been provided on the OSP under the two specific PREFORMA licenses so that the complete software can be used and distributed (in a cascade) as required in PREFORMA. We make the following observations to support the outcome of this initial analysis.

[28] http://www.preforma-project.eu/verapdf-download.html

[29] http://www.preforma-project.eu/dpfmanager-download.html

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

[31] For MediaArea, the date format is used to name each release.

[32] For example, the stable release for the source code under “MPL v2.0 or later” for Debian is provided via the link: http://www.preforma-project.eu/downloads/veraPDF/src/all-platforms/veraPDF-0.6-20151104-MPL.zip. Further, the stable release for the source code under “GPL v3 or later” for Debian is provided via the link: http://www.preforma-project.eu/downloads/veraPDF/src/all-platforms/veraPDF-0.6-20151104-GPL.zip

[33] For example, the stable release for the source code for Debian is provided via the link: http://www.preforma-project.eu/downloads/DPFManager/Release-1.2/Debian/src17-2015-10-28.zip

[34] For example, the stable release for the source code for Debian is provided via the link: http://www.preforma-project.eu/downloads/MediaConch/2015-10-31/src17-2015-10-31.zip

First, we observe that the supplier has chosen to provide software under two different branches (one provided under “GPLv3 or later” and another provided under “MPL v2”) on the OSP. In acknowledging that this choice fulfils PREFORMA requirements provided that software in both branches is identical in both branches, we do not recommend this as it increases complexity. Instead we recommend that the supplier provides one branch (dual licensed under the two specific PREFORMA licenses, “GPLv3 or later” and “MPLv2 or later”). Second, in acknowledging that the MPLv2 implicitly is to be interpreted with “or later” we recommend the supplier to explicitly clarify that all software is provided under “MPLv2 or later” to thereby avoid uncertainty[35] in the future. Third, we observe that the supplier has chosen to include software in the MPL branch under a different license (Apache 2.0), which inhibits distribution of the complete software (in a cascade) as required in PREFORMA. Fourth, we observe that the complete source code provided on the OSP is not provided under the PREFORMA licenses. To promote longevity in order to successfully address the PREFORMA R&D challenge it is critical that the complete source code can be incorporated in any future version of the PREFORMA licenses. For example, PREFORMA requires that the complete source code can be incorporated in other software (developed much later) to be provided under GPLv4, GPLv5, or any later version of this license, and it cannot be assumed that software provided under other open source licenses can be included in such later versions of the PREFORMA licenses. Consequently, software under Apache v2.0 cannot be used[36] in PREFORMA. Thereby, distribution of software under other licenses will inhibit longevity of software as required in PREFORMA and inhibit successful addressing of the PREFORMA R&D challenge. Fifth, even if the supplier has obtained all necessary rights for implementation of a file format in software[37] so that it can be distributed under the Apache 2.0 license it does not necessarily follow that such rights have been obtained for distribution of software under “MPLv2 or later” and under “GPLv3 or later”. Consequently, for the broader open source community and any potential external contributor it is essential to clarify that all necessary rights have been obtained. Sixth, we observe that there is a lack of information concerning how anyone can obtain the corresponding source code which relate to executables[38] provided on the supplier’s own web site. It is essential to provide information concerning how anyone can obtain corresponding source code

[35] For example, it may be that MPLv3, MPLv4, or any later version of this license may be drafted differently in the future (perhaps beyond the existence of the Mozilla Foundation). If, and when, software from this open source project will be incorporated in other projects, it is advantageous to be as clear as possible on licensing conditions (especially since there may be a complex interplay with existing and future licenses for SEPs that may inhibit use of the software under certain versions of the MPL license).

[36] If there are any additional restrictions (which implies a need for inclusion of the Apache 2.0 license) it does not conform to MPLv2.

[37] For example, we observe that the supplier includes Apache PDFBox (licensed under Apache 2.0) as part of the software provided on the OSP.

[38] For example, the supplier makes an executable available on its own web page without clear information concerning under which license the software is provided and how the corresponding source code can be obtained (see http://verapdf.org/software/). We recommend that this web page clarifies that software is provided under the PREFORMA licenses and how corresponding source code can be obtained.

on the relevant web pages controlled by the supplier. Therefore, it is essential that the supplier addresses this fundamental PREFORMA requirement in order to meet the PREFORMA R&D challenge and provide software which may be of significant value for memory institutions and other stakeholders. Seventh, we observe that the supplier has not provided a stable release each month as required in PREFORMA.

From our assessment of the work conducted by **Easy Innova** concerning this aspect we make the following observations. First, we note that source code (for stable releases) has been provided on the OSP. From our initial analysis of source code provided on the OSP we note that the source code has not yet been provided on the OSP under the two specific PREFORMA licenses so that the complete software can be used and distributed (in a cascade) as required in PREFORMA. We make the following observations to support the outcome of this initial analysis.

First, we observe that the licensing conditions for the software provided on the OSP are unclear. For example, the file “README.md” which is included in the file “src17-2015-12-10.zip” provided on the OSP refers to two files (“LICENSE.GPL” and “LICENSE.MPL”) that do not exist[39] in the file “src17-2015-12-10.zip”. Further, the supplier has chosen not to include any license text in the Java source code files in the file “src17-2015-12-10.zip”. To clarify that the software is provided under the PREFORMA licenses (“GPLv3 or later” and “MPL v2 or later”), it is necessary to include license text in the header of each source code file. Second, we observe that the supplier has chosen to include software under a different license (Apache 2.0), which inhibits distribution of the complete software (in a cascade) as required in PREFORMA. Third, we observe that the complete source code provided on the OSP is not provided under the PREFORMA licenses. To promote longevity in order to successfully address the PREFORMA R&D challenge it is critical that the complete source code can be incorporated in any future version of the PREFORMA licenses. For example, PREFORMA requires that the complete source code can be incorporated in other software (developed much later) to be provided under GPLv4, GPLv5, or any later version of this license, and it cannot be assumed that software provided under other open source licenses can be included in such later versions of the PREFORMA licenses. Consequently, software under Apache v2.0 cannot be used[40] in PREFORMA. Thereby, distribution of software under other licenses will inhibit longevity of software as required in PREFORMA and inhibit successful addressing of the PREFORMA R&D challenge. Fourth, even if the supplier has obtained all necessary rights for implementation of a

[39] In addition, license conditions for software maintained on the open collaboration platform (GitHub) are also unclear. For example, the file “https://github.com/EasyinnovaSL/DPFManager/blob/develop/README.md” which is maintained on GitHub includes a link to the file “LICENCE.GPL” (with a reference to https://github.com/EasyinnovaSL/DPFManager/blob/develop/LICENSE.GPL) and a link to the file “LICENCE.MPL” (with a reference to https://github.com/EasyinnovaSL/DPFManager/blob/develop/LICENSE.MPL). However, none of these license files are available and license information in source files maintained on GitHub are missing (8 February 2016).

[40] If there are any additional restrictions (which implies a need for inclusion of the Apache 2.0 license) it does not conform to MPLv2.

file format in software[41] so that it can be distributed under the Apache 2.0 license it does not necessarily follow that such rights have been obtained for distribution of software under “MPLv2 or later” and under “GPLv3 or later”. Consequently, for the broader open source community and any potential external contributor it is essential to clarify that all necessary rights have been obtained. Fifth, we observe that there is a lack of information concerning how anyone can obtain the corresponding source code which relate to executables[42] provided on the supplier’s own web site. It is essential to provide information concerning how anyone can obtain corresponding source code on the relevant web pages controlled by the supplier. Therefore, it is essential that the supplier addresses this fundamental PREFORMA requirement in order to meet the PREFORMA R&D challenge and provide software which may be of significant value for memory institutions and other stakeholders. Sixth, we observe that the supplier has not provided a stable release each month as required in PREFORMA.

From our assessment of the work conducted by **MediaArea** concerning this aspect we make the following observations. First, we note that source code (for stable releases) has been provided on the OSP. From our initial analysis of source code provided on the OSP we note that the source code has not yet been provided on the OSP under the two specific PREFORMA licenses so that the complete software can be used and distributed (in a cascade) as required in PREFORMA. We make the following observations to support the outcome of this initial analysis.

First, we observe that software is provided on the OSP under several different (open source and non-open source) licenses. For example, in addition to software provided under “GPLv3 or later” and “MPLv2 or later” there is also software distributed on the OSP under the BSD 2-clause[43], BSD 3-clause[44], zlib/libpng license[45], and Public Domain[46]. Second, we observe that the

[41] For example, we observe that the supplier includes Apache PDFBox (licensed under Apache 2.0) and Apache Camel (licensed under Apache 2.0) as part of the software provided on the OSP.

[42] For example, the supplier makes an executable available on its own web page (e.g. an executable for Windows is available via http://www.easyinnova.com/dpfmanager/Downloads/Current-release/Windows/dpf_manager-1.4.exe) without clear information concerning how the corresponding source code can be obtained. Further, there is a need to clarify if the software provided on the supplier´s own web site is identical to the software provided on the OSP. We note that the supplier use the phrase “Alpha release” related to the executables whereas the software provided on the OSP shall be monthly stable releases. In assuming that the executable provided on the supplier’s own web site constitutes the latest version of a stable release, it is essential that the corresponding source code can be easily found. Hence, it is not sufficient to provide a general link to the open collaboration platform. We expect that the web page (http://www.dpfmanager.org/#download) clarifies how the complete source code which corresponds to each executable can be obtained.

[43] For example, the file aes.h which is provided in "src01-2015-12-31.zip” on the OSP is provided under this license. From this it is clear that this file is not provided under the required PREFORMA licenses (“GPLv3 or later” and “MPLv2 or later”).

[44] For example, the file sha2.c which is provided in "src01-2015-12-31.zip” on the OSP is provided under this license. From this it is clear that this file is not provided under the required PREFORMA licenses (“GPLv3 or later” and “MPLv2 or later”).

software provided on the OSP relies on software (licensed under different open source and proprietary licenses) which has not been distributed as required by PREFORMA. For example, there are dependencies to software licensed under the MIT license[47]. Third, we observe that the supplier claims (in the document “Supplier Updated Response to Feedback on the final release – Oct 2015”) that the software provided on the OSP relies on proprietary licensed software (“Visual C++ library” by Microsoft), something which does not conform with PREFORMA licensing requirements. Fourth, in acknowledging that the supplier stresses on the open collaboration platform that there are plans for providing software under the two specific PREFORMA licenses (“GPLv3 or later” and “MPLv2 or later”) it is currently unclear when the supplier will fulfil[48] this PREFORMA requirement. Fifth, we observe that there is a lack of information concerning how anyone can obtain the corresponding source code which relate to executables[49] provided on the supplier’s own web site. It is essential to provide information

---

[45] For example, the file tinyxml2.cpp which is provided in "src01-2015-12-31.zip” on the OSP is provided under this license. From this it is clear that this file is not provided under the required PREFORMA licenses (“GPLv3 or later” and “MPLv2 or later”).

[46] For example, the file base64.h which is provided in "src01-2015-12-31.zip” on the OSP lacks licensing information. We acknowledge that the supplier claims (in the document “Supplier Updated Response to Feedback on the final release – Oct 2015”) that this source code is in the Public Domain and is “by Bob Withers”. However, the file provided on the OSP states that “Ideas taken from work done by Bob Withers” and that “R1 2002-05-07 by Markus Ewald”. From this it is clear that this file is not provided under the required PREFORMA licenses (“GPLv3 or later” and “MPLv2 or later”).

[47] For example, the file “Reader_libcurl.cpp” contains a statement “#include "curl/curl.h” which refers to MIT-licensed open source software that has not been included in the file "src01-2015-12-31.zip” on the OSP. As the complete software must be provided on the OSP under the PREFORMA licenses (“GPLv3 or later” and “MPLv2 or later”) the supplier must include this software in the file "src01-2015-12-31.zip” which must be distributed on the OSP.

[48] We note that the supplier states the following intention concerning open source licensing on the open collaboration platform “All software and source code developed by MediaArea during the PREFORMA project will be provided under the following two open source licenses: GNU General Public License 3.0 (GPLv3 or later), Mozilla Public License (MPLv2 or later).” (https://github.com/MediaArea/MediaConch/blob/master/README.md). However, for fulfilment of PREFORMA requirements it is critical that the complete software is provided on the OSP under the PREFORMA licenses (“GPLv3 or later” and “MPLv2 or later”) so that it can be distributed (and redistributed) in a cascade to memory institutions and other relevant stakeholders.

[49] For example, the supplier makes an executable available on its own web page (e.g. executables for Windows are available via https://mediaarea.net/MediaConch/downloads/windows.html) without clear information concerning how the corresponding source code can be obtained. Further, there is a need to clarify if the software provided on the supplier´s own web site is identical to the software provided on the OSP. In assuming that the executable provided on the supplier’s own web site constitutes the latest version of a stable release, it is essential that the corresponding source code can be easily found. We expect that the web page (https://mediaarea.net/MediaConch/downloads/windows.html) clarifies how the complete source code which corresponds to each executable can be obtained.

concerning how anyone can obtain corresponding source code on the relevant web pages controlled by the supplier. Therefore, it is essential that the supplier addresses this fundamental PREFORMA requirement in order to meet the PREFORMA R&D challenge and provides software which may be of significant value for memory institutions and other stakeholders.

For the continued work, we expect that all the suppliers address the fundamental PREFORMA licensing requirements and thereby consequently fulfil PREFORMA tendering requirements. A fundamental prerequisite for successfully addressing the PREFORMA R&D challenge is 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”, something which is currently unclear. 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”.

3.4 ASSESSING PROVISION OF BUILD ENVIRONMENT

The Table below shows an overview of how each supplier has provided monthly releases of build environment[50] on the OSP. Each row shows different releases expected to be provided by each supplier in order to fulfill the PREFORMA requirement for time based (monthly) stable releases[51]. The two rows representing months during which the PREFORMA consortium has reviewed the work[52] conducted by the suppliers is highlighted in bold face. Each cell in the table identifies each release with associated date for when it was released.

[50] In noting that requirements for provision of an open source tool chain is included in deliverable D4.3, we acknowledge that suppliers will fulfil these requirements no later than early 2016 (before the open source workshop) as agreed with PREFORMA partners. Each open source project will be assessed accordingly once these requirements have been addressed according to deliverable D4.3.

[51] For details on PREFORMA requirements concerning provision of time based (monthly) stable releases, see deliverable D4.3.

[52] It should be noted that PREFORMA partners considered reports (“PREFORMA Prototyping Phase 1 – Intermediate Report” provided to report on achievements made until July 2015 and “PREFORMA Prototyping Phase 1 – Final Report” provided to report on achievements made until October 2015) from each supplier and all achievements made so far (including software made available by suppliers on GitHub and their own websites) at time for each review.

Month \ Supplier	veraPDF consortium[53]	Easy Innova[54]	MediaArea[55]
December 2015	Not provided	Not provided	Not provided
November 2015	Not provided	1.2.3 (10 Dec. 2015)	Not provided
October 2015	Not provided	**1.2**[56] (28 Oct. 2015)	Not provided
September 2015	Not provided	1.1.1 (02 Oct. 2015)	Not provided
August 2015	Not provided	1.1 (29 Sep. 2015)	Not provided
July 2015	Not provided	**1.0** (31 Jul. 2015)	Not provided

Provision of an open source licensed build environment by which the complete source code can be compiled is a PREFORMA requirement (as detailed in deliverable D4.3). However, as identified in the document "Feedback on the intermediate release" from the PREFORMA consortium to the suppliers, no such build environment was provided at that stage. In subsequent discussions concerning achievements made so far in light of the PREFORMA "Feedback on the intermediate release" it was agreed between the PREFORMA consortium and all three suppliers that fulfilment of this PREFORMA requirement could be deferred until the Open Source Workshop[57] in Stockholm in April 2016.

So far, one supplier has provided a build environment on the OSP. From our assessment of the work conducted by **Easy Innova** concerning provision of an open source licensed build environment on the OSP we observe that the build environment provided is not open source licensed. Hence, the conditions under which the build environment is provided does not fulfil the PREFORMA requirement and the supplier needs to resolve this.

3.5 ASSESSING PROVISION OF EXECUTABLES

The Table below shows an overview of how each supplier has provided monthly releases of executables on the OSP. Each row shows different releases expected to be provided by each supplier in order to fulfill the PREFORMA requirement for time based (monthly) stable

[53] http://www.preforma-project.eu/verapdf-download.html

[54] http://www.preforma-project.eu/dpfmanager-download.html

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

[56] For example, a non-open source licensed build environment for Debian is provided via the link: http://www.preforma-project.eu/downloads/DPFManager/Release-1.2/Debian/buildenv17-2015-10-28.zip

[57] http://opensourceworkshop.preforma-project.eu/

releases[58]. The two rows representing months during which the PREFORMA consortium has reviewed the work[59] conducted by the suppliers is highlighted in bold face. Each cell in the table identifies each release with associated date for when it was released.

Month \ Supplier	veraPDF consortium[60]	Easy Innova[61]	MediaArea[62]
December 2015	Not provided	Not provided	31 December 2015[63]
November 2015	0.8.5 (11 Dec. 2015)	1.2.3 (10 Dec. 2015)	30 November 2015
October 2015	**0.6.46**[64] (4 Nov. 2015)	**1.2**[65] (28 Oct. 2015)	**31 October 2015**[66]
September 2015	Not provided	1.1.1 (02 Oct. 2015)	39 September 2015
August 2015	0.4.11 (16 Sep. 2015)	1.1 (29 Sep. 2015)	31 August 2015
July 2015	**0.2.9** (16 Jul. 2015)	**1.0** (31 Jul. 2015)	**31 July 2015**

[58] For details on PREFORMA requirements concerning provision of time based (monthly) stable releases, see deliverable D4.3.

[59] It should be noted that PREFORMA partners considered reports ("PREFORMA Prototyping Phase 1 – Intermediate Report" provided to report on achievements made until July 2015 and "PREFORMA Prototyping Phase 1 – Final Report" provided to report on achievements made until October 2015) from each supplier and all achievements made so far (including software made available by suppliers on GitHub and their own websites) at time for each review.

[60] http://www.preforma-project.eu/verapdf-download.html

[61] http://www.preforma-project.eu/dpfmanager-download.html

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

[63] For MediaArea, the date format is used to name each release.

[64] For example, the stable release for the executable for Debian is provided via the link: http://www.preforma-project.eu/downloads/veraPDF/bin/all-platforms/verapdf-0.6.46-04112015.zip

[65] For example, the stable release for the executable for Debian is provided via the link: http://www.preforma-project.eu/downloads/DPFManager/Release-1.2/Debian/exec17-2015-10-28.zip

[66] For example, the stable release for the executable for Debian is provided via the link: http://www.preforma-project.eu/downloads/MediaConch/2015-10-31/exec17-2015-10-31.zip

Concerning provision of executables, we make the following observations.

From our assessment of the work conducted by the **veraPDF consortium** concerning this aspect we make the following observations[67]. First, we note that executables have been provided on the OSP. However, it should be noted that for each platform specific executable there shall always be an up-to-date corresponding open source code that can be downloaded as a single file from the OSP and the open collaboration platform. At time of writing[68], it has not been confirmed that this PREFORMA requirement has been fulfilled. The PREFORMA consortium[69] has so far been unable to compile[70] the source code provided on the OSP and therefore no resulting executable has been produced from the source code provided on the OSP. Consequently, the PREFORMA consortium has been unable to compare the executable provided by the supplier on the OSP with an executable produced from the source code on the OSP. Second, we note that the supplier has not made the software available for use via a standard web browser[71] as required in D4.3. It is therefore important that the supplier makes it possible to use the software via a standard web browser without any need for registration. These two observations have been communicated to the supplier.

From our assessment of the work conducted by **Easy Innova** concerning this aspect we make the following observations[72]. First, we note that executables have been provided on the OSP. However, from the information provided it is unclear[73] if the provided executables have been

[67] as observed 8 February 2016.

[68] as observed 8 February 2016.

[69] The Skövde partner and other partners in the PREFORMA consortium.

[70] The outcome of our analysis shows that we are unable to compile the source code using an open source licensed build environment as required in PREFORMA (the outcome of our analysis is based on use of OpenJDK, which is an open source licensed implementation of the Java platform that can be used for compilation of Java source code).

[71] We note (as observed 8 February 2016.) that the supplier does not provide online access to the software. Failure to provide online access to the software is a violation of PREFORMA requirements of open access to the software as a service as detailed in D4.3.

[72] as observed 8 February 2016.

[73] Since the supplier provides a build environment which fails to fulfill PREFORMA requirements concerning licensing of the build environment, it may be an indication of that the executables provided fail to fulfill PREFORMA requirements (as it is required in PREFORMA that all executables must be created by use of an open source licensed build environment, e.g. OpenJDK). Consequently, it is unclear if provided executables can be distributed and used under the two specific PREFORMA licenses (“MPL v2.0 or later” and “GPLv3 or later”).

created[74] using an open source licensed build environment, something which is required in PREFORMA. Hence, for each platform specific executable produced[75] using an open source licensed build environment there shall always be an up-to-date corresponding open source code that can be downloaded as a single file from the OSP and the open collaboration platform. Consequently, when the supplier provides executables produced using an open source licensed build environment on the OSP it will be meaningful to compare the executable which will be provided by the supplier on the OSP with an executable produced from the source code on the OSP. Second, we note that the supplier has not made the software available for use via a standard web browser[76] as required in D4.3. It is therefore important that the supplier makes it possible to use the software via a standard web browser without any need for registration. These two observations have been communicated to the supplier.

From our assessment of the work conducted by **MediaArea** concerning this aspect we make the following observations. First, we note that executables have been provided on the OSP. However, it is unclear if conditions for provision of executables on the OSP (as required in PREFORMA) are the same for executables provided on the supplier's own website. For example, the supplier provides executables for several platforms on its own site[77] but the license conditions for those executables are unclear. Second, we note that the supplier has not made the software available for use via a standard web browser[78] as required in D4.3. It is therefore important that the supplier makes it possible to use the software via a standard web browser without any need for registration. These two observations have been communicated to the supplier.

[74] It should be noted that the supplier provides a build environment on the OSP. However, the build environment provided is licensed under conditions which fail to fulfill PREFORMA licensing requirements (as the build environment is not open source). For this reason, the provided build environment cannot be used in PREFORMA, and there is a need for the supplier to instead provide an open source licensed build environment which fulfill PREFORMA requirements.

[75] This is currently (8 February 2016) unclear.

[76] We note (as observed 8 February 2016) that the supplier does not provide online access to the software. Failure to provide online access to the software is a violation of PREFORMA requirements of open access to the software as a service as detailed in D4.3.

[77] For example, the supplier provides executables for several different platforms, including windows (https://mediaarea.net/MediaConch/downloads/windows.html) and Debian (https://mediaarea.net/MediaConch/downloads/debian.html) without providing clarifying license conditions. It is important to clarify that executables are provided under the PREFORMA licenses ("MPL v2.0 or later" and "GPLv3 or later") on each webpage where an executable can be downloaded.

[78] We note (as observed 8 February 2016) that the supplier does not provide online access to the software. Failure to provide online access to the software is a violation of PREFORMA requirements of open access to the software as a service as detailed in D4.3.

3.6 ASSESSING OPEN SOURCE IMPLEMENTATION OF STANDARDS

Concerning implementation of file formats in open source software, we make a number of observations.

From our assessment of the work conducted by the **veraPDF consortium** concerning this aspect we make the following observations based on the information provided by the supplier. We note that the veraPDF consortium has been active related to international standardisation (ISO).

From our analysis of the content provided[79] on open collaboration platform (GitHub) it is evident that the veraPDF consortium provides[80] synthetic test files. However, there is a need to clarify licensing conditions for those files[81] on the open collaboration platform and fulfil PREFORMA licensing requirements for synthetic test files (see deliverable D4.3).

From our assessment of the work conducted by **Easy Innova** concerning this aspect we make the following observations based on the information provided by the supplier. We note that the Easy Innova has been active related to international standardisation (ISO).

From our analysis of the content provided on open collaboration platform (GitHub) it is evident that Easy Innova provides[82] test files. In acknowledging that file names used provide an indication of what aspect of the file format specific files[83] are supposed to test. However, there is a need to clarify meta data for each file and details concerning what aspects of each file format each different synthetic file it is supposed to test. Further, there is also a need to clarify licensing

[79] https://github.com/veraPDF/veraPDF-corpus/tree/staging/PDF_A-1b

[80] We note that section 2 in the “Prototyping Phase 1 Final Report” from the veraPDF consortium contains detail concerning synthetic test files provided on the open collaboration platform. Further, section 2 also contains references to other websites containing test files. However, there is a lack of clarity concerning licensing of synthetic test files provided on the open collaboration platform, and synthetic test files provided via other web sites fail to fulfil PREFORMA licensing requirements (as detailed in D4.3). For example, conditions for use of test files in the Bavaria suite (http://www.pdflib.com/knowledge-base/pdfa/validation-report/, specifically the files are available via http://www.pdflib.com/fileadmin/pdflib/Bavaria/2009-04-03-Bavaria-pdfa.zip) referred to in the “Prototyping Phase 1 Final Report” is licensed under the following conditions: “Redistributing all or parts of the Bavaria report or the accompanying test documents requires written permission of PDFlib GmbH.”

[81] There is a need clarify licensing conditions for all files in each test file directory and in the meta data of the file itself.

[82] Test files are available at https://github.com/EasyinnovaSL/DPFManager/tree/master/src/test/resources

[83] From the information provided it is currently unclear which of the files provided are synthetic files.

conditions for all files[84] on the open collaboration platform in order to fulfil PREFORMA licensing requirements for synthetic test files (see deliverable D4.3).

From our assessment of the work conducted by **MediaArea** concerning this aspect we make the following observations based on the information provided by the supplier. We note that MediaArea has been active related to standardisation in the context of IETF.

From our analysis of the content provided on open collaboration platform (GitHub) it is evident that the MediaArea provides synthetic test files[85] and a dedicated web page[86] for demonstration of various features of conformance checking. However, there is a need to provide licensing conditions for all files[87] on the open collaboration platform and fulfil PREFORMA licensing requirements for synthetic test files (see deliverable D4.3).

Further, there are two important aspects related to implementation of file formats which need to be addressed by all open source projects. First, each project needs to address 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, each supplier needs to address 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). These are important aspects for all projects which are key for successfully addressing the PREFORMA R&D challenge and all suppliers (and associated open source projects) need to increase attention[88] to these aspects.

3.7 ASSESSING ACHIEVEMENT OF SUSTAINABLE OPEN SOURCE PROJECTS

Concerning achieving sustainable open source projects with associated vibrant business ecosystems and communities, we make a number of observations. In this section we initially report on general observations for the three suppliers and thereafter provide observations and specific recommendations for each supplier.

---

[84] There is a need clarify licensing conditions for all files in each test file directory and in the meta data of the file itself.

[85] Test files are available at https://github.com/MediaArea/MediaConch_SampleFiles

[86] https://github.com/MediaArea/MediaConch/tree/master/Demo

[87] There is a need clarify licensing conditions for all files in each test file directory and in the meta data of the file itself.

[88] It is essential that suppliers successfully manage to engage the broader developer and user communities related to each file format (and media type) implemented in software in order to promote improved quality in technical specifications of file formats and in quality in software implementations of technical specifications of file formats.

From our assessment of the work conducted by all three suppliers concerning this aspect we make the following observations. First, we acknowledge the limited time frame during which software has been available on the open collaboration platform for each open source project. However, we note that roadmaps focused on external potential contributors and with content addressing a time frame well beyond the PREFORMA R&D project are lacking. For example, one would expect some indication concerning plans until December 2020. Second, from observation it is still unclear to what extent external contributions have so far been attracted. For future planning and action, we anticipate increased attention on sustainability of each project beyond the PREFORMA R&D project. Third, from our assessment of software provided so far on the OSP we note that there is scope for improvement concerning code transparency[89] and software architecture[90] for all open source projects[91]. For example, in case a memory institution

[89] Open source software which is developed and maintained in open source projects can be distributed to anyone for use, scrutiny, improvement, and redistribution according to its licensing conditions. When open source software is provided on open collaboration platforms and available via the web and other distribution channels, this promotes transparency and aid open collaboration. Previous research shows that open source projects may significantly promote transparency, both in terms of access to the source code and also in terms of a transparent open development model, something which facilitates scrutiny and external audit of open source software (von Krogh and Spaeth, 2007). There are a number of dimensions of transparency, which can be separately (or in combination) analysed. For example, coding practices is one important dimension of code transparency which has been analysed in previous research (e.g. Gamalielsson et al., 2012). See: Gamalielsson, J., Grahn, A. and Lundell, B (2012) Learning through analysis of coding practices in FLOSS projects, In Robles, G., González Barahona, J., Tebbens, W. and Hammouda, I. (Eds.) Proceedings of FLOSSEdu 2012: FLOSS Education - Long-term Sustainability, Tampere University of Technology, Department of Software Systems, Report 21, Tampere, ISBN 978-952-15-2938-2, pp. 13-19.; von Krogh, G. and Spaeth, S. (2007) The open source software phenomenon: Characteristics that promote research, The Journal of Strategic Information Systems, Vol. 16 (3), pp. 236-253.

[90] Previous research shows the importance of a well-designed modular software architecture for open source projects which are developed in open collaboration. For example, Crowston et al. (2012) stresses that “Modularity has been seen as key to the feasibility of distributed development” (p. 7:16) and it is widely acknowledged that development of open source software is a successful exemplar of distributed development (Fitzgerald, 2006). Further, research shows that a clear, transparent, and well-designed software architecture is essential for attracting external contributions since many developers typically contribute to just a single module (Scacchi, 2007). See: Crowston, K., Kangning, W., Howison, J., and Wiggins, A. (2012) Free/Libre open-source software development: what we know and what we do not know, ACM Computing Survey, Vol. 44(2), Article 7.; Fitzgerald, B. (2006) The transformation of open source software, MIS Quarterly, Vol. 30(4), pp. 587–598.; Scacchi, W. (2007) Free/Open Source Software Development: Recent Research Results and Methods, Advances in Computers, Vol. 69, pp. 243-295.

wishes to deploy a software component in which only one specific tendered file format for each media type (e.g. PDF/A-1 for text, or TIFF/IT for image) is implemented, it is essential that only the specific subset of the software developed in which the specific file format is implemented can easily be identified and reused (e.g. if a memory institution is only interested in the subset of the software implementing TIFF/IT, it is critical that the software architecture promotes easy reuse of this subset of the software without a need for incorporation of any software that is specific for implementing TIFF/EP). Fourth, concerning opportunities for integration, we acknowledge that some of these requirements are mandatory[92] whereas some are optional[93]. At the same time, it is clear that successful integration of software from the two other suppliers provides increased business opportunities. At this stage, progress on this aspect is (naturally) limited, but once developed software become more stable and provided as required so that it can be distributed to memory institutions (and other organisations), it is essential that suppliers take an active part in evolving a business ecosystem related to developed software.

Concerning the work performed by the veraPDF consortium, we provide the following recommendation in order to provide long-term sustainability of the open source project. From our analysis we note that the veraPDF consortium is an active contributor to PDF/A and its further development[94] (currently termed "PDF/A-next") within relevant working groups for standardisation of the file format within ISO. In this role, the veraPDF consortium may have unique opportunities to influence the further evolution of the file format in a way to ensure its future relevance for longevity of files produced in this format. One approach for achieving sustainability of "PDF/A-next" would be to promote that the ISO WG adopts a work practice for development of this new file format which is supplemented with development and maintenance of an open source-licensed reference implementation (licensed under "GPLv3 or later") within the ISO standardisation process. If an open reference implementation of "PDF/A-next" is deployed in software licensed under "GPLv3 or later" it becomes an inherent part of the ISO standardisation process for new versions of the PDF/A file format within ISO. Such an open reference implementation would constitute a significant step towards ensuring long-term relevance of the file format for memory institutions. Even if the ISO WG would not adopt a work-practice involving use of an open reference implementation licensed under "GPLv3 or later" in

---

[91] Without code transparency and a clear software architecture, open source projects may be less attractive for external stakeholder and more difficult to reuse software from the open source project. In the words of Gacek and Arief (2004): "An open source software system's architecture might be available or not. An unintentionally unavailable software architecture suggests that the structure exists in some people's minds only." (p. 37) Such a situation is something which all open source projects in PREFORMA must avoid. See: Gacek, C. and Arief, B. (2004) The Many Meanings of Open Source, IEEE Software, Vol. 21(1), pp. 34-40.

[92] Mandatory requirements are options 1, 5, 9, 13, 17, and 21 as specified in deliverable D4.3.

[93] Optional requirements are options 2, 3, 4, 6, 7, 8, 10, 11, 12, 14, 15, 16, 18, 19, 20, 22, 23, and 24 as specified in deliverable D4.3.

[94] For example, the veraPDF consortium reports (in "PROTOTYPING PHASE 1 FINAL REPORT") that they 'led the effort to create a new Part of PDF/A, currently termed "PDF/A-next".'

the standards development process it is of uttermost importance for the longevity of files (which are created in the file format) that the IPR conditions (specifically concerning patents) are such that anyone can implement “PDF/A-next” in software which is provided and distributed under “GPLv3 or later”. In case this cannot be achieved the relevance of “PDF/A-next” would be significantly reduced, especially for application areas with requirements for long-term maintenance of files. However, as the veraPDF consortium has a unique opportunity (as leaders of the WG in ISO standardisation) we expect a positive development and for the PREFORMA consortium we would very much support and welcome initiatives for development of an open source reference implementation of “PDF/A-next” which is provided and distributed under “GPLv3 or later” as an inherent part of the standardisation process.

Further, from our analysis of the software provided by the veraPDF consortium on the OSP, it seems that the software provided implements files formats which have not been included in the tender (e.g. ISO 32000-1). In noting that it is fine for a supplier to provide software beyond PREFORMA requirements (implementation of ISO 32000-1 in software is not a tender requirement) it is of uttermost importance that the supplier clearly communicates and convinces any potential external contributor that the supplier has obtained all necessary rights for implementing and distributing software under the PREFORMA licenses (i.e. “MPLv2 or later” and “GPLv3 or later”). One strategy which contributes to such clear communication concerns improvement of the content in the roadmap so that any potential contributor is convinced that the supplier has obtained all necessary rights.

Concerning the work performed by Easy Innova, we provide the following recommendation in order to provide long-term sustainability of the open source project. From our analysis we note that Easy Innova has taken action to establish “TI/A” as a new file format within ISO. However, before taking further actions concerning the new initiative “TI/A”, it is critical to ensure that the name “TI/A” can be used (for legal reasons) and to communicate its relation to the TIFF/IT and TIFF/EP standards (which are the two standards being tendered in the PREFORMA project). Overall, it is essential to extend the information concerning these issues to the broader community (including the road-map provided by the supplier which aims at external contributors). For attracting external contributors to the open source project it is important to clarify the conditions[95] for use of these and the new file format TI/A to the broader community. Therefore, it is critical to clarify precisely the overlap between these three file formats and also clarify any potential impact (which potentially) may inhibit use of software in which one (or several) of these file formats is implemented to the broader open source community. The fact that the supplier has ensured (as part of the format contract) the PREFORMA consortium that it

[95] There are a number of important reasons for this. For example, there are a number of patent declarations for TIFF/EP in the ISO database, whereas there are no such (at time of writing) for TIFF/IT. Therefore, in case a memory institution wishes to deploy a software component for checking conformance with only one of the file formats (e.g. TIFF/IT), it is essential that the software architecture easily allows this and for this reason code transparency is of paramount importance for developed software. Further, since many members of broader open source communities are extremely sensitive to patent related issues it is critical to clarify conditions for involvement in the open source project for potential external contributors.

has obtained all necessary rights for the work in PREFORMA (which includes that the supplier has obtained all necessary rights for implementing and distributing software under "GPLv3 or later") is not transparent (and therefore of no relevance) to any external potential contributor. Clarification of the software architecture concerning which subset of the software implements TIFF/IT and which subset of the software implements TIFF/EP is essential for achieving code transparency with respect to the tendered file formats. Further, if the supplier aims to also implement TI/A, it is also essential to clarify which subset of the software implements TI/A for achieving code transparency.

Concerning the work performed by MediaArea, we provide the following recommendations in order to provide long-term sustainability of the open source project. First, from our analysis of the source code provided by MediaArea on the OSP we note that there are functions implementing the file format MPEG-4 and also other file formats (e.g. JPEG 2000) not explicitly requested in the PREFORMA tender. For this reason, it is critical (for community and legal reasons) to clarify with respect to any potential external contributor (i.e. beyond the PREFORMA consortium) that the supplier has obtained all necessary rights (including all necessary patent licenses) for all implemented file formats. Overall, it is essential to extend the information concerning these issues to the broader community (including the road-map provided by the supplier which aims at external contributors). For attracting external contributors to the open source project it is important to clarify the conditions[96] for use of these file formats especially since these file formats are not open standards[97] and many organisations have declared that they control standards essential patents related to these file formats. The fact that the supplier has ensured (as part of the format contract) the PREFORMA consortium that it has obtained all necessary rights for the work in PREFORMA (which includes that the supplier has

[96] There are a number of important reasons for this. For example, there are a number of patent declarations made to ISO from several organisations for (several parts of) the MPEG-4 file format standard (ISO/IEC 14496). Therefore, it is essential to provide clarity on this issue for use of software which implements the MPEG-4 file format. Further, in case a memory institution wishes to deploy a software component for checking conformance with only open file formats implemented in the software, it is essential that the software architecture easily allows this and for this reason code transparency is of paramount importance for developed software. Further, it is important to recognise that many members of broader open source communities are extremely sensitive to patent related issues.

[97] The requirements for the PREFORMA tender (see "Deliverable D2.1 Overall Roadmap" and "Deliverable D2.2 Tender Specifications") define an open standard as follows: "The standard is adopted and will be maintained by a not-for-profit organisation, and its ongoing development occurs on the basis of an open decision-making procedure available to all interested parties (consensus or majority decision etc.).; The standard has been published and the standard specification document is available either freely or at a nominal charge. It must be permissible to all to copy, distribute and use it for no fee or at a nominal fee.; The intellectual property - i.e. patents possibly present - of (parts of) the standard is made irrevocably available on a royalty-free basis.; There are no constraints on the re-use of the standard." This definition is also used by the Swedish Governmental organisation "Statens inköpscentral vid Kammarkollegiet" which is responsible for establishing framework agreements for public sector procurement when expressing requirements for which standards may be referenced in procurement.

obtained all necessary rights for implementing and distributing software under “GPLv3 or later”) is not transparent (and therefore of no relevance) to any external potential contributor. For the continued work it is critical that the supplier provides clarity on these issues and details precisely which file formats and parts thereof are included in the software provided on the OSP.

Second, from our analysis we note that MediaArea refers to Debian in their feedback provided to the PREFORMA consortium on 14 December 2015 and 8 January 2016. In this feedback it is mentioned that MediaConch has been reviewed by Debian maintainers and it is mentioned[98] that it “is being accepted in the official Debian repository”. However, it should be noted that Debian licensing requirements are different from the PREFORMA licensing requirements as expressed in the DoW, deliverables D2.2, D2.3, the tender, D4.3, and the tender (in the second round). Consequently, having MediaConch in the Debian repository is not relevant with respect to fulfilment of the PREFORMA requirements. From our analysis, we observe that MediaConch has not yet been provided on the open source platform under the “GPLv3 or later” and “MPLv2 or later” licenses as required by PREFORMA.

To achieve long term sustainable open source projects, it is essential for any potential external contributor to be convinced that any contribution is well received and can be contributed without any legal and licensing issues. For example, it is important to recognise that any agreement between rights-holders, suppliers and the PREFORMA consortium is of limited (or no[99]) value (especially when such are not publicly disclosed). Therefore, clarifying licensing conditions and information that the supplier has obtained all necessary rights is especially important concerning software which handle synthetic test files as such test files have deliberately been designed to deviate from the technical specification of such file formats. This is fundamentally important for file formats for which it is known that organisations have declared standard essential patents which are necessarily infringed when such a file format is implemented in software. One example of such a file format for which such details are necessary to communicate in the road-map is ISO 32000-1. Lack of such details may significantly inhibit contributions from external contributors as many community members are very sensitive with respect to unclear conditions concerning potential patent infringements and the potential need for obtaining patent licenses.

[98] See “Supplier Updated Response to Feedback on the final release - Oct 2015” dated 8 January 2016.

[99] What is important in this respect is the perception of conditions for contributing amongst potential contributors.

4 CONCLUSION AND FUTURE OUTLOOK

4.1 OVERVIEW

This deliverable reports on monitoring of the Open Source Project implementations. Based on development efforts for each supplier, this deliverable provides feedback on their use of: an open work practice for development; frequent open releases; and promotion activities aiming towards a sustainable community.

The monitoring is focused on assessment of the extent to which suppliers address establishment of sustainable communities. Specifically, an evaluation is presented of how each open source project implementation adheres to requirements expressed in deliverable D4.3. In so doing, the deliverable provides an evaluation of the extent to which best practices from community driven open source projects have been adopted with adherence to full transparency for all digital assets.

Besides an assessment of achievements made so far, outcomes from assessment reported in this deliverable may also provide valuable guidance for suppliers in their efforts towards establishing long-term sustainable open source communities.

4.2 SUMMARY OF RECOMMENDATIONS

This deliverable contains several observations and recommendations based on our assessment of the work performed so far which require further attention amongst suppliers. In order to fulfil PREFORMA requirements and successfully address the PREFORMA R&D challenge it is critical that suppliers address all issues identified and recommendations, and we consider the following issues to be of particular short-term importance.

First, PREFORMA requires that a supplier provides the complete source code (i.e. a single zip-file containing all necessary files) under the two specific PREFORMA licenses ("MPLv2 or later" and "GPLv3 or later") on the open source portal. This is critical for provision of software according to standard practice in public sector procurement of open source software and clarity concerning provision of software (through distribution of software via the OSP) to memory institutions is of uttermost importance for successful community development and delivery of software to users. At time of writing, no supplier adheres to this requirement. However, all suppliers have communicated that they are committed to fulfil this PREFORMA requirement and we are confident that they will do so.

Second, PREFORMA requires that the supplier provides, on a monthly basis, releases which have been exposed to a certain level of QA. As we are unable to produce an executable from the source code provided on the OSP by use of an open source licensed build environment this issue needs further attention amongst suppliers. Further, as only one supplier has provided software on the OSP on a monthly basis (and the other two have done so for 4 of the 6 months) which has been exposed to a certain level of QA, there is scope for improvements and all suppliers need to devote increased attention to this PREFORMA requirement.

Third, PREFORMA requires that an executable shall be provided for each platform. Based on observations from the Open Source Portal and websites provided by each supplier, there is a need to ensure that software is provided under the PREFORMA licenses (i.e. “MPLv2 or later” and “GPLv3 or later”) on all sites.

Fourth (at time of writing), software cannot be used via any major web browser as suppliers have not yet provided such.

Fifth, PREFORMA requires that a supplier provides an up-to-date roadmap for the different versions of the software, targeted at external contributors, on the development platforms. This is an issue which needs considerable attention from now on since the current content in the roadmap is focused on PREFORMA partners instead of external contributors. This is a critical issue for the success of the PREFORMA R&D challenge as its current content lacks essential information for promoting external contributions.

Sixth, for reasons of community development and long-term sustainability of open source projects it is essential to improve code transparency and clarity concerning how specific subsets of software can be reused under the PREFORMA licenses. For memory institutions and other users it is essential that the specific subset of the software which implements a specific file format can be easily reused and distributed under the PREFORMA licenses for use in other applications and organisations. For example, an organisation may want to reuse only the specific subset of the software which implements the open file format PDF/A-1 without having to incorporate any specific implementation of PDF/A-2 or PDF/A-3.

Seventh, to promote external code contributions it is essential that all suppliers increase their attention to this issue. For example, we recommend that suppliers provides ‘easy hacks’ and provides increased clarity concerning interpretation of file formats in the code. This relates to requirements (as detailed in D4.3) for handing of synthetic test files directed to the broader external open source community beyond the life-cycle of PREFORMA (i.e. beyond specific to PREFORMA initiatives for handling of test files via cloud storage). Further, we also recommend that suppliers provide the complete source code (with test files, and associated digital assets), build environments, and executables via Live CDs/DVDs in order to promote distribution and use of the software without a need for installation.

Eighth, for the continued work it is essential that suppliers increase their attention for attracting external code contributions as well as other types of contributions from the broader open source communities and potential business partners. This is key for development of long-term sustainable open source projects beyond the PREFORMA project.