

DELIVERABLE

Project Acronym: PREFORMA

Grant Agreement number: 619568

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

D3.4 – Open Source Workshop

Revision: version 1.0 – final

Authors:

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

Contributors:

**Jonas Gamalielsson (University of Skövde)
 Björn Lundell (University of Skövde)
 Claudio Prandoni (Promoter)**

Reviewers:

**Stefan Rohde-Enslin (SPK)
 Nicola Ferro (University of Padua)
 Ruth Montague (LGMA)**

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	2016-04-20	J. Gamalielsson & B. Lundell	HS	Initial draft
0.2	2016-07-08	J. Gamalielsson & B. Lundell C. Prandoni	HS PROMOTER	Revised draft after peer review
0.3	2016-07-12	R. Montague	LGMA	Revised draft for English
1.0	2016-07-25	C. 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This deliverable contains original unpublished work except where clearly indicated otherwise.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	4
1 INTRODUCTION	5
2 ON PREPARATION AND CONDUCT OF THE WORKSHOP	6
2.1 ON PREPARATION OF THE WORKSHOP	6
2.2 ON CONDUCT OF THE WORKSHOP.....	8
3 REPORT FROM THE MORNING SESSION.....	11
3.1 OPENING THE PREFORMA WORKSHOP	11
3.2 AN OVERVIEW OF THE PREFORMA PROJECT	11
3.3 KEYNOTE BY PETER BUBESTINGER.....	13
3.4 KEYNOTE BY TILL JAEGER.....	17
3.5 INVITED PRESENTATION BY MELANIE IMMING	21
4 REPORT FROM THE AFTERNOON SESSION.....	23
4.1 PRESENTATION BY VERAPDF	23
4.2 PRESENTATION BY EASYINNOVA	25
4.3 PRESENTATION BY MEDIAAREA.....	27
4.4 EXHIBITION AND NETWORKING	28
5 SUMMARY.....	32
APPENDIX A.....	33
APPENDIX B.....	34

EXECUTIVE SUMMARY

Deliverable D3.4 reports on the Open Source Workshop that was based on the availability of the first prototype on the Open Source Portal and the results of the first demonstration organised by 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 for the full day workshop was to devote the morning session to presentations and the afternoon session to interaction and discussion amongst workshop participants. To address the fundamental community aspects of open source in the archival domain and in memory institutions involved in (or planning) digital preservation initiatives, the PREFORMA team invited Peter Bubestinger as a keynote speaker. To address the licensing aspects of open source, the PREFORMA team invited Dr. Till Jaeger as a keynote speaker.

The aim for 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. The aim for the afternoon session was to report on development efforts undertaken by PREFORMA suppliers, with highlights on open source tools being developed, and to offer suppliers an opportunity for exhibiting their tools to workshop participants. A further goal during the afternoon was to further stimulate interaction and dialogue between suppliers' representatives and other workshop participants.

The workshop facilitated a unique opportunity for attendees to raise different issues and challenges of specific interest with suppliers' representatives, PREFORMA partners' representatives, the invited speakers, and other workshop participants. During discussions, several attendees established valuable contacts and networking amongst participants was highly appreciated. For the suppliers, the interactive networking session gave many opportunities for disseminating and communicating their efforts to the broader communities, including potential adopters of open source software developed in PREFORMA.

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 to work on one of the main challenges memory institutions are facing nowadays: the long-term preservation of digital data. The project develops three open source conformance checkers that control whether a file complies with standard specifications and with the acceptance criteria of the memory institution. They offer full control of the process of conformity testing of files to be created, migrated and ingested into archives.

The objective of deliverable D3.4 is to provide a report about the Open Source Workshop that was based on the availability of the first prototype on the Open Source Portal and the results of the first demonstration organised by the suppliers.

The Open Source Workshop highlighted important aspects of open source development in relation to the key challenges addressed by the PREFORMA project. The workshop offered an internationally recognised and well received high-profile event focused on how open source solutions can effectively address these challenges which is of major concern for memory institutions in their mission to future-proof longevity of files. This PREFORMA event provided an important opportunity for promotion and discussion of development efforts that each supplier had undertaken in relation to each open source project. The workshop was organised by the PREFORMA project on 7 April 2016 in Stockholm and was hosted by Kungliga Biblioteket.

The rest of this document is organised as follows. First, an overview of the preparation and conduct of the workshop is provided (section 2). Thereafter, the document provides a report from the presentations during the morning session (section 3) and the interactive afternoon session (section 4). Finally, the document provides a summary of the open source workshop (section 5).

2 ON PREPARATION AND CONDUCT OF THE WORKSHOP

2.1 ON PREPARATION OF THE WORKSHOP

An important goal for the workshop was to provide a high-profile event focused on open source aspects of the PREFORMA project and to organise an event during which significant development efforts undertaken by PREFORMA suppliers could be recognised and exposed for the broader communities. To achieve this goal, the PREFORMA team recognised the importance of highlighting two specific key challenges for successful open source development in the context of digital archives, which were considered important to address during the event: *fundamental community aspects of open source in the archival domain and in preserving institutions* and *licensing aspects of open source*. This was addressed by inviting two internationally recognised keynote speakers.

To address the fundamental community aspects of open source in the archival domain and in memory institutions involved in (or planning) digital preservation initiatives, the PREFORMA team invited Peter Bubestinger. Mr Bubestinger is an internationally recognised expert on open source and open file formats, who has extensive experience from involvement in and deployment of open source projects in organisations in the digital archiving domain.

To address the licensing aspects of open source, the PREFORMA team invited Dr. Till Jaeger. Dr. Jaeger is an internationally recognised world-leading expert on licensing of open source software. He is a certified copyright and media law attorney who has extensive experience and know-how from involvement with clients from many organisations and open source projects concerning copyright and licensing issues, including involvement in lawsuits to enforce the General Public License (GPL) in court.

The content of the open source workshop was organised in two main sessions. The aim for 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. The aim for the afternoon session was to report on development efforts undertaken by PREFORMA suppliers with highlights on open source tools being developed and to offer suppliers an opportunity to exhibit their tools to workshop participants. A further goal during the afternoon was to further stimulate interaction and dialogue between suppliers' representatives suppliers and other workshop participants.

As well as providing the PREFORMA suppliers with an opportunity to communicate their development efforts, it was agreed to present a brief overview of the PREFORMA project and related aspects of open source for scientific communities which provided a basis for interactive sessions and dialogue between participants.

To maximise attendance amongst representatives for memory institutions and other stakeholders in the broader interest sphere for open source solutions, it was agreed amongst project partners that the historical buildings the Swedish National Library would constitute an ideal venue for hosting such a high-profile event as the PREFORMA open source workshop. Further, the selected venue is well connected with a number of governmental agencies and other organisations represented in Stockholm, which increases the likelihood of a well-attended workshop by potential participants based in other European locations. When preparing the programme for the workshop, PREFORMA partners aimed to attract a mix of attendees representing several different stakeholder groups, including potential users and decision makers

that may consider solutions developed in PREFORMA for potential organisational adoption. Another important target group was developers and companies which potentially would be interested in collaboration in open source projects developed in PREFORMA.

As part of preparing and promoting the workshop, PREFORMA developed promotional material which partners used for reaching out to potential attendees in different countries. Through already established networks, PREFORMA representatives also actively promoted the workshop. Social media and the custom made website for the workshop were also actively used in this promotion.

The number of registrations for the workshop exceeded expectations and a week before the event, the number of registrations went beyond the number of available seats (100) in the auditorium at the venue. However, it was agreed to allow a few additional registrations beyond the number of available seats since it was expected that a few registered individuals would most likely not attend the event due to unforeseen circumstances.

On 6 April PREFORMA partners met at a face-to-face preparation meeting in Stockholm in order to finalise details concerning the workshop (Figure 1). As part of this preparation meeting, representatives for the PREFORMA project and representatives for all suppliers scrutinised draft presentations for the three open source projects. Key aspects and messages in each draft presentation were discussed and each supplier received feedback from PREFORMA representatives in order for suppliers to polish and finalise their presentation for the workshop (Figure 2). The preparation meeting with the suppliers also included opportunities for detailed discussions between PERFORMA representatives and suppliers' representatives (Figure 3).



Figure 1. Discussions amongst PREFORMA partners during the preparation meeting



Figure 2. Draft presentations discussed with suppliers during the preparation meeting



Figure 3. PREFORMA technical coordinator and MediaArea representative in discussions

2.2 ON CONDUCT OF THE WORKSHOP

The workshop was organised by the PREFORMA project on 7 April 2016 in Stockholm and was hosted by Kungliga biblioteket. The overall structure for the full day workshop was to devote the morning session to presentations and the afternoon session to interaction and discussion amongst workshop participants. The morning session was organised in an auditorium at Humlegården, which is the main building of Kungliga biblioteket (Figure 4). For the afternoon session, the workshop was held in a room suitable for interactive discussion in an open space (the Sibelius Hall located in “Finlandshuset”, a building in the vicinity of Humlegården).



Figure 4. Kungliga biblioteket at Humlegården

The morning session gave an overview of the PREFORMA project and highlighted some of its key open source issues through presentations in an auditorium (Figure 5). Specifically, the morning session included the two keynote presentations. The afternoon session was initiated by presentations of PREFORMA solutions from the suppliers. Thereafter, discussions and interaction continued in an open space amongst workshop participants and the suppliers. The three suppliers exhibited and demonstrated their open source solutions, and there were plenty of opportunities for in-depth discussions.



Figure 5. Participants arriving to the auditorium before the start of the morning session

All presenters from the morning session, including the two keynote speakers, joined the afternoon session and participated in discussions. The session facilitated a unique opportunity for attendees to raise different issues and challenges, that were of their specific interests, with suppliers' representatives, PREFORMA partners' representatives, the invited speakers, and other workshop participants. During these discussions, several attendees established valuable contacts and networking amongst participants was highly appreciated. For the suppliers, the interactive networking session gave many opportunities for disseminating and communicating their efforts to the broader communities, including potential adopters of open source software developed in PREFORMA.

The sessions during the workshop attracted many (77) attendees from several countries who obtained a set of high-profile presentations on core issues and challenges addressed in PREFORMA, and during the workshop engaged in networking opportunities and in-depth discussions. The popularity of the event was indicated by the fact that many attendees arrived early as shown by an early crowded auditorium before the start of the event (Figure 6).



Figure 6. The auditorium before the start of the morning session

3 REPORT FROM THE MORNING SESSION

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

3.1 OPENING THE PREFORMA WORKSHOP

Lars Ishammar, Deputy National Librarian, welcomed the attendees (Figure 7). He addressed openness in various forms and the connection to memory institutions, and highlighted the importance of PREFORMA and open source software.



Figure 7. Welcome address by the deputy national librarian

3.2 AN OVERVIEW OF THE PREFORMA PROJECT

Börje Justrell, coordinator of PREFORMA, thanked the former speaker and welcomed the audience to the session (Figure 8). The presenter introduced the agenda for the day and specifically mentioned Bengt Neiss of the Swedish National library who was a key person in the local arrangement of the event. PREFORMA is an EU FP7 pre-commercial procurement project which started on 1 January 2014 and has a duration of 48 months and a total budget for the procurement of 2.805.000 EUR. The project is coordinated by Riksarkivet in Sweden with

technical and communication support from the Italian partner Promoter SRL. Further, there are four technical partners and eight memory institutions which are engaged in the project. The presenter explained that memory institutions are facing increasing transfers of electronic documents and other media content for long-term preservation. Data is normally stored in specific file formats for different media types (documents, images, sound, video etc.), and these files are usually produced by software from different vendors. Even if the transferred files are in standard formats, the correct implementation of standards cannot be guaranteed. The presenter argued that this poses problems for long-term preservation, and that the aim of PREFORMA therefore is to develop good quality implementations of selected file formats for preservation of digital content in the long term. This will give memory institutions full control of the process of performing conformity tests of files to be ingested into archives.

PREFORMA adopts an open source approach with the aim of establishing a sustainable community comprising a wide range of contributors and users from different stakeholder groups. The open source nature ensures long-term availability of the software, beyond the memory institutions and suppliers involved in PREFORMA. The developed software will be licensed under the two specific open source licenses “GPLv3 or later” and “MPLv2 or later”. The presenter stated that the PREFORMA challenge involves development of an open source conformance checker that checks if a file complies with standard specifications, checks if a file complies with the acceptance criteria of the memory institution, reports back to human and software agents, and performs simple fixes. The challenge also involves establishment of ecosystems around the open source implementation that generate useful feedback for those who control the software, advances improvement of standard specifications, and advances development of new business cases for managing preservation files. The presenter explained that target users and stakeholders of PREFORMA software are memory institutions, developers, research organisations, standardisation bodies, funding agencies, and other organisations planning for long-term preservation of digital content. The PREFORMA project comprises three main phases (design, prototyping, and testing), and the three suppliers are currently active in the prototyping phase: veraPDF consortium (implementing the text format PDF/A), EasyInnova (implementing the image format TIFF), and MediaArea (Implementing the A/V format combination MKV|FFV1|LPCM). Further, PREFORMA open source projects are provided via three channels: an open collaboration platform (GitHub) where all code is maintained and openly available; the Open Source Portal where stable (monthly) releases of source code, executables, and build environment are available for easy download and use under the PREFORMA licenses; and via each supplier’s own channels. Finally, the speaker announced the forthcoming experience workshop will take place in Berlin in November 2016, in connection with the second prototype demonstration, and the final conference will take place in Stockholm in the end of 2017 to present the results of the project.

For further details concerning the presentation “PREservation FORMAts for culture information/e-archives” by Börje Justrell, please see Appendix A.



Figure 8. The coordinator of PREFORMA presents an overview of the project

3.3 KEYNOTE BY PETER BUBESTINGER

Antonella Fresa, the technical coordinator of PREFORMA thanked Börje Justrell, the project coordinator of PREFORMA, and attendees before introducing the first keynote speaker, Peter Bubestinger who is an internationally recognised digital preservation expert who has specialised in open file formats and open source software.

The overall message from the first keynote speech, by Peter Bubestinger, addressed fundamental community aspects of open source in the archival domain and in memory institutions involved in (or planning) digital preservation initiatives. Mr Bubestinger elaborated on various conceptions (and misconceptions) related to core concepts of free and open source software and drew from his experience of involvement in specific projects which have implemented various file formats of relevance for longevity of files in the digital archiving domain. As part of this, he elaborated on his own experience of challenges related to implementation of file formats in systems that deliver professional quality and illustrated his experiences with a real-world example from the audio-visual media domain.

Peter Bubestinger started his presentation with an elaboration of the core concept of 'open source' and asked himself: 'What actually is open source?' In elaborating on the concept he shared details from his own background and experience as a programmer, starting to learn how to program at the age of 11. Later, during his studies at a technical university, he was still unexposed to open source and its inherent licensing model which emphasises sharing. From his experience of having been exposed to an overarching message which had amplified how to lock people down and to disallow people to do certain things with software, he shared with the

audience how he, as a passionate developer, was stimulated to learn more and address tricky things and technical challenges related to software. While studying media and computer science, he continued, he needed a job and became involved with a company that developed systems for digital archiving that worked with mass digitisation of content for broadcast. He revealed that development in this context was not open source and from this experience he gained insights concerning opportunities and challenges, with several different concepts about open source.

Amongst the misconceptions elaborated it was clarified that open source is not the same as freeware ('freeware' implies no-cost, whereas 'open source' does not). Further, it was clarified that open source is often mistakenly perceived as non-commercial, sometimes even amongst representative of companies that produce open source software. From this, he clarified that open source is about user freedom and elaborated on the 'four freedoms' (as defined by the Free Software Foundation, i.e. the freedom to use, study, share and improve the software¹). As part of this, the lack of freedom was illustrated by a number of personal experiences from the use of specific closed devices (a digital camera and a remote control device for a VCR). It was clarified that the concept of freedom in this context refers to 'freedom as in free speech' and the term 'free software' was clarified with reference to the four freedoms. Based on an illustrative example from the world of furniture, it was stressed that it is common to meet these four freedoms in the non-software world, whereas in the software-world it is an exception that we have them.

The presentation continued with a coverage of long-term preservation, starting with an illumination of what is meant by 'long-term'. As part of this, he posed the question to the audience: Is it 10 years, is it 100 years, or is it longer? In elaborating on this question, he referred to an experience where a question was posed by an attendee at another conference on long-term preservation who asked 'What do you consider long-term?' and revealed that the response was 3-5 years. In clarifying why software was only available for 3-5 years, it was explained that this is the typical market lifetime of software. In contrast, he commented that when we speak about archiving the requirement is really infinity and to be sure, we add one day. He continued to explain the concept of virtual immortality and related this to recording devices. In so doing, he clarified the importance of making it work in future conditions as software and conditions for execution of software changes.

The presentation continued by posing an open question on why open source is not more widely used in the archival domain and in memory institutions involved in (or planning) digital preservation initiatives, elaborating on perceived professionalism and quality. From this it was

¹ The four Freedoms (from Free Software Foundation)

1. USE it for any purpose
2. STUDY how the program works and understand it
3. SHARE copies of the software
4. IMPROVE the program and distribute these changes

commented that a commonly held assumption is that often perceived quality is positively correlated to price. Instead, it was commented that quality of software is independent of price and independent of license under which the software is provided. Issues of support and management of risk were covered, and it was mentioned that some organisations do not want to use anything different to others. Some examples of organisations using open source in the media domain was presented and it was commented that the BBC is one of the organisations that has published software under open source licenses, and that they have said: 'If we get public funding we should put it into public solutions if possible.' Other examples of open source covered included GlusterFS (a network file system that can be used for distributed storage solutions for media streaming, initially developed by Gluster Inc. and later by RedHat Inc., and used for Hollywood movies) and OpenEXR (an image file format initially developed by Industrial Light & Magic and used for film productions in Hollywood).

The issue of independence vs. dependence for the domain was addressed and it was mentioned that the archival domain is a relatively small (niche) market with highly specialised demands. Hence, it was commented, there is high potential for vendor dependence. The presentation elaborated on an example of faulty software that produced incorrect video conversion in a specific device provided by a vendor to a memory institution. As the device was used for other needs (than preservation), the vendor was not interested to solve the problem with the faulty software. From this experience, a lesson learnt is that there is a high risk of vendor dependence when using off-the-shelf components for a non-primary use-case.

Approaches for using open source software to address issues of dependence were elaborated upon. Aspects covered included the extent to which it is possible to reuse existing solutions and seek support from commercial suppliers. Further, the value of transparency of code provided by open source implementations of file formats was elaborated upon. It was stressed that if you find a problem you can make someone to fix it as there are no 'black-boxes' with open source. Also, even if the archival domain is a niche domain for large parts of a solution, several commonly used tools and widely deployed code is typically adopted which often has a larger user base, something which contributes to increased exposure of errors in the codebase. Finally, he commented on the benefit of maintaining control concerning when to upgrade as a contrast to vendor-forced upgrades which are common for proprietary software.

The presentation continued with an elaboration of professionalising with open source which includes paying for open source. As part of this it was commented that the more users there are for a specific piece of software, the more people there will be that are likely to find any potential errors in it (so that it promotes quality). Further, when several user organisations pool resources and agree to pay for development by sharing cost for the development of open source, the cost for an individual organisation is decreased and yet all users will have access to what has been developed. This implies that all improvements and changes made locally in a single organisation will be made available to the main project ('upstream'). The benefits (compared to proprietary software) of having the possibility to adapt open source software to the specific needs of an individual organisation were also mentioned. Further, the issue of justifying the need for sharing developed software that has been paid for by public (tax-payer) money was commented on. Thereafter he elaborated on how individuals and organisations can contribute to an increased professionalising of open source, and covered organisations that can govern and support use of open source (such as Free Software Foundation Europe).

Finally, a real-world example was presented to illustrate how open source can be used in the digital archiving domain. The illustration covered how the lossless video codec FFV1 has been

used in the Austrian National Archive (Mediathek) for digital video archiving. The presentation commented on the consequences for Mediathek of the fork in the open source project used (FFmpeg was forked to LibAV) and it was mentioned that the changes to the code base that the work at Mediathek undertook are now contributed and maintained in both open source projects. As a final comment, it was stressed that 'we archives have common interests and common challenges, so why not go for common solutions'.

For details of the presentation "Free Software and Open Formats: Virtual immortality and independence for digital archives" by Peter Bubestinger, please see Appendix A.



Figure 9. Peter Bubestinger during his keynote address



Figure 10. Peter Bubestinger during Q&A in connection with his keynote address

3.4 KEYNOTE BY TILL JAEGER

Antonella Fresa, the technical coordinator of PREFORMA introduced the second keynote speaker, Dr. Till Jaeger who is an internationally recognised world-leading expert on the licensing of free and open source software and certified copyright, and a media law attorney with extensive experience from involvement with clients in lawsuits to enforce the GPL in court.

The overall message from the second keynote speech, by Dr. Jaeger, conveyed insights and an in-depth exposure to challenges and solutions concerning free and open source license compatibility. Core concepts, such as derivative works and principles for copyleft and non-copyleft free and open source licenses were elaborated upon including implications in different situations. Several illustrations and examples of situations brought to the attention of the attendees.

Dr. Jaeger started the presentation with a reference to the four freedoms and commented that these are based in the licenses and stressed that it is about copyright law. So, more or less, it was stressed that you have the same grant of rights for all open source licenses, but that you have to deal with the differences concerning different license obligations. It was explained that infringing a license obligation is a copyright violation, and we know that copyright law is very strict so this is something that is not just 'easy doing'. Hence, we have the broad grant of rights (we have the four freedoms) but we also have rules.

The presentation continued with an elaboration of the history of free and open source licenses and stressed that some licenses were written more than 20 years ago, when the practical situation was rather different. It was commented that, at the time, it was typical that we had one single project and one single license that fit the project. Dr. Jaeger mentioned that nowadays software development is much more modular and there is much more reuse of code and modules from other projects, sometimes from other licenses. In this scenario, it was commented, it is essential to know about the pitfalls and also the solutions. It was stressed that the starting point to understand how to make modifications is in copyright law,. Dr. Jaeger explained that from a copyright perspective, if you want to modify code or the program you need the permission of the original author as this is just a fact of copyright. It was commented that this is precisely what open source licenses are providing, they are providing the right to modify.

But there is a dependency on this right from the original author. It was further explained that – with the creation of a modified work – the owner of the copyright of the derivative work is depending on the copyright of the original work. From this it was commented that, for each and every new modification of the work, there is a new copyright on the derivative work so that all the persons are depending from each other and this dependency is something that is in the license.

Dr. Jaeger continued by stressing that the main issue in this context is the term copyleft, and posed the question: 'what is copyleft?' It was explained that copyleft is something that was invented by Richard Stallman and that this is a tremendous 'legal hack' to understand this concept, which is the basis for nearly all important questions on free and open source software law. It was commented that to understand this concept one has to understand the legal situation in the 1980s when this concept was invented, since at that time software was not a product; it was something that you received with your hardware. At that time, the source code was shipped with the hardware and it was possible to modify the code for the specific purposes of the recipient. The evolution of the software industry was elaborated upon and it was commented that some companies realised that it was possible to have the software protected as a product independently of the hardware. It was explained that these companies, on the one hand put pressure on lawmakers for copyright protection of software, and on the other hand these companies did not provide the source code any more (so just binaries) which made it impossible for the recipient of the software to modify the software. The presentation continued with the history of how Stallman started to develop free software (and the GNU-project) based on the idea that all software should be free so that anyone can use, modify, and share the source code. The legal challenge that Stallman addressed was to ensure that not only all the code I write will be available to anyone (because this can be easily controlled), but also that all the modifications and all the code written by third-parties will also be available to anyone. It was commented that Stallman used copyright (which introduced all the problems) to do it 'vice versa' and this is why it is called copyleft (instead of copyright). In this context copyright is not used to collect royalties, but used to establish a system that ensures free access to the software. Dr. Jaeger explained that a clause in the license states that if software is distributed under this copyleft license you have all the rights as in the four freedoms and also that if you distribute a modified work of this software it has to be distributed under the very same license and also provide the source code for it.

From this background on copyleft, the presentation continued with an elaboration of the problems related to copyleft. If we have one license and one program, this is not so much a problem. It was explained that there are two categories of copyleft licenses, strong copyleft licenses (GPL – GNU General Public License, versions 2 and 3; AGPL – GNU Affero General Public License; EPL – Eclipse Public License, EUPL – European Union Public License) and weak copyleft licenses (LGPL – GNU Lesser General Public License, versions 2, 2.1 and 3; MPL – Mozilla Public License, versions 1, 1.1 and 2). Dr. Jaeger elaborated upon the implications for strong copyleft licenses, stating that for strong copyleft licenses – all modifications have to be licensed under the original license if distributed, and for weak copyleft licenses – some modifications have to be licensed under the original license if distributed. The presentation continued with an elaboration upon the case of Netscape which triggered the development of the weak copyleft license MPL which has copyleft obligations restricted to the file that contains the MPL-licensed code. It was further explained that for non-copyleft licenses (also referred to as permissive licenses) it is the case that derivative works may be licensed under different license conditions, even proprietary.

The presentation continued with a coverage of the impact of copyleft and it was clarified that copyleft is relevant for licensing own developments and modifications to software, and also for compatibility with other open source software components as well as remixing code. It was stressed that it is not the case that you can just mix code from any source and that it is an open question as to precisely what is a derivative work (as there is currently no case law as of today that involve GPL interpretation, even though there is currently an ongoing case). It was explained that the reason why there is a lack of explanation in current copyright law is the fact that in the past this did not matter (as there is no room for dispute when a company licenses the entire code based under a specific license, so there has been no room for dispute). Dr. Jaeger continued by stating that in the world of free and open source software where there are licenses that refer to derivative works, it is an issue and this has been an on-going issue of discussion (what is, and what is not, a derivative work) amongst legal experts involved with free and open source software for several years.

Two cases were elaborated upon concerning license compatibility. The first case concerns mixing code (i.e. putting code into one file, deleting and adding code in that file) and the second case concerns combining components (modules, plugins and libraries), something which is becoming more and more important and also raising more and more legal questions, it was argued. The presentation continued with explanations of examples of combining code, which included citing examples concerning both static linking (with a scenario in which the library is combined during compilation) and dynamic linking (with a scenario in which the library is loaded during execution). It was commented that it is heavily disputed if this example of dynamic linking is creating a derivative work.

The presentation continued with an elaboration upon how to license a derivative work and it was stressed that one has to choose a license (i.e. it is not possible to use more than one license, except in the specific case of dual-licensing). Hence, it was explained that license compatibility is a question of the license that you choose for your derivative work, is a license that is allowed by all the code that your program is based on. Dr. Jaeger continued to elaborate upon issues to be addressed in legal compliance work, and commented that the first question is if two components to be considered is a derivative work, and if it is not a derivative work (i.e. if the two components are independent programs) then license compatibility is a non-issue since each program has its own license. It was further explained that if it is a derivative work (i.e. if the copyleft is triggered) then you have to check compatibility.

Thereafter, three different scenarios of combining code were presented, namely: a) combining non-copyleft and non-copyleft; b) combining copyleft and copyleft, and c) combining copyleft and non-copyleft. It was stressed that the easy case is if there is no copyleft, no license requires to license derivative works under a specific license, so you are always able to combine derivative work under different non-copyleft licenses, and in this scenario you are just adding all the license obligations. It was commented that the Apache license requires an indemnity clause and that the BSD-license requires that the name of the copyright owner is provided. Dr. Jaeger continued to elaborate on the scenario of combining two different components that are both under copyleft licenses, and commented that since a copyleft license requires that the derivative work is licensed under the same license, there will be a clash since this implies that the two different copyleft licenses will require two different things (with the consequence of not being able to fulfil both obligations, which implies that this scenario will breach one license in the case the other license is fulfilled). It was further commented that some licenses have an explicit compatibility clause which constitutes an exception from this (and an example being the compatibility clause in LGPL 2.1 which makes it possible to combine with code under GPL

version 2). Otherwise, if combining code that is not compatible it would imply that there is a copyright infringement. The presentation continued to elaborate on the scenario in which components under copyleft and non-copyleft are combined, and it was commented that this may be more complex since the copyleft license implies that the code has to be licensed under the very same license conditions. It was explained that this does not allow that any additional license obligations is added. From this, it was explained, that before combining components, one must ensure that there is no obligation in the non-copyleft license that does not already exist in the obligations of the copyleft license. Dr. Jaeger illustrated this complexity with the example of considering combining a code under the 4-Clause BSD license with code under GPL version 2.0 (something which is not allowed due to the, so called, advertising clause in the 4-Clause BSD license). It was commented, that since this obligation does not exist in the GPL these two licenses are not compatible and therefore it is not possible to use code licensed under the 4-Clause BSD license in the GPL code.

Finally, Dr. Jaeger elaborated upon compatibility issues between different licenses and commented on the rationale for certain projects to select specific licenses and versions thereof.

For details of the presentation “The Relevance of FOSS License Compatibility” by Dr. Till Jaeger, please see Appendix A.



Figure 11. Dr. Till Jaeger during his keynote address



Figure 12. Dr. Till Jaeger during Q&A in connection with his keynote address

3.5 INVITED PRESENTATION BY MELANIE IMMING

Antonella Fresa, the technical coordinator of PREFORMA introduced the invited presentation by Melanie Imming who is involved in open science initiatives at European and International levels through the LIBER organisation.

The overall message from the invited presentation, by Melanie Imming, stressed the importance of open source software and open file formats as enablers of open science.

First, Melanie Imming (Figures 13 and 14) introduced the LIBER Europe organisation, its activities, geographical locations, conferences, and EU projects. It was explained that a central part of LIBER's mission is to provide an information infrastructure that enables research in LIBER institutions to be world class. Further, it was stated by the presenter that open science involves different aspects of openness such as open policies, open access publications, open APIs, open data, open licenses, open source, and open standards. A definition of open science was presented that emphasises the importance of openness of data and tools used to derive research results. It was stressed that there is the need to move from "publish or perish" to open science.

The science code manifesto was presented which contains five principles that can be adopted to address the challenge of making and sustaining software as a cornerstone of science. Related to the PREFORMA challenge, the presenter highlighted the curation principle and that "making sure data is stored in a controlled way and can be (re)used today and in the future is an important element in Open Science". Further it was stressed that standardisation of file formats promotes re-usability, enables processing and preservation of data in a controlled way, ensures outputs that are really open and accessible in the long term, and improves interoperability of new tools and services.

The presenter discussed possibilities for libraries to enable open science by addressing typical issues such as availability, findability, interpretability, re-usability, citability, and curation of data. It is also important for libraries to focus on research data management and to create awareness, trust and community building to support open science. Finally, the presenter encouraged involvement and engagement in activities related to open science.

For details of the presentation "Open Science, Open Data: towards a new transparent and reproducible ecosystem" by Melanie Imming, please see Appendix A.



Figure 13. Melanie Imming during her invited presentation



Figure 14. Melanie Imming during Q&A in connection with her invited presentation

4 REPORT FROM THE AFTERNOON SESSION

This section reports on the afternoon session.

Bitr. Prof. Björn Lundell chaired the afternoon session and introduced the overview presentation from each first supplier. Before the start of the session, the chair for the afternoon session discussed the format with the speakers representing suppliers (Figure 15). After each overview presentation, the audience was given an opportunity to ask clarifying questions which were moderated by the chair.



Figure 15. The chair for the afternoon session (second from left) in dialogue with speakers

4.1 PRESENTATION BY VERAPDF

The veraPDF consortium was represented by Joachim Jung, Boris Doubrov and Carl Wilson during the presentation entitled “veraPDF: definitive, open source PDF/A validation for digital preservationists”. First, Joachim Jung (Figure 16) introduced the veraPDF consortium, which is jointly lead by Open Preservation Foundation and PDF Association. The veraPDF software implements a conformance checker for the text file format PDF/A. Further, the presenter explained that the veraPDF consortium contributes to standardisation efforts related to ISO 19005 variants through the PDF Association technical working group. VeraPDF is described as being a consortium, a set of FLOSS applications, a test corpus for different variants of PDF/A, and potentially a framework for developing other format validators. The presenter argued that VeraPDF is useful for different stakeholders: digital preservation practitioners, in order to ensure that newly curated documents comply to standards and institutional policies; PDF technology developers, to be able to reduce software development and support costs; document technology developers in order to assess PDF technology options, improve reliability and performance and provide new options for products, and to reduce support costs; and customers creating and using documents who may reduce operational risks and legal liabilities, and receive more capable and interoperable software. Next, the different ISO standards, conformance levels, and variants of PDF/A were explained.

After this, 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 conformance

checker. Specifically, it was shown how a file can be tested and a report can be generated and inspected.

Carl Wilson (Figure 17) 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. Further, the speaker presented the installer and a progress report illustrating the status of different activities in the project. Finally, 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 Appendix B.



Figure 16. Representative (Joachim Jung) for VeraPDF during the presentation



Figure 17. Representative (Carl Wilson) for VeraPDF during the presentation

4.2 PRESENTATION BY EASYINNOVA

EasyInnova 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”. First, Miquel Montaner (Figure 18) 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 (Figure 19) explained that DFP manager is conformance checker for different versions of the TIFF image format. Further, the speaker stated that the software can be executed on multiple platforms in different ways, exhibits a modular architecture, and uses a number of technologies (e.g. Java, Maven, Spring, Travis CI, JavaFx, and Jacpfx). The different steps for use of the software was elaborated which involve selection of files, defining criteria for acceptance of files, management of metadata and application of fixes, and report generation. The presenter showed what the GUI looks like for the aforementioned steps in a typical use case. Different use scenarios were explained, which include standalone use, client-server, web application, and high availability mode. The integration with the OAIS model was also discussed. Further, the current and planned future features of the conformance checker, policy checker, metadata fixer were presented. It was also explained that DFP manager is an open source project released under licenses MPLv2+ and GPLv3+ which is provided on a company website, on the PREFORMA portal, and on GitHub.

Next, Peter Fornaro explained file format preservation and described important properties of file formats. The TIFF format, 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 standardisation 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. Further, the presenter stated that there are many TIFF experts involved (69 experts in 16 countries), and support from 61 memory institutions. Finally,

the consortium announced their interest in further extending collaboration with new early adopters, TIFF experts, and software developers.

For details concerning the presentation “DPF Manager – The open source Community” by EasyInnova, please see Appendix B.



Figure 18. Representative (Miquel Montaner) for EasyInnova during the presentation



Figure 19. Representative (Xavier Tarrés Bonet) for EasyInnova during the presentation

4.3 PRESENTATION BY MEDIAAREA

MediaArea was represented by Jérôme Martinez (CEO), during the presentation entitled “MediaConch – Implementation and policy checking on FFV1, Matroska, LPCM, and more”. First, Jérôme Martinez (Figure 20) 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 plugin, 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 Appendix B.



Figure 20. Representative (Jérôme Martinez) for MediaArea during the presentation

4.4 EXHIBITION AND NETWORKING

At the end of the session the chair invited all participants to coffee and subsequently introduced the exhibition and the project corners where the three different suppliers gave demonstrations of their conformance checkers. Each supplier was provided a corner in the room for the afternoon session in which they exhibited their open source solutions. The corners for VeraPDF (Figure 21), EasyInnova (Figure 22), and MediaArea (Figure 23) were all well attended during the exhibition and provided many opportunities for networking with participants interested in the specific solution exhibited. There was also a separate corner dedicated to the PREFORMA consortium (Figure 24). Overall, during the afternoon session there were plenty of opportunities for informal networking and discussions with other participants including PREFORMA partners' representatives (Figure 25).



Figure 21. The VeraPDF corner during the exhibition



Figure 22. The EasyInnova corner during the exhibition



Figure 23. The MediaArea corner during the exhibition



Figure 24. The PREFORMA consortium corner during the exhibition



Figure 25. Networking during the afternoon session

5 SUMMARY

This deliverable reports on the Open Source Workshop that was organised at Kungliga Biblioteket in Stockholm on 7 April 2016. The event featured two invited internationally recognised experts as keynote speakers focusing on: open source software and open file formats (Peter Bubestinger); open source licensing (Dr. Till Jaeger); an invited presentation (by Melanie Imming) during the morning session; and presentations from each of the three PREFORMA suppliers (VeraPDF, EasyInnova, and MediaArea) during the afternoon session. The afternoon session continued with a software exhibition and networking session featuring the software for conformance checking provided by the three suppliers.

The sessions during the workshop attracted many (77) attendees from several countries who obtained a set of high-profile presentations on core issues and challenges addressed in PREFORMA, and during the workshop engaged in networking opportunities, in-depth discussions and interaction.

APPENDIX A

This appendix contains slides used by presenters in the morning session of the Open Source Workshop.

1. Slides from the presentation “PREservation FORMAts for culture information/e-archives” by Börje Justrell
2. Slides from the keynote presentation “Free Software and Open Formats: Virtual immortality and independence for digital archives” by Peter Bubestinger
3. Slides from the keynote presentation “The Relevance of FOSS License Compatibility” by Dr. Till Jaeger
4. Slides from the invited presentation “Open Science, Open Data: towards a new transparent and reproducible ecosystem” by Melanie Imming

PREservation FORMAts for culture information/e-archives

*PREFORMA
and the Open Source Community
Borje Justrell,
Swedish National Archives
Coordinator*

PREFORMA General Presentation



Project Identity Card

- PREFORMA is a **Pre-Commercial Procurement** project co-funded by the European Commission under FP7-ICT Programme.
- Start date:** 1 January 2014
- Duration:** 48 month (end date: 31 December 2017)
- Total budget for the procurement:** 2.805.000 EUR
- Website:** www.preforma-project.eu
- Contacts**
 - Project Coordinator: Borje Justrell, Riksarkivet, borje.justrell@riksarkivet.se
 - Technical Coordinator: Antonella Fresa, Promoter Srl, fresa@promoter.it
 - Communication Coordinator: Claudio Prandoni, Promoter Srl, prandoni@promoter.it

PREFORMA General Presentation



Project Partners



- RIKSARKIVET, Sweden **Project Coordinator and memory institution**
- PROMTER SRL, Italy **Technical and Communication Coordinator**

- Technical partners**

- PACKED EXPERTISECENTRUM DIGITAAL ERFGOED VZW, Belgium
 - FRAUNHOFER-GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V., Germany
 - HOGSKOLAN I SKOVDE (University of Skovde), Sweden
 - UNIVERSITA DEGLI STUDI DI PADOVA, Italy

- Memory institutions**

- STICHTING NEDERLANDS INSTITUUT VOOR BEELD EN GELUID, Netherlands
 - Koninklijk Instituut voor het Kunstpatrimonium, Belgium
 - GREEK FILM CENTRE AE, Greece
 - LOCAL GOVERNMENT MANAGEMENT AGENCY-AN GHNIOMHAIREACHT BAINISTIOCHTA RIALTAIS ATIUIL, Ireland
 - STIFTUNG PREUSSISCHER KULTURBESITZ, Germany
 - AYUNTAMIENTO DE GIRONA, Spain
 - Eesti Vabariigi Kultuuriministeerium, Estonia
 - KUNGLIGA BIBLIOTEKET, Sweden

PREFORMA General Presentation



Project Concept



- Memory institutions are facing **increasing transfers** of electronic documents and other media content for long-term preservation.
- Data content are normally stored in **specific file formats** for documents, images, sound, video etc., and these files are usually produced by software from different vendors.
- Even if the transferred files are in standard formats, the correct **implementation of standards** cannot be guaranteed:
 - The software used for the production of the electronic files is not in control neither by the institutions that produce them nor by the memory institutions;
 - Conformance tests of transfers are done by memory institutions, but are not totally reliable; different software for testing could end up in different results.
- This poses problems in **long-term preservation**. Data objects meant for preservation, passing through an uncontrolled generative process, can jeopardise the whole preservation exercise.

PREFORMA General Presentation



Project Aim and Objectives



- The aim:** to implement good quality files in various standard formats for preserving digital content long-term.
- The main objective:** to give memory institutions full control of the process of conformity tests of files to be ingested into archives.
- The main objective of the PCP launched by PREFORMA:** to develop an *open source software* for the management of the whole conformance test process, supporting a range of standards, addressing the needs of any memory institution or other organisation with a preservation task.

PREFORMA General Presentation



Open Source approach



- PREFORMA is following an **open source approach**, with the aim of establishing a sustainable research and development community comprising a wide range of contributors and users from different stakeholder groups.
- The open source nature **ensures long-term availability of the software**, beyond the memory institutions and suppliers involved in PREFORMA.
- Licenses
 - All **software** developed during the PREFORMA project will be provided under two specific open source licenses: "GPLv3 or later" and "MPLv2 or later".
 - All **digital assets** developed during the PREFORMA project will be provided under Creative Commons CC-BY v4.0, and in open file formats.

PREFORMA General Presentation



Overall R&D Objective

(The PREFORMA Challenge)



- Develop an **open source conformance checker** that:
 - checks if a file complies with standard specifications
 - checks if a file complies with the acceptance criteria of the memory institution
 - reports back to human and software agents
 - perform simple fixes
- Establish an ecosystem around an **open source reference implementation** that:
 - generates useful feedback for those who control software
 - advances improvement of the standard specification
 - advances development of new business cases for managing preservation files

PREFORMA General Presentation



Target users and stakeholders

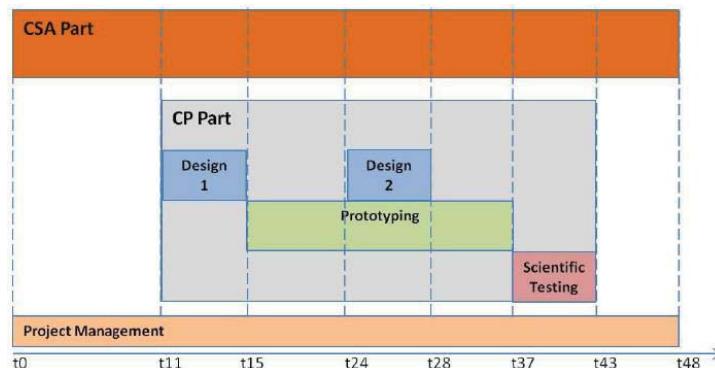


- Memory institutions** and cultural heritage organisations, involved in (or planning) digital culture initiatives.
- Developers** contributing code for the PREFORMA open source tools.
- Research organisations** providing technical advice to cultural stakeholders.
- Standardisation bodies** maintaining the technical specifications of the preservation formats covered in PREFORMA.
- Funding agencies**, such as Ministries of Culture and national/regional administrations, that own and manage digitisation programmes and may endorse the use of the PREFORMA tools in the digitisation process.
- Other **projects** in the digital cultural heritage domain.
- Any other organisation** planning for long-term preservation of digital content.

PREFORMA General Presentation



Project implementation schedule



- **Design phase** (4 months): November 2014 – February 2015
- **Prototyping phase** (22 months): March 2015 – December 2016
 - First prototypes: March 2015 – October 2015
 - Re-design: November 2015 – February 2016
 - Second prototype: March 2016 – December 2016

- **Testing phase** (6 months): January 2017 – June 2017

PREFORMA General Presentation



PREFORMA Suppliers in the Prototyping Phase



1. veraPDF Consortium (led by Open Preservation Foundation and PDF Association)
 - The PDF/A conformance checker, accepted industry-wide (PDF/A)
2. EasyInnova
 - Digital Preservation Formats Manager (TIFF)
3. MediaArea
 - PREFORMA MediaConch: CONformance CHecking for audiovisual files (MKV|FFV1|LPCM)

PREFORMA General Presentation



Requirements for Long-term Sustainability of Open Source Communities



To achieve long-term sustainable open source communities of relevance for memory institutions and other stakeholder groups, contracted organisations need to successfully manage and conduct a number of fundamental activities:

- **effective communication** of long-term vision of goals and plans for how the project will evolve over time;
- use of an **open collaboration platform** (such as GitHub) and use of open source tools with associated work practices;
- provision of **source code** under **clear licensing and IPR conditions**;

PREFORMA General Presentation



Requirements for Long-term Sustainability of Open Source Communities



- **provision of executables** - there shall always be executables for several different deployment platforms;
- **provision of build environment and its source code** - the specific tool chain used for creation of a running instance of the open source code (the build environment) must be provided under an open source license, i.e. a license approved by the Open Source Initiative (www.opensource.org).
- **long-term sustainability of a technical specification of a file format** - promoted through implementation in open source software for which there is transparent information on how the specification of the file format has been interpreted.

PREFORMA General Presentation



PREFORMA Open Source Projects Provided via Three Channels



1. Development releases provided via an open collaboration platform (GitHub)
 - all code maintained and openly available
2. Stable (monthly) releases of source code, executables, and build environment available at the Open Source Portal for easy download and use under the PREFORMA licenses.
 - all code maintained long-term (after being provided by the suppliers)
3. Releases of developed software provided via each supplier's own channels

PREFORMA General Presentation



Events



Today: Open Source Preservation Workshop in Stockholm, in connection with the first Prototype Demonstration.

Forthcoming:

- **Experience Workshop** that will take place in Berlin in November 2016, in connection with the second Prototype Demonstration
- **Final Conference** that will take place in Stockholm in the end of 2017 to present the results of the project.

PREFORMA General Presentation



PREFORMA Open Source Workshop

Free Software and Open Formats: Virtual immortality and independence for digital archives

Stockholm, 07. April 2016

Hello ;-)

Well, what actually is "Open Source"?

Common misassumptions:

- Open Source = Freeware
- Must be non-commercial
- Is nerdy and non-user friendly
- This strange "Lunax" system...?

So Open Source is actually about **YOUR user freedom...**

Free Software

Free as in free speech!

Free Software

Defined by "The 4 Freedoms"

The right to:

- **USE** it for any purpose
- **STUDY** how the program works and understand it
- **SHARE** copies of the software
- **IMPROVE** the program and distribute these changes

FOSS is defined by providing the user at least 4 digital freedoms.

So what?

Benefits for long-term preservation?

Let's look at it in detail...

How long is "long term"?

- 10 years?
- 100 years?
- longer?
- or: 3-5 years?
= market-lifetime of software



Let's define "long term" for preservation/archiving as = Infinity! :)



+1

Just to be sure, let's add one day ;)

Long-term = Infinity + 1 day

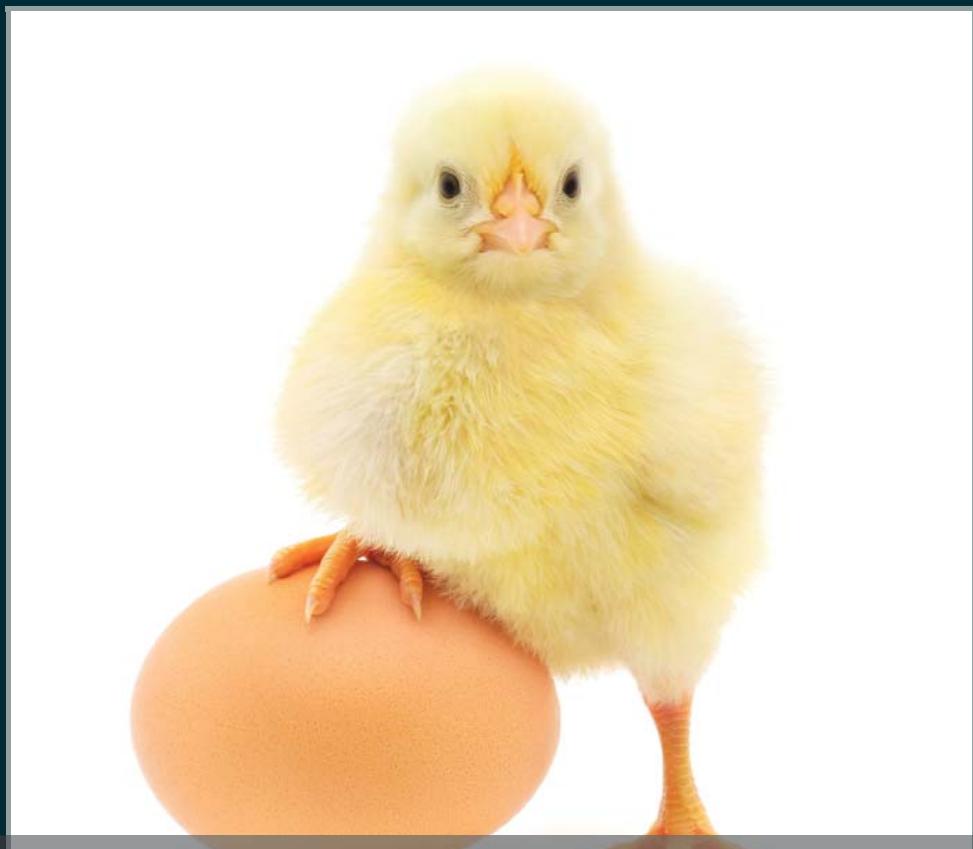
Virtual Immortality?

- Archive your replayer
- Including schematics
- Building components
- Exchange it with others
- Make it work under future conditions

Free Software allows you this, by license definition. Irrevocably = for infinity. This counteracts issues of tool/format obsolescence.

Why isn't it more widely used or known in archival domains?

Sounds too good to be true. So why is the norm software still proprietary in archiving?



It's a **chicken-egg-problem**:

- Vendors don't offer FOSS, because customers/archives don't request FOSS.
- Archives don't request FOSS, because vendors don't offer FOSS.

Perceived professionalism and quality

- Price = quality?
- GUI = quality?
- Professionals not using it?
- Ongoing support?
- Responsibilities?

"Professionals" often consider brand and price as an indicator for quality. This is why "free" is often mistakenly thought to be "unprofessional" or of lower quality.

With these doubts, and because "*noone has ever done this before...*", archives currently rather stay with proprietary products.

Usability concerns or lack of: GUI, ease of install, support, etc.

Understandable - but this can be improved.

Price = Quality?

- You can get buggy software under all licenses: proprietary and free.
- And within any price-range ;)

Price = Quality?

Quality of software is price- and license-independent.

This should always be kept in mind when dealing with software. Also the fact that "just because everybody else is using it" doesn't mean it's good (for you).

Professional vs. Free?

Professionals don't use Free Software...?

- BBC:
 - Dirac: <http://diracvideo.org/>
 - Ingex: <http://ingex.sourceforge.net/>
 - and many more:
<http://www.bbc.co.uk/opensource/projects/>
- GlusterFS: Used for Hollywood movies
- Open EXR: Developed and used by "Industrial Light & Magic"
- And GNU/Linux (which also powers the Internet) of course ;)

Independence?

The archival domain

- Relatively small market (niche)
- Off-the-shelf products often focus on different use cases
- Highly specialized demands
- High potential for vendor dependence...

With Free Software

- Reusability of existing solutions
- Free choice of (local) support/suppliers
- No black-box
- Common tools/codebase = larger userbase
- Less "forced" upgrades

The next step:
Professionalization of Free Software

Paying for Free Software: The best of both worlds

Currently, funding goes to vendors who do not offer you choice/freedom. Is it in your interest to fund entities that profit from your dependence, lock-in and black-boxes you can't control?

Paying for Free Software Benefits:

- Better support/updates
- Pooling resources
- Improvements available "upstream"
- Designed for *your* use-cases
- Overall better cost-effectiveness
- Public money = public solutions

FOSS offers **you** freedom - and control over your tech-environment. If you don't support FOSS with resources, you will lose these freedoms/control - because: Who should offer it to you, then?

Professionalizing Free Software: How?

- Make support contracts
- Paid installation/integration
- Hire developers
- Non-financial contributions:
 - Documentation
 - Testing / bug-reports
 - Helping others

If you like a FOSS tool - especially if you use it at work:

Consider paying at least some amount. It will keep the project healthy - and you profit from that.

Professionalizing Free Software: Whom to ask?

- Open Preservation Foundation
- Free Software Foundation Europe (FSFE)
- Maintainer of FOSS-project of your interest
- Companies that support/develop Free Software

Professionalizing Free Software: A real-world example

The lossless video codec FFV1

How it all began...

- Austrian National Archive (Mediathek) wanted to do lossless digital video archiving
- Not satisfied with existing products (Interoperability issues)
- Found FFV1 in FFmpeg
- Excellent codec, but we wanted/needed more...

How we used the "4 freedoms"...

- Contacted and hired FFV1 maintainer Michael Niedermayer
- Other parties involved (pooled resources)
- Budget calculated in reference to costs of proprietary alternatives
- Had FFV1 improvements in progress...

How it continued...

- FFmpeg forked: LibAV
- Patches then submitted to LibAV, too
- FFV1.3: faster and integrity-aware
- Important: Published our experiences with FFV1
- So: Other archives using FFV1 now profit from improvements, too!

PREFORMA's role

- Upcoming conformance/policy checker
- Standardization of FFV1/MKV
- Increased credibility of these formats
- Reduced skepsis of users/vendors
- Encouraging similar projects in the future

A growing Free Software ecosystem for long-term preservation

- VeraPDF
- DPF Manager
- MediaInfo, MediaConch
- JHOVE
- DROID
- VLC
- FFmpeg / LibAV
- SoX, ImageMagick
- Archivematica
- CollectiveAccess
- DVA-Profession
- AnyArk
- QCTools
- dvdisaster
- and many more...

Growing ecosystem for professional archiving with FOSS

Collaboration welcome

We archives have:

- Common interests.
- Common challenges.
- Common solutions?

- We all profit from investing in FOSS solutions.
- FOSS allows pooling resources.
- FFV1 just one example for what is possible.

- Fin -

About myself

Peter Bubestinger-Steindl
(p.bubestinger@das-werkstatt.com)

- Studied Media Computer Science at the TU Vienna
- Developer and Tech-Consultant since 2000
- Hands-on experience with archives since 2002:
 - Austrian Mediathek
 - University of Notre Dame Archives
 - The National Archives U.K.
 - ORF (National broadcaster, Austria)
 - VoV (National broadcaster, Vietnam)
 - RTV (National broadcaster, Slovenia)
 - SRTC (National broadcaster, Sudan)
 - Fonoteca Nacional (Mexico)
 - Memnon Archiving Services (Belgium)
 - SRF (Sweden), YLE (Finland), SRR (Romania), ...

License and Credits

This presentation is available under a Free License:

*“Creative Commons "Attribution-ShareAlike"
(CC-BY-SA) ”*



Peter Bubestinger-Steindl
(p.bubestinger@das-werkstatt.com)

The Relevance of FOSS License Compatibility

Open Source Preservation Workshop, PREFORMA, Stockholm, April 7, 2016

Dr. Till Jaeger
Certified Copyright and Media Law Attorney

JBB

Rechtsanwälte
Jaschinski Biere Brexl Partnerschaft

www.jbb.de

Derivative Works

Modifications of preexisting software result in a copyright on the modified software (if modification is individual rather than trivial) - „derivative works“

The copyright holder of the modified work needs the permission of the copyright holder of the original work to use the modified work

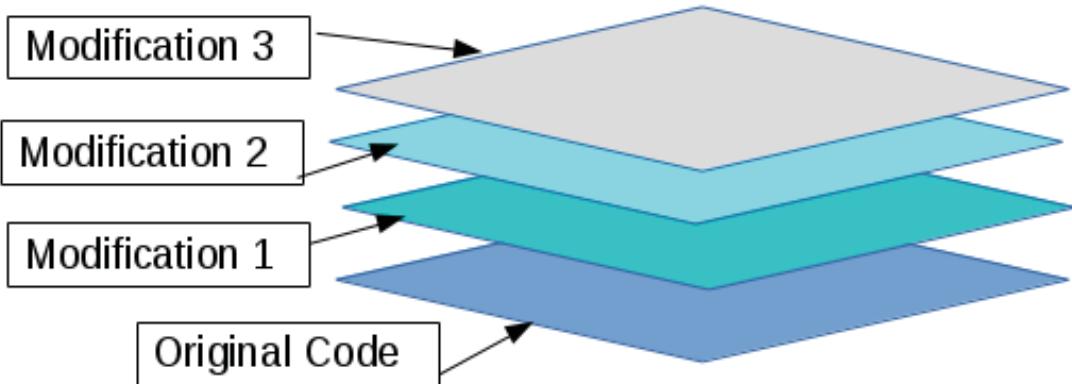
The copyright holder of the original work can control if modifications are made and how to use them

JBB

Rechtsanwälte
Jaschinski Biere Brexl Partnerschaft

www.jbb.de

Copyright on Derivative Works



Copyleft and License Categories

“Copyleft”: Obligation to the licensee to license derivative works under the original OSS license (in case of distribution)

“Legal hack” of Richard Stallman to avoid that licensees do not provide the source code of their modifications and proprietary use of the modified software

The concept of Copyleft: using copyright to ensure free use of a software instead of collecting royalties

Copyleft and License Categories

Copyleft licenses

- Strong copyleft licenses: all modifications have to be licensed under the original license if distributed
- Weak copyleft licenses: some modifications have to be licensed under the original license if distributed

Non-Copyleft licenses (permissive licenses): derivative works may be licensed under different license conditions, even proprietary

Copyleft and License Categories

Licenses with strong Copyleft:

- GPL - GNU General Public License, versions 2 and 3
- AGPL - GNU Affero General Public License
- EPL – Eclipse Public License
- EUPL – European Union Public License

Licenses with weak Copyleft

- LGPL – GNU Lesser General Public License, versions 2, 2.1 and 3
- MPL – Mozilla Public License, versions 1, 1.1 and 2

Copyleft and License Categories

Non-Copyleft licenses:

- BSD License (2-Clause, 3-Clause and 4-Clause)
 - Apache License 2.0
 - MIT License
 - Academic Free License
 - PHP License
- ...

JBB

Rechtsanwälte
Jaschinski Biere Brexl Partnerschaft

www.jbb.de

7

The Impact of Copyleft

The Copyleft is relevant for:

- Licensing your own developments/modifications
- Compatibility with other OSS-components and remixing code

Mostly a problem of copyright law: what is a “derivative work” and triggers Copyleft?

Currently no case law that involve GPL interpretation or concerning the concept of adaptation of software, but: VMware case currently at court (District Court of Hamburg)

JBB

Rechtsanwälte
Jaschinski Biere Brexl Partnerschaft

www.jbb.de

8

The Impact of Copyleft

Relevant for the question whether code from one Open Source component can be copied in another component

Relevant for the question whether Open Source components under various licenses can be combined together within one work

```
vp8_cx_iface.c
3-Clause-BSD

if (!ctx->priv)
{
    struct vpx_codec_alg_priv *priv =
        (struct vpx_codec_alg_priv *)
        vpx_callon(1, sizeof(*priv));
}

if (!priv)
{
    return VPX_CODEC_MEM_ERROR;
}

ctx->priv = (vpx_codec_priv_t *)priv;
ctx->priv->init_flags = ctx->init_flags;

if (ctx->config.enc)
{
    /* Update the reference to the config
     * structure to an
     * internal copy.
     */
    priv->cfg = *ctx->config.enc;
    ctx->config.enc = &priv->cfg;
}

priv->vp8_cfg = default_vp8_cfg;
priv->vp8_cfg.pkt_list = &priv->pkt_list.head;

priv->cx_data_sz = priv->cfg.q_w * priv-
>cfg.q_h * 3 / 2 * 2;

if (priv->cx_data_sz < 32768) priv-
>cx_data_sz = 32768;

priv->cx_data = malloc(priv->cx_data_sz);

getmsg.c
GPL-2.0

#include <errno.h>
#include <stropts.h>

int
getmsg (fdlist, ctptr, dataptr, flagsp)
    int fdlist;
    struct strbuf *ctptr;
    struct strbuf *dataptr;
    int *flagsp;
{
    set_errno (ENOSYS);
    return -1;
}

stub_warning (getmsg)
#include <stub-tag.h>
```

```

vp8_cx_iface.c
3-Clause-BSD

if (!ctx->priv)
{
    struct vpx_codec_alg_priv *priv =
        (struct vpx_codec_alg_priv *)
        vpx_calloc(1, sizeof(*priv));

    if (!priv)
    {
        return VPX_CODEC_MEM_ERROR;
    }

    ctx->priv = (vpx_codec_priv_t *)priv;
    ctx->priv->init_flags = ctx->init_flags;

    if (ctx->config_enc)
    {
        /* Update the reference to the config
        structure to an
        * internal copy
        */
        priv->cfg = &ctx->config_enc;
        ctx->config_enc = &priv->cfg;
    }

    priv->vp8_cfg = default_extractcfg;
    priv->vp8_cfg.pkt_list = &priv-
    >pkt_list.head;

    priv->cx_data_sz = priv->cfg.q_w * priv-
    >cfg.q_h * 3 / 2 * 2;

    if (priv->cx_data_sz < 32768) priv-
    >cx_data_sz = 32768;

    priv->cx_data = malloc(priv->cx_data_sz);
}

```

```

getmsg.c
GPL-2.0

#include <errno.h>
#include <stropts.h>

int
getmsg (fildes, ctptr, dataptr, flagsp)
    int fildes;
    struct strbuf *ctptr;
    struct strbuf *dataptr;
    int *flagsp;
{
    _set_errno (ENOSYS);
    return -1;
}

stub_warning (getmsg)
#include <stub-tag.h>

```

```

vp8_cx_iface.c
3-Clause-BSD

if (!ctx->priv)
{
    struct vpx_codec_alg_priv *priv =
        (struct vpx_codec_alg_priv *)
        vpx_calloc(1, sizeof(*priv));

    if (!priv)
    {
        return VPX_CODEC_MEM_ERROR;
    }

    ctx->priv = (vpx_codec_priv_t *)priv;
    ctx->priv->init_flags = ctx->init_flags;

    if (ctx->config_enc)
    {
        /* Update the reference to the config
        structure to an
        * internal copy
        */
        priv->cfg = &ctx->config_enc;
        ctx->config_enc = &priv->cfg;
    }

    priv->vp8_cfg = default_extractcfg;
    priv->vp8_cfg.pkt_list = &priv-
    >pkt_list.head;

    priv->cx_data_sz = priv->cfg.q_w * priv-
    >cfg.q_h * 3 / 2 * 2;

    if (priv->cx_data_sz < 32768) priv-
    >cx_data_sz = 32768;

    priv->cx_data = malloc(priv->cx_data_sz);
}

```

```

getmsg.c
GPL-2.0

#include <errno.h>
#include <stropts.h>

int
getmsg (fildes, ctptr, dataptr, flagsp)
    int fildes;
    struct strbuf *ctptr;
    struct strbuf *dataptr;
    int *flagsp;
{
    _set_errno (ENOSYS);
    return -1;
}

if (ctx->config_enc)
{
    /* Update the reference to the config
    structure to an
    * internal copy
    */
    priv->cfg = &ctx->config_enc;
    ctx->config_enc = &priv->cfg;
}

stub_warning (getmsg)
#include <stub-tag.h>

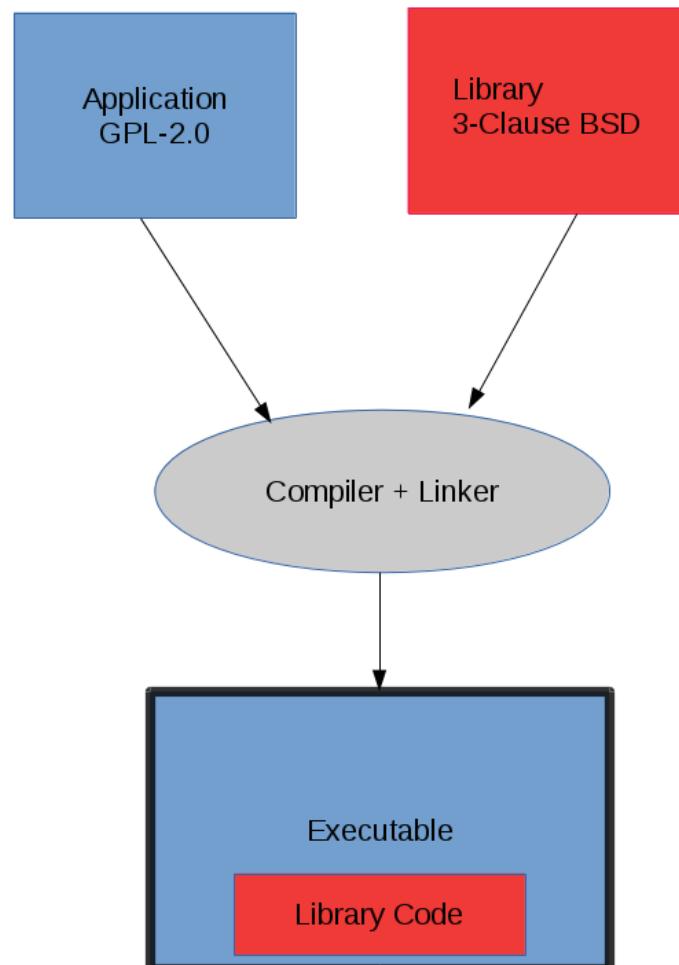
```

The Impact of Copyleft

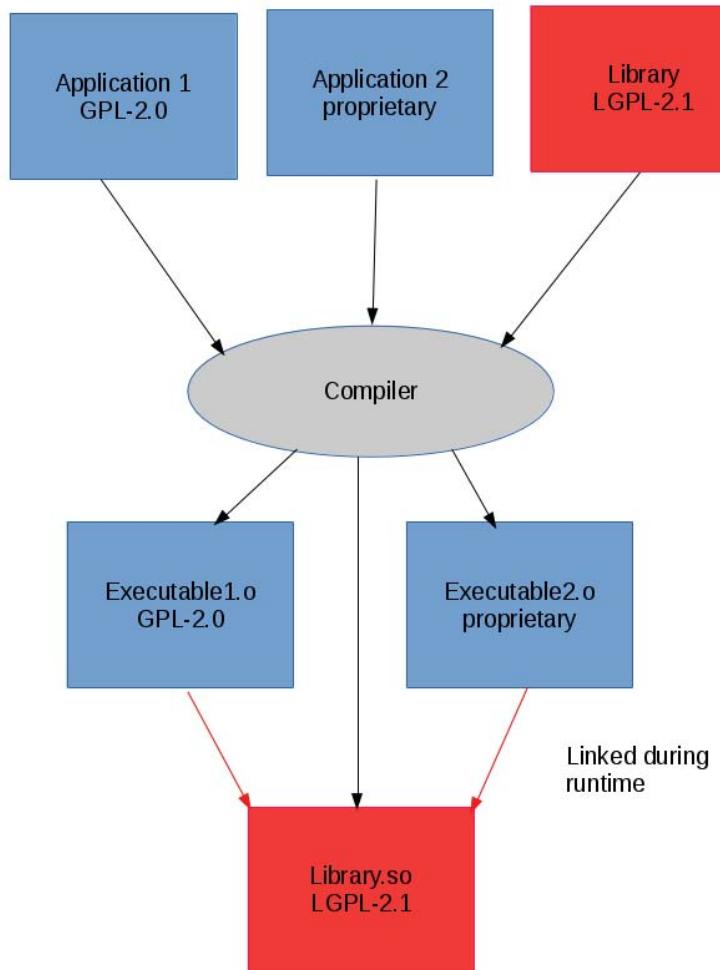
Combination of components:

- Libraries
- Plugins
- Modules

Static Linking



Dynamic Linking



The Impact of Copyleft

Derivative works cannot be licensed under more than one license at the same time (except in the case of dual licensing)

Which licenses are compatible (allowing the use of another license for the modified code)? Which license prevails?

Compliance process: a) are two components to be considered a derivative work (as defined in the respective licenses)?, b) if yes: are the FOSS licenses compatible?

Non-Copyleft and Non-Copyleft

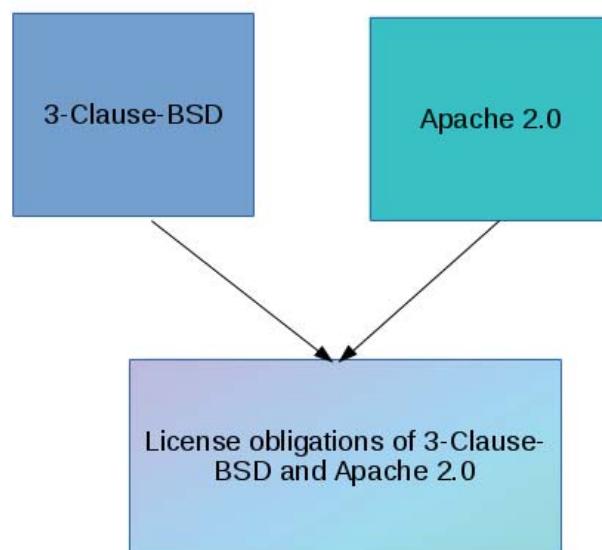
Mixing code under two or more Non-Copyleft licenses

License obligations of all licenses have to be fulfilled altogether

Non-Copyleft licenses are always compatible among each other

Choosing a proprietary license or a Non-Copyleft license for the derivative work

Non-Copyleft and Non-Copyleft



Copyleft and Copyleft

Mixing code under two or more copyleft licenses

Code licensed under two different copyleft licenses cannot be combined in one program, as one of the licenses would necessarily have to be violated

Copyleft licenses are not compatible except...

Copyleft and Copyleft

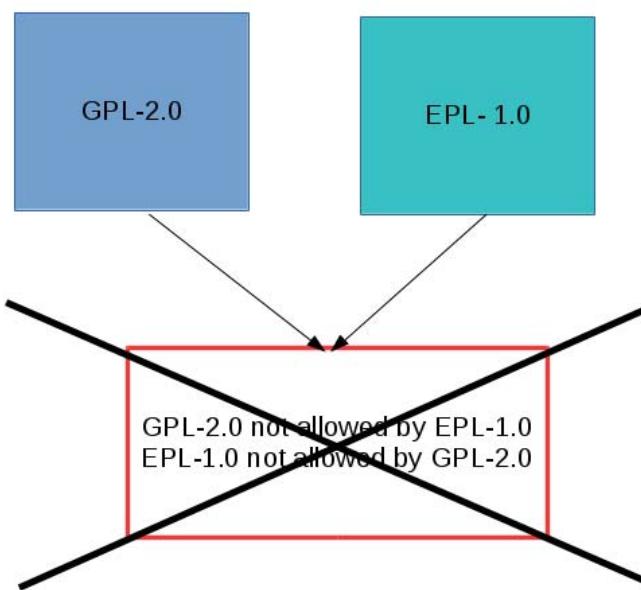
... a specific compatibility clause is allowing the combination

E.g. Sec. 3 LGPL-2.1: „*You may opt to apply the terms of the ordinary GNU General Public License instead of this License to a given copy of the Library. To do this, you must alter all the notices that refer to this License, so that they refer to the ordinary GNU General Public License, version 2, instead of to this License.*“

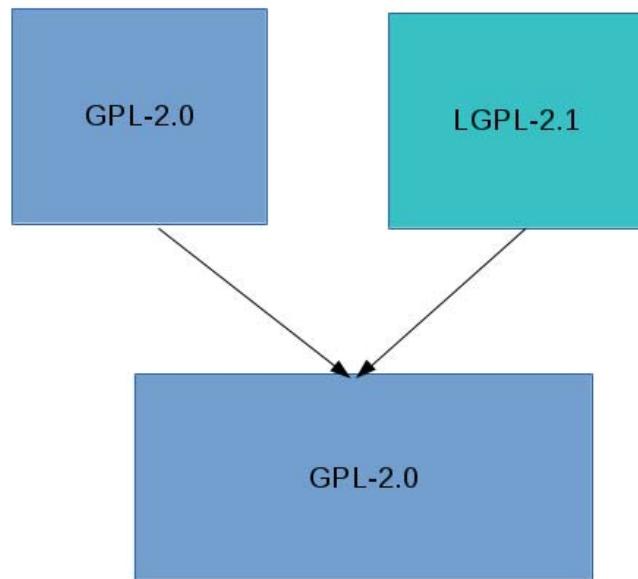
Copyleft and Copyleft

Sec. 5 EUPL: „*Compatibility clause: If the Licensee Distributes and/or Communicates Derivative Works or copies thereof based upon both the Original Work and another work licensed under a Compatible Licence, this Distribution and/or Communication can be done under the terms of this Compatible Licence. For the sake of this clause, “Compatible Licence” refers to the licences listed in the appendix attached to this Licence. Should the Licensee’s obligations under the Compatible Licence conflict with his/her obligations under this Licence, the obligations of the Compatible Licence shall prevail.*“

Copyleft and Copyleft



Copyleft and Copyleft



Copyleft and Non-Copyleft

Mixing code under a Copyleft license with code under a Non-Copyleft license

Non-Copyleft and Copyleft licenses are compatible but in the case that the Non-Copyleft license provides an obligation the Copyleft license does not provide

Copyleft licenses do not allow additional restrictions (e.g. Sec. 6 GPL-2.0: „*You may not impose any further restrictions on the recipients' exercise of the rights granted herein.*“)

Copyleft and Non-Copyleft

Example – 4-Clause-BSD license:

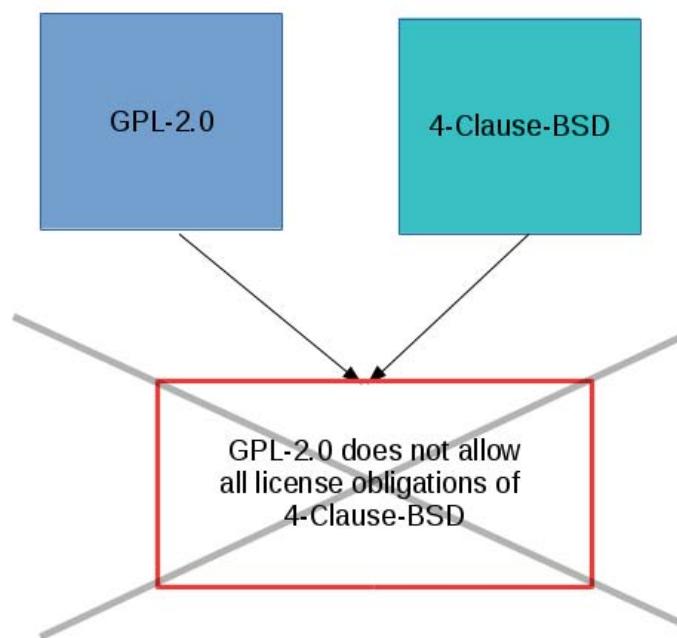
All advertising materials mentioning features or use of this software must display the following acknowledgement:

“This product includes software developed by the organization.“

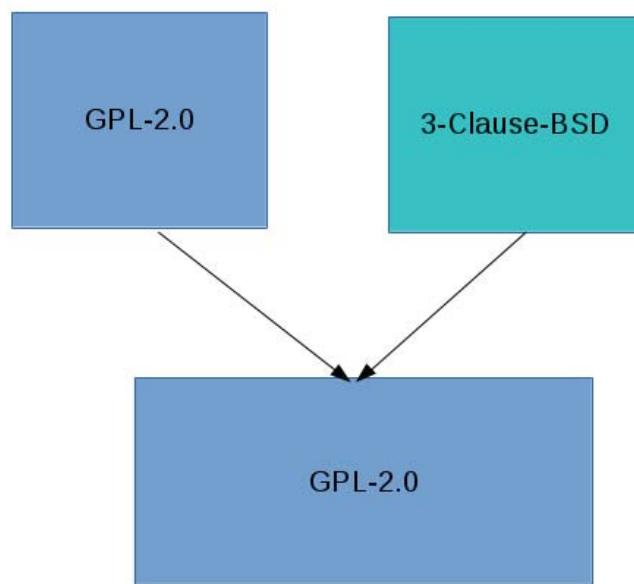
GPL-2.0 does not provide for such an obligation

GPL-2.0 and 4-Clause-BSD are not compatible

Copyleft and Non-Copyleft



Copyleft and Non-Copyleft



License Compatibility

Examination:

- Is there a Copyleft license? Is the Copyleft applicable? (e.g., derivative work) - compatibility of the licenses involved
- More than one Copyleft license?
- If yes: compatibility clause applicable?
- If not: contain(s) the Non-Copyleft license(s) restrictions not contained in the Copyleft license

License Compatibility

GPL-2.0 (Linux) and GPL-3.0 – internal compatibility

Sec. 9 GPL-2.0 provides as follows:

- "any later version" – license under the GPL-3.0 possible
- Program does not specify a version number – license under all versions (GPL-1.0, GPL-2.0 and GPL-3.0) possible
- GPL-2.0 only? FSF tolerates this licensing for a long time – GPL-2.0 and GPL-3.0 are not compatible in this case

License Compatibility

New solution for the compatibility problem in the GPL-3.0

Sec. 7 GPL-3.0

"Notwithstanding any other provision of this License, for material you add to a covered work, you may (if authorized by the copyright holders of that material) supplement the terms of this License with terms:

- a) Disclaiming warranty or limiting liability differently from the terms of sections 15 and 16 of this License; or*
- b) Requiring preservation of specified reasonable legal notices or author attributions in that material or in the Appropriate Legal Notices displayed by works containing it; or"*

License Compatibility

	GPL-2.0	GPL-3.0	LGPL-2.1	LGPL-3.0	Apache License 2.0	MIT License and BSD License (3 clause)	MPL-2.0 ¹	CPL/EPL
GPL-2.0		-	+	-	-	+	+	-
GPL-3.0	-		+	+	+	+	+	-
LGPL-2.1	+	+		+	-	+	+	-
LGPL-3.0	-	+	+		+	+	+	-
Apache License 2.0	-	+	-	+		+	-	-
MIT License and BSD License (3 clause)	+	+	+	+	+		+	+
MPL-2.0 ¹	+	+	+	+	-	+		-
CPL/EPL	-	-	-	-	-	+	-	

¹ Depending on Exhibit B

JBB

Rechtsanwälte
Jaschinski Biere Brexl Partnerschaft

www.jbb.de

31

Thank you for your attention!

JBB Rechtsanwälte

Dr. Till Jaeger

Christinenstraße 18/19

10119 Berlin

www.jbb.de

JBB

Rechtsanwälte
Jaschinski Biere Brexl Partnerschaft

www.jbb.de

32



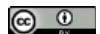
Ligue des Bibliothèques Européennes de Recherche
Association of European Research Libraries

Open Science, Open Data:

**towards a new transparent and
reproducible ecosystem**

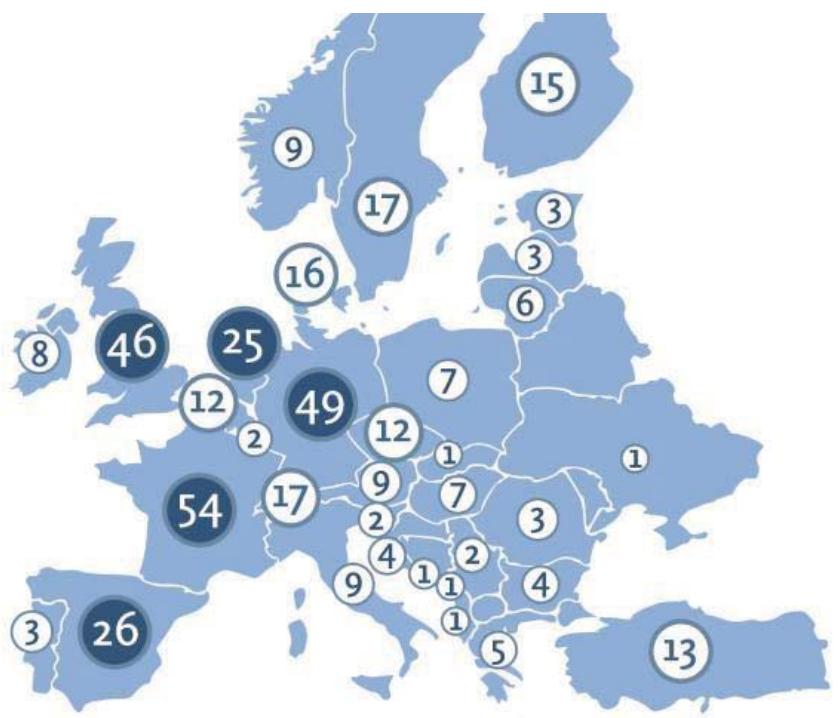


Melanie Imming
EU Projects manager, LIBER



LIBER Europe

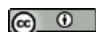
- Research Libraries
- Founded in 1971
- More than 400 national, university and other libraries from over 40 countries



LIBER Conference



 LIBER
Ligue des Bibliothèques Européennes de Recherche
Association of European Research Libraries



LIBER EU Projects



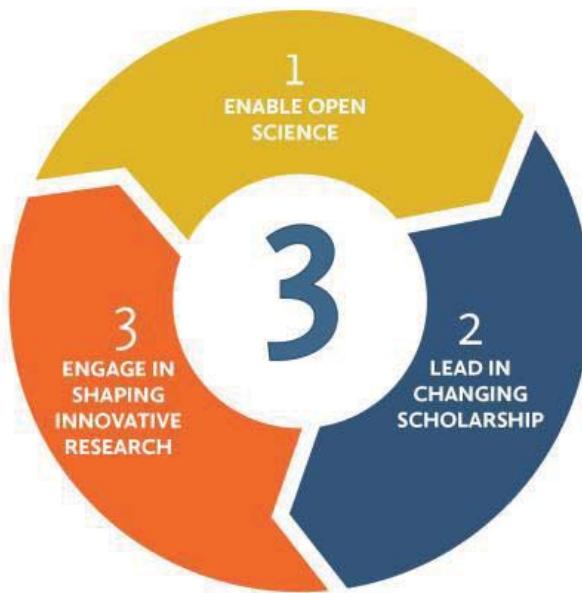
 LIBER
Ligue des Bibliothèques Européennes de Recherche
Association of European Research Libraries



LIBER Europe

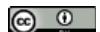
STRATEGIC DIRECTIONS

2015-2017



- **BUILDING CAPACITY** AND BEST PRACTICE IN AREAS SUCH AS **OPEN SCIENCE & METRICS**
- **SUPPORTING RESEARCH DATA MANAGEMENT** DEVELOPMENT
- **EQUIPPING LIBRARY LEADERS** TO PLACE THEIR LIBRARIES AT THE HEART OF **INNOVATIVE AND OPEN RESEARCH ENVIRONMENTS**
- **PROMOTING POSITIONS** ON OPEN ACCESS, RDM, COPYRIGHT, TEXT AND DATA MINING & OTHER TOPICS
- **SUPPLYING TOOLS** TO ENGAGE STAKEHOLDERS ON THESE ISSUES

Read more: www.libereurope.eu/strategy



*“LIBER is Re-inventing
the Library
for the Future”*

1
ENABLE OPEN
SCIENCE

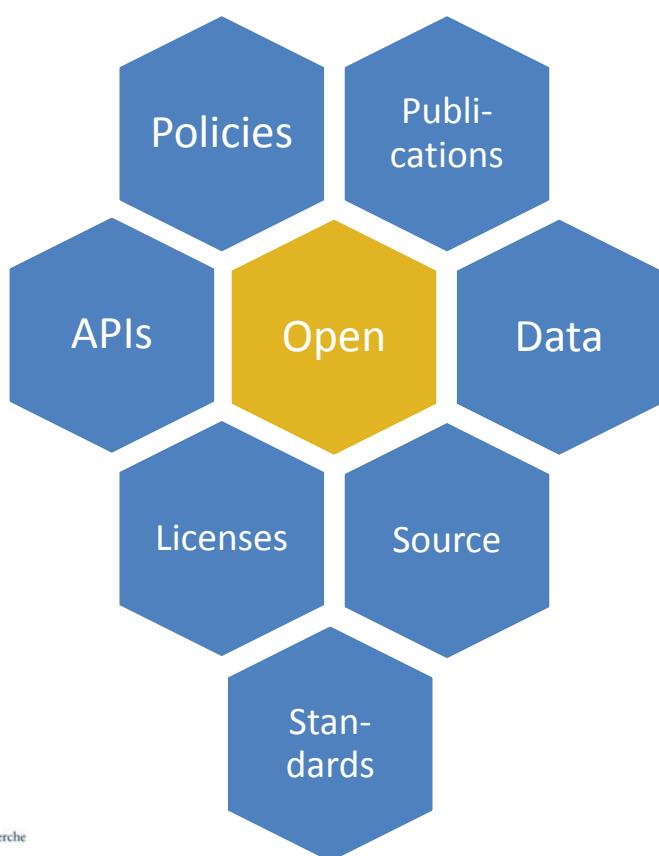
LIBER Europe

A central part of LIBER's mission is to provide an information infrastructure that enables research in LIBER institutions to be world class.

For this infrastructure to thrive, it must be part of an ecosystem that can accommodate and nurture the changing nature of research and innovation in the digital age.



Open Science



Open Science Definition

“The conduction of science in a way that others can **collaborate** and **contribute**, where research data, lab notes and other research processes are **freely available**, with terms that allow **reuse**, **redistribution** and **reproduction** of the research”

<https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>



CC BY Openaccessbutton.org



LIBER and Open Science

So, we need alternative mechanisms for the recognition of excellence in Open Science, e.g. ranking systems, to Open up Science.

- From publish or perish to Open Science
- Scientific tools used need to be cited, and, in order to make scientific experiments recreatable, there need to be incentives to create open and sustainable software



Science Code Manifesto

[Manifesto](#) [Discussion](#) [Endorse](#) [Resources](#) [About](#)

Software is a cornerstone of science. Without software, twenty-first century science would be impossible. Without better software, science cannot progress.

But the culture and institutions of science have not yet adjusted to this reality. We need to reform them to address this challenge, by adopting these five principles:

Code	All source code written specifically to process data for a published paper must be available to the reviewers and readers of the paper.
Copyright	The copyright ownership and license of any released source code must be clearly stated.
Citation	Researchers who use or adapt science source code in their research must credit the code's creators in resulting publications.
Credit	Software contributions must be included in systems of scientific assessment, credit, and recognition.
Curation	Source code must remain available, linked to related materials, for the useful lifetime of the publication.

Science Code Manifesto

•Code

source code written specifically to process data for a published paper must be available to the reviewers and readers of the paper.



Ligue des Bibliothèques Européennes de Recherche
Association of European Research Libraries



Science Code Manifesto

•Code

•Copyright

The copyright ownership and license of any released source code must be clearly stated.



Ligue des Bibliothèques Européennes de Recherche
Association of European Research Libraries



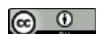
Science Code Manifesto

- Code
- Copyright
- Citation

Researchers who use or adapt science source code in their research must credit the code's creators in resulting publications



Ligue des Bibliothèques Européennes de Recherche
Association of European Research Libraries



Science Code Manifesto

- Code
- Copyright
- Citation
- Credit

Software contributions must be included in systems of scientific assessment, credit, and recognition.



Ligue des Bibliothèques Européennes de Recherche
Association of European Research Libraries



Science Code Manifesto

- Code
- Copyright
- Citation
- Credit
- Curation**

Source code must remain available, linked to related materials, for the useful lifetime of the publication.



LIBER and Open Science

Curation

‘Making sure data is stored in a controlled way and can be (re) used today and in the future is an important element in Open Science’.

LIBER and Open Science

Standardisation of file formats will ensure (re-)usability today and in the future, as it:

- enables processing and preservation of data in a controlled way
- ensures outputs that are really open and accessible in the long term
- improves interoperability of new tools and services



Workshop Nov 2015: “Text and Data Mining in Europe: Challenges and Action”



Participants: content providers (publishers, data centers, museums and libraries)



Technical challenges identified:

- Quality of datasets
- Lack of a secure infrastructure

Workshop Nov 2015: “Text and Data Mining in Europe: Challenges and Action”



Solutions:

- Develop and use **open standards**
- Develop templates for metadata and content
- Allow for peer review of data quality, develop **validation tools**,
- Appraise good quality data
- Organisations should invest resources to improve the quality of their data



Workshop Feb 2016: European Open Science Cloud



Opening paragraph of The European Open Science Cloud for Research
Rome Workshop Report:

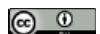
‘The creation of a trusted environment for hosting and processing research data (...) will help overcome many key challenges currently facing scientific disciplines. These challenges include a huge lack of awareness of the value of data and the incentives for data sharing, a continued lack of and urgent need for common standards to ensure interoperability of data...’

Netherlands EU Presidency Open Science Conference

Amsterdam, 4/5 April 2016



LIBER
Ligue des Bibliothèques Européennes de Recherche
Association of European Research Libraries



Libraries enabling Open Science

Data Issue Libraries and data centres opportunities

Availability

- Lower barriers to researchers to make their data available
- Integrate data sets into retrieval services

Findability

- Support of persistent identifiers
- Engage in developing common meta description schemas and common citation practices
- Promote use of common standards and tools among researchers

Interpretability

- Support crosslinks between publications and datasets
- Provide and help researchers understand meta-descriptions of datasets
- Establish and maintain a knowledge base about data and their context

Libraries enabling Open Science

Data Issue Libraries and data centres opportunities

Re-usability	-Curate and preserve datasets -Archive software needed for re-analysis of data -Be transparent about conditions under which data sets can be re-used (expert knowledge needed, software needed)
Citability	-Engage in establishing uniform data citation standards -Support and promote persistent identifiers
Curation/ Preservation	-Transparency about curation of submitted data -Promote good data management practice -Collaborate with data creators -Instruct researchers on discipline specific best practices in data creation (preservation formats, documentation of experiment,...)



Libraries enabling Open Science

Focus on Research Data Management:

- Growing variety of data types and volume
- Curation of data from the planning stage of research projects

Libraries enabling Open Science

Awareness, trust and community building

- Institutions - develop policies and roadmaps
- Researchers - highlight benefits of open science
- (Other) Stakeholders at institutional level and internationally



Libraries enabling Open Science

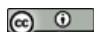
- Stay in control!
- Unite!
- Be active in projects like Preforma
- Advocate & Engage



Open Science

What can you do?

- Release data under CC0
- media components and arrangements of data under CC BY
- Work from what is already working
- Use what is really open: freely available, can be freely adopted, implemented and extended (no license fees)
- Sign The Hague Declaration!



[Home](#) ▾ [Declaration](#) ▾ [Background](#) ▾ [Original Participants](#) ▾ [Contact](#) ▾



[Download PDF](#) ▾

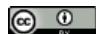
THE HAGUE DECLARATION ON KNOWLEDGE DISCOVERY IN THE DIGITAL AGE

Thank you!

- The Hague Declaration: <http://thehaguedclaration.com>
- LERU Roadmap for Research Data
<http://www.leru.org/index.php/public/news/press-release-leru-roadmap-for-research-data>
- <http://sciencecodemanifesto.org>
- Research Data Alliance <https://rd-alliance.org>
- LIBER 10 Recommendations on Getting Started in RDM
<http://libereurope.eu/wp-content/uploads/The%20research%20data%20group%202012%20v7%20final.pdf>
- OpenAire <https://www.openaire.eu>
- San Francisco Declaration
- <http://www.ascb.org/dora-old/files/SFDeclarationFINAL.pdf>



Ligue des Bibliothèques Européennes de Recherche
Association of European Research Libraries



APPENDIX B

This appendix contains slides used by presenters representing suppliers in the afternoon session of the Open Source Workshop.

1. Slides from the presentation “veraPDF: definitive, open source PDF/A validation for digital preservationists” by VeraPDF
2. Slides from the presentation “DPF Manager – The open source Community” by EasyInnova
3. Slides from the presentation “MediaConch – Implementation and policy checking on FFV1, Matroska, LPCM, and more” by MediaArea

veraPDF: definitive, open source PDF/A validation for digital preservationists

Open Preservation Foundation
PREFORMA Open Source Workshop 2016, Stockholm



Presenters



Joachim Jung, Open Preservation Foundation



Boris Doubrov, Dual Lab



Carl Wilson, Open Preservation Foundation



Supplier

veraPDF Consortium (Memory Institution - Industry Partnership) jointly lead by Open Preservation Foundation and PDF Association

Content type

Text

Standard specifications

PDF/A-1 (ISO 19005 - 1:2005); PDF/A-2 (ISO 19005 - 2:2011); PDF/A-3 (ISO 19005 - 3:2012); [PDF 1.7 (ISO 32000 - 1:2008)]

Standardisation efforts

Consolidation of ISO 19005 variants through PDF Association TWG

Website

<http://verapdf.org/>



3

The veraPDF consortium



4

The veraPDF consortium

Cultural Heritage

Document Industry



The veraPDF consortium

Cultural Heritage

Document Industry



Consortial
Lead



Open Preservation Foundation



- Founded in 2010 as the Open Planets Foundation
- International, not-for-profit membership organisation
- Sustain the results of R&D projects
- Steward open-source digital preservation software
- Enable knowledge exchange and the development of best practice through
 - interest groups
 - community events
 - training
- **Members** participate, guide and support its digital preservation activities
- **Supporters** ensure the sustainability and steer the development of individual digital preservation software products
- Members & supporters are libraries, archives, universities, companies who provide digital preservation services

<http://openpreservation.org/>



The veraPDF consortium

Cultural Heritage



Document Industry

Consortial
Lead



The veraPDF consortium

Cultural Heritage



Consortial
Lead



ISO TC 171 SC 2



9

PDF Association



- Initiative of the Association for Digital Document Standards e.V.
- Both founded in 2006
- First named the PDF/A Competence Center
- Promote of international standards for PDF technology
 - PDF (ISO 32000)
 - PDF/E (ISO 24517) creation of documents in engineering workflows
 - PDF/UA (ISO 14289) universally accessible PDF
 - PDF/VT (ISO 16612-2) Variable ("V") and Transactional ("T") workflows
- The work of the PDF Association is aimed at
 - developers of PDF solutions
 - companies that work with PDF in document management systems (DMS) and electronic content management (ECM)
- The PDF Association includes members from over 20 countries

<http://www.pdfa.org/>



10

The veraPDF consortium

Cultural Heritage

Community &
Testing Support



Consortial
Lead



Document Industry



ISO TC 171 SC 2



The veraPDF consortium

Cultural Heritage

Community &
Testing Support



Consortial
Lead



Industry
Partner

Solutions for
Memory Institutions

Document Industry



ISO TC 171 SC 2



The veraPDF consortium

Cultural Heritage

Community &
Testing Support



Consortial
Lead



Solutions for
Memory Institutions

Industry
Partner



ISO TC 171 SC 2



Solutions for PDF and beyond
Member of PDF Association



So what is veraPDF?

veraPDF is

- a consortium
- a set of open source (GPL3+ / MPL2+) applications:
 - Java software library for definitive PDF/A validation
 - CLI
 - GUI
 - REST
 - PREFORMA Conformance checker
- an authoritative test corpus for all flavours of PDF/A validation (CC-BY-4)

and potentially

- an open source framework for developing format validators

Why veraPDF?

- Digital preservation practitioners
 - Ensure newly curated documents comply to standards
 - Ensure compliance with institutional policies
- PDF technology developers
 - Reduce software development costs
 - Reduce support costs
- Document technology developers
 - Assess PDF technology options
 - Improve reliability and performance, and new options for products
- Customers creating and using documents
 - Reduce operational risks and legal liabilities
 - Receive more capable and interoperable software



PDF/A ISO standard

PDF/A is a standard that comprises three parts:

- PDF/A-1
 - ISO 19005-1:2005 Part 1: based on PDF 1.4
- PDF/A-2
 - ISO 19005-2:2011 Part 2: based on ISO 32000-1 (PDF 1.7)
- PDF/A-3
 - ISO 19005-3:2012 Part 3: based on ISO 32000-1 with support for embedded files

Note that while PDF/A-2 and PDF/A-3 are based on the ISO standard for PDF 1.7, PDF/A-1 is based on PDF 1.4 which is **not** an ISO standard.



PDF/A levels

The standards are subdivided into conformance levels

Parts 1 - 3 define:

- Level b - reliable **reproduction of visual appearance**
- Level a - Level b plus **document structure, tagging, and language**

Parts 2 & 3 additionally define:

- Level u - Level b plus **unicode character mappings (text extraction)**



PDF/A flavours

This gives us 8 PDF/A flavours:

- PDF/A-1b - PDF 1.4, visual reproduction
- PDF/A-1a - PDF 1.4, level b + structure
- PDF/A-2b - PDF 1.7, visual reproduction
- PDF/A-2a - PDF 1.7, level b + structure
- PDF/A-2u - PDF 1.7, level b + unicode map
- PDF/A-3b - PDF 1.7, visual reproduction + attachments
- PDF/A-3a - PDF 1.7, level b + structure + attachments
- PDF/A-3u - PDF 1.7, level b + unicode + attachments



veraPDF: Knowledge

Traceability, from standards to software



PDF/A Validation: The Challenge

- PDF/A relies on two different versions of PDF, with specifications 1000+ pages long each
- Contains references to a number of external standards: fonts, image compression, XMP metadata, ICC profiles, digital certificates
- Comes in three versions, three levels and two technical corrigenda for PDF/A-1
- Requires very formal analysis of the requirements

Isartor Test Suite: Building on Previous Success

The Isartor test suite was assembled by the PDFA's TWG, it was constructed according to these principles:

- **Atomic**: test files examine a particular aspect of the standard.
- **Complete**: provides full coverage of all requirements in the PDF/A-1 standard.
- **Self-documenting**: test files contain intrinsic information about the expected result of performing the test.
- **Traceable**: a test is based on a specific clause in PDF/A-1 which can easily be identified.
- **Reproducible**: All test files created programmatically with a documented change history.

<http://www.pdfa.org/2011/08/isartor-test-suite/>



Establishing Ground Truth

- Isartor, Bavaria, BFO: 323 test atomic self-documented test files.
- Initial test case analysis shows about 1000 extra files needed for proper test coverage of all PDF/A versions and levels.
- Subject to adjustment based on real-world cases and practical considerations.
- To compare: the W3C corpus for XML 1.1 (specified on 38 pages) contains over 600 test files



veraPDF: Test Corpus

In order to ensure that the software was accepted by industry we decided to produce a corpus of test files that:

- are traceable back to the PDF/A specification clauses
- provide pass and fail cases for all specification tests
- are open (CC-BY-4) licensed and freely reusable
- reviewed and ratified by the veraPDF TWG and (eventually) the PDF/A ISO committee

<https://github.com/veraPDF/veraPDF-corpus>



veraPDF: Validation Profiles

We didn't want to obfuscate our interpretation of PDF standards in code where only programmers can find them. Our solution has been to produce Validation Profiles that are:

- in plain text form so anyone can read, if not understand them
- traceable back to the PDF/A specification clauses
- open licensed and freely reusable
- reviewed and ratified by the veraPDF TWG and (eventually) the PDF/A ISO committee

<https://github.com/veraPDF/veraPDF-validation-profiles>



Knowledge: Starting with Standards

ISO 19005-1:2005 AKA PDF/A-1b - page 13

■ 6.1.10 Filters

The **LZWDecode** filter shall not be permitted.

NOTE. The use of the LZW compression algorithm has been subject to intellectual property constraints.

■ 6.1.11 Embedded files

A file specification dictionary, as defined in PDF 3.10.2, shall not contain the **EF** key. A file's name dictionary, as defined in PDF Reference 3.6.3, shall not contain the **EmbeddedFiles** key.

NOTE. These keys are used to encapsulate files containing arbitrary content within a PDF file. The explicit prohibition of these keys has the implicit effect of disallowing embedded files that can create external dependencies and complicate preservation efforts.



Validation Rules: LZW Compression

```
<rule object="CosIIFilter">
  <id specification="ISO_19005_1" clause="6.1.10" testNumber="1"/>
  <description>The LZWDecode filter shall not be
permitted</description>
  <test>internalRepresentation != "LZWDecode" &&
        internalRepresentation != "LZW"</test>
  <error>
    <message>LZW compression is used in the inline image</message>
    <arguments/>
  </error>
  <references/>
</profile>
```



veraPDF Corpus: LZW Compression

veraPDF / veraPDF-corpus

Code Issues 4 Pull requests 5 Boards Burndown Wiki

Branch: **staging** veraPDF-corpus / PDF_A-1b / 6.1 File structure / 6.1.10 Filters /

DmitryRemezow Add 5 test files for 6.1 / 6.2 clauses

..

veraPDF test suite 6-1-10-t01-fail-a.pdf

veraPDF test suite 6-1-10-t01-fail-b.pdf

veraPDF test suite 6-1-10-t01-fail-c.pdf

veraPDF test suite 6-1-10-t01-pass-a.pdf



Resolution of ambiguities

In some cases the specifications leave room for different interpretations of requirements. Such as:

■ **6.1.6 Strings**

Hexadecimal strings shall contain an even number of non-white-space characters.

Hexadecimal strings in PDF: <4E6F 7620 7368>

It is not clear, which characters are considered as white spaces. In particular, if a NULL (00h) character is allowed inside hexadecimal strings.

A special **PDF Validation Technical Workgroup** is formed by PDF Association dedicated to resolving ambiguities in various PDF standards and test corpora. Meets once per month to discuss the issues.



Collecting all information together: the Wiki

veraPDF Validation Rules

The veraPDF validation engine implements the PDF/A specification using formalizations of each "shall" statement (i.e., each requirement) in PDF/A-1, PDF/A-2 and PDF/A-3.

This wiki distinctly identifies each rule used by the software and provides details on the error(s) triggering a failure of the rule.

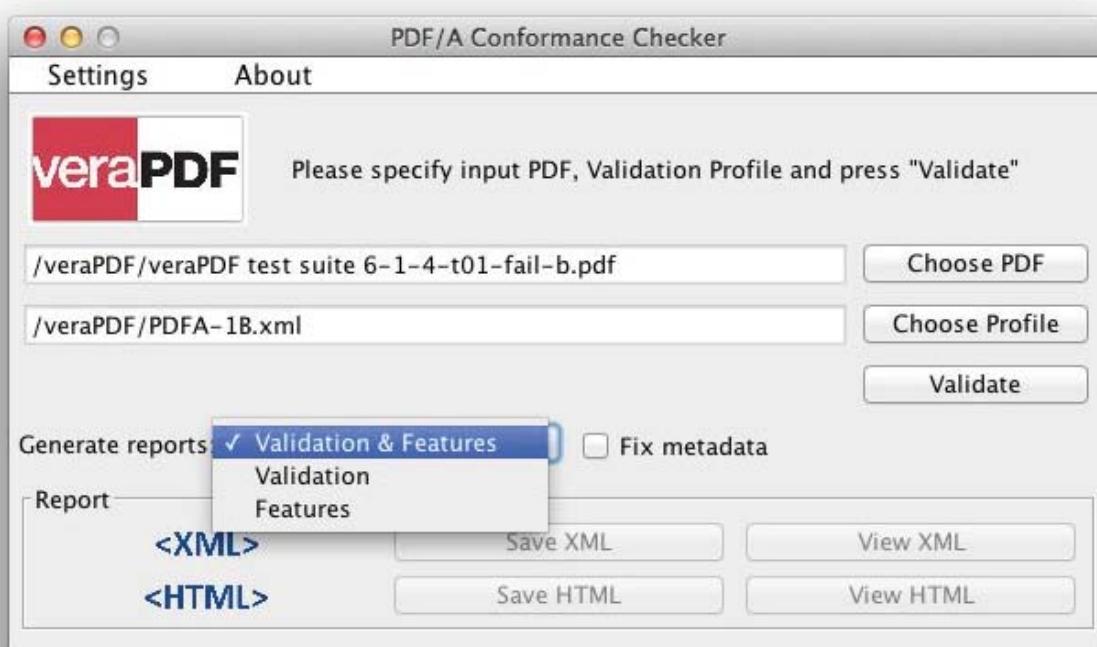
For each error the Object type, test condition, applicable specification and conformance level, as well as additional references, are provided.

Understandings based on the discussions of the PDF Validation Technical Working Group are included as appropriate.

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



Demonstrations

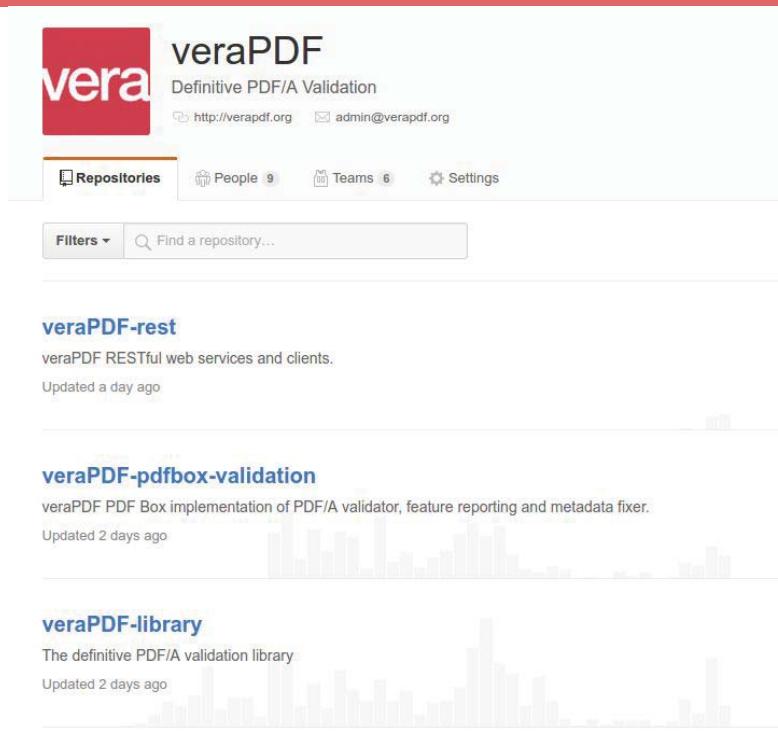


veraPDF: Software

Building a definitive PDF/A validator



veraPDF: GitHub



veraPDF
Definitive PDF/A Validation
<http://verapdf.org> admin@verapdf.org

Repositories 9 People 9 Teams 6 Settings

Filters Find a repository...

veraPDF-rest
veraPDF RESTful web services and clients.
Updated a day ago

veraPDF-pdfbox-validation
veraPDF PDF Box implementation of PDF/A validator, feature reporting and metadata fixer.
Updated 2 days ago

veraPDF-library
The definitive PDF/A validation library
Updated 2 days ago

veraPDF-apps
veraPDF GUI, CLI and installer for PDF Box validation
Updated 2 days ago

veraPDF: Travis CI

Travis CI Blog Status Help Carl Wilson

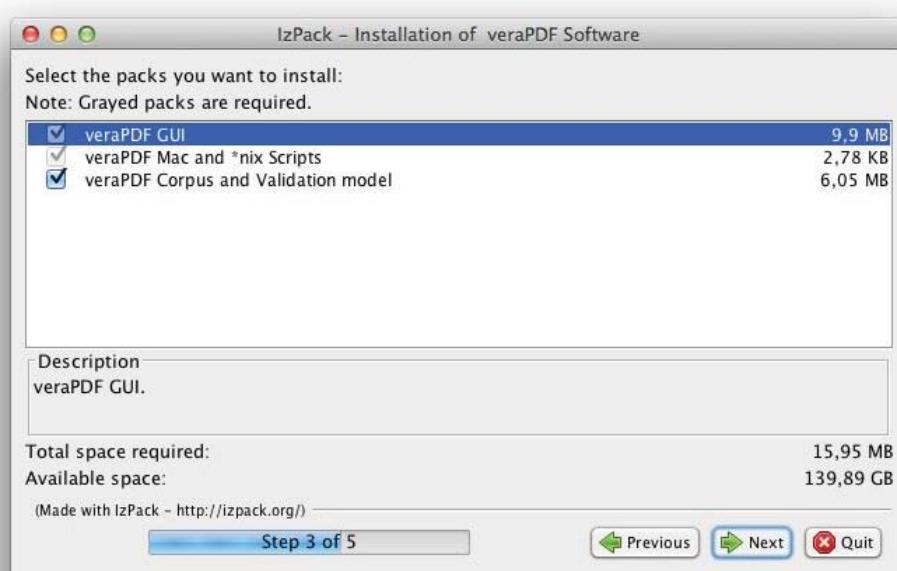
veraPDF veraPDF

Project	Build #	Branch	Commit	Last Run
✓ veraPDF-rest	# 9	-o- master	266a180	Passed a day ago
✓ veraPDF-pdfbox-validation	# 2	-o- master	b294f70	Passed 2 days ago
✓ veraPDF-library	# 1444	-o- integration	3fdf03b	Passed 2 days ago
✓ veraPDF-xmp	# 25	-o- integration	dde04f7	Passed 3 days ago
✓ veraPDF-pdfbox	# 137	-o- trunk	37ef1a5	Passed 4 days ago
✓ veraPDF-model	# 319	-o- integration	813a484	Passed 6 days ago
✓ veraPDF-model-syntax	# 131	-o- master	ed765a6	Passed 3 months ago
✗ old-veraPDF-pages	# 12	-o- master	9ecbbe	Failed 6 months ago
✓ veraPDF-library-GUI	# 26	-o- master	b8e8432	Passed 9 months ago
✓ veraPDF-framework	# 8	-o- master	91fc9b8	Passed 10 months ago
✓ xttextmodel-mavenized	# 12	-o- master	2244e41	Passed 10 months ago

veraPDF

EU PREFORMA 33

Getting veraPDF software



<http://downloads.verapdf.org/re/verapdf-installer.zip>



Progress Report

- PDF/A validation profiles
- Java API and generic validator
- PDF Parser based Apache PDFBox
- GUI, CLI and REST-WS interfaces
- Full support for PDF/A 1b
- Prototypes for PDF/A 2b, 3b, 1a (80% complete)
- Test corpora for
 - PDF/A 1b, XMP metadata (reviewed)
 - PDF/A 2b, 3b (under review)



35

Community

veraPDF is building community as well as software:

- Sustainability through community
 - Users
 - Developers
- Sustainability through integration
 - DSPACE
 - RODA
 - JHOVE
 - Commercial software vendors: Archivematica, Exlibris, LibNova, Preservica



36

Outreach 2016

- Jan 2016 **Curategear**, Chapel Hill
- Mar 2016 **PASIG Conference**, Prague
- Mar 2016 **UNESCO PERSIST meeting**, Abu Dhabi
- Apr 2016 **PREFORMA Open Source Workshop**, Stockholm
- Jun 2016 **Open Repositories**, Dublin
- Jun 2016 **PDF Days Europe**, Berlin
- Jun 2016 **DLM Forum**, The Hague
- Oct 2016 **iPRES**, Bern



What's next?

- Greenfield PDF parser (GPL/MPL)
- Completion of PDF/A flavours support
- Completion of veraPDF validation corpus
- Plugins for validation and characterisation of embedded files, fonts, images, digital certificates
- Integrations with 3rd party software
- Performance testing and tuning



Get involved!

- Issues
- User requirements
- Documentation
- Test files
- Testing
- Fixes and enhancements



Come and see us

You can come can talk to us during the exhibition after the presentation.

- Get a demo of the veraPDF applications.
- Talk about our plans for the coming year.
- Find out how you can go about integrating veraPDF into your institution's systems or workflows.



Questions?

Stay in touch:

-  <http://verapdf.org/>
-  <http://verapdf.org/subscribe/>
-  https://twitter.com/_verapdf
-  <https://github.com/veraPDF>
-  info@verapdf.org





Initial Presentation

- 1. Consortium & People Involved
- 2. Scope of the Project

• DPF Manager

• TI/A initiative

• Community

INDEX



Dr. MIQUEL MONTANER

CTO at Easy Innova

miquel@easyinnova.com



EASY INNOVA

www.easyinnova.com

Consortium & People Involved



Easy Innova, S.L. (Spain)
Spin-off of the University of Girona

Certified by:



Member of:



Know-how:



Dr. Miquel Montaner
CTO



Robert Sallo
R&D Manager



Xavi Tarrés
Project Manager



Antonio López
Senior Developer



Dr. Víctor Muñoz
R&D Manager



University of Girona (Spain)
Agents Research Lab

Know-how:



Prof. Dr. Josep Lluís de la Rosa
Full Professor and Researcher



Dr. Albert Trias
R&D Manager



University of Basel (Switzerland)
Digital Humanities Lab

Know-how:



Prof. Dr. Lukas Rosenthaler
Full Professor and Researcher



Dr. Peter Fornaro
Managing Director

Scope of the Project

DPF Manager
Open source software
Main features

DPF Manager



Standardization process
TI/A Standard Initiative
Research



Early adopters
Developers
TIFF Experts



INDEX

- Initial presentation
- **DPF Manager**
 - 1. What is DPF Manager?
 - 2. DPF Manager Features
 - 3. How it Works?
 - 4. Use Scenarios
 - 5. DPF Manager current version
 - 6. What's Next?
 - 7. Open Source Project

• T/A initiative

• Community



Xavier Tarrés Bonet

Project Manager at Easy Innova
xavitarres@easyinnova.com



EASY INNOVA

www.easyinnova.com

What is DPF Manager?



DPF Manager is the most advanced TIFF conformance checker for digital preservation

In considering the suitability of particular image for the purposes of digital preservation, relies on the use of a stable, open and well documented file format as well as some data object properties acceptance criteria.

- TIFF Baseline 6.0
- TIFF/EP
- TIFF/IT
 - Tiff/IT P1
 - TIFF/IT P2
- TIFF for Archival (TI/A) **to be accepted as ISO Recommendation**
- Custom policy rules

DPF Manager Features

Multi- platform



Flexible



Modular



Technologies



How it Works?

Select the files to process

1



Files i

Select

Define your acceptance criteria

2

TIFF baseline 6.0
ISO TIFF/EP and ISO TIFF/IT
User defined policy rules

New policy



How it Works?

3

Manage files metadata and apply fixes



Automatically add metadata to files



copyright

▼

PREFORMA

Add new field

4

Get the Report

Multiple report format
Acceptance criteria result
Metadata information
Tiff file structure



palette-1c-8b.tiff

D:\Desktop\DOCS\images\palette-1c-8b.tiff

⚠ This file does NOT conform to conformance checker

	Errors	Warnings
Baseline	0	0
Tiff/EP	41	0
Tiff/IT	1	0
Policy Checker	3	0

Home Screen

New configuration

- Implementation checker
- Policy checker
- Reporter
- Metadata fixer
- Periodical checks
- Summary

Check files

- Multiple file report
- Single file report

View historical data

- List of previous reports

New Configuration > Implementation Checker



- Implementation checker
- Policy checker
- Reporter
- Metadata fixer
- Periodical checks
- Summary

Check files

- Multiple file report
- Single file report

View historical data

- List of previous reports

CONFORMANCE CHECKER | REPORTS | ABOUT

DPF Manager

NEW CONFIGURATION

IMPLEMENTATION CHECKER

Standards

- TIFF / EP (ISO 12234-2:2001)
- TIFF / IP (ISO 12369)
- TIFF / IT-P1 (ISO 12639:1998)
- TIFF / IT-P2 (ISO 12639:2004)
- TIFF / A (To be standarized)

Continue

New Configuration > Policy Checker



- Implementation checker
- Policy checker
- Reporter
- Metadata fixer
- Periodical checks
- Summary

Check files

- Multiple file report
- Single file report

View historical data

- List of previous reports

CONFORMANCE CHECKER | REPORTS | ABOUT

DPF Manager

NEW CONFIGURATION

POLICY CHECKER

Policies

- Default
- 3200px sRGB
- Another custom policy

IMPORT POLICY NEW POLICY

New policy

Color profile: sRGB

Image width: 3200 px

Add rule

Continue

New Configuration > Reporter

- Implementation checker
- Policy checker
- Reporter**
- Metadata fixer
- Periodical checks
- Summary

Check files

- Multiple file report
- Single file report

View historical data

- List of previous reports

The screenshot shows the 'NEW CONFIGURATION' screen for the 'Reporter' configuration. At the top, there's a navigation bar with links for 'CONFORMANCE CHECKER', 'REPORTS', and 'ABOUT'. Below the navigation is the 'DPF Manager' logo. The main area is titled 'NEW CONFIGURATION' and contains several sections:

- Format:** Options for output format include HTML (with a dropdown menu 'Select HTML template'), Expert mode, PDF (with a dropdown menu 'Select PDF template'), Expert mode, XML, and JSON.
- Check files:** Buttons for 'Multiple file report' and 'Single file report'.
- View historical data:** A link to 'List of previous reports'.
- Buttons:** A large blue 'Continue' button at the bottom center.

New Configuration > Metadata Fixer

- Implementation checker
- Policy checker
- Reporter
- Metadata fixer**
- Periodical checks
- Summary

Check files

- Multiple file report
- Single file report

View historical data

- List of previous reports

The screenshot shows the 'NEW CONFIGURATION' screen for the 'Metadata fixer' configuration. The interface is similar to the Reporter configuration, with a 'CONFORMANCE CHECKER', 'REPORTS', and 'ABOUT' navigation bar at the top. The 'DPF Manager' logo is present. The main area is titled 'NEW CONFIGURATION' and includes the following sections:

- Format:** Options for output format include HTML (with a dropdown menu 'Select HTML template'), Expert mode, PDF (with a dropdown menu 'Select PDF template'), Expert mode, XML, and JSON.
- Check files:** Buttons for 'Multiple file report' and 'Single file report'.
- View historical data:** A link to 'List of previous reports'.
- Metadata Fixer settings:** Includes checkboxes for 'Automatically fix metadata errors' (with a note about removing private data) and 'Automatically add metadata to files' (with a dropdown menu 'Copyright' and a 'PREFORMA project' field).
- Buttons:** A large blue 'Continue' button at the bottom center.

New Configuration > Periodical Checks

New configuration

- Implementation checker
- Policy checker
- Reporter
- Metadata fixer
- Periodical checks**
- Summary

Check files

- Multiple file report
- Single file report

View historical data

- List of previous reports

The screenshot shows the 'NEW CONFIGURATION' interface with the 'PERIODICAL CHECKS' tab selected. It includes fields for selecting a check frequency ('Run this check once' or 'Run this check every: Week') and a specific day ('Tuesday'). A time selector shows '12:00 h'. An email checkbox is checked, pointing to 'admin@archive.org'. A large blue 'Continue' button is at the bottom.

New Configuration > Summary

New configuration

- Implementation checker
- Policy checker
- Reporter
- Metadata fixer
- Periodical checks
- Summary

Check files

- Multiple file report
- Single file report

View historical data

- List of previous reports

The screenshot shows the 'CONFIGURATION SUMMARY' section. It lists the configured tools: 'Implementation checker' (TIFF/EP ISO 12234-2:2001), 'Policy checker' (Policy: 3200px aRGB, Color profile: + sRGB, Image width: = 3200px), 'Report' (Format: HTML / Default template, expert mode, XML), and 'Metadata fixer' (Fix errors automatically, Remove fields, Author, Date, Add fields, Copyright = PREFORMA project). At the bottom, it shows 'Periodical Checks' set to 'Every week (Tuesday, Thursday at 12:00 h)' and 'Send report by email to: admin@archive.org'. A large blue 'Save configuration' button is at the bottom.

Home Screen

New configuration

- Implementation checker
- Policy checker
- Reporter
- Metadata fixer
- Periodical checks
- Summary

Check files

- Multiple file report
- Single file report

View historical data

- List of previous reports

The screenshot shows the DPF Manager Conformance Checker interface. At the top, there's a navigation bar with links for 'CONFORMANCE CHECKER', 'REPORTS', and 'ABOUT'. Below the navigation is the 'DPF Manager' logo. The main area is titled 'CONFORMANCE CHECKER'. It features a 'Files' section with a text input field containing 'localhost:8888/path/file.tif' and a 'Select' button. Below this is a 'Configuration' section with three options: 'Default', 'B&W TIFF / IT', and 'Another custom configuration'. At the bottom of this section are 'IMPORT CONFIGURATION' and 'NEW CONFIGURATION' buttons. A large blue button labeled 'Check files' is centered at the bottom.

Check Files > Multiple File Report

New configuration

- Implementation checker
- Policy checker
- Reporter
- Metadata fixer
- Periodical checks
- Summary

Check files

- Multiple file report
- Single file report

View historical data

- List of previous reports

The screenshot shows the DPF Manager Multifile Report screen. At the top, it displays '6 files processed'. Below this is a summary table:

4 conform to TIFF / EP and policy checker	1 conforms to TIFF / EP only	1 conforms to policy checker only
---	------------------------------	-----------------------------------

A pie chart indicates '4 passed' (75%) and '2 failed' (25%). Below the summary are six individual file reports:

File Name	Type	Errors	Warnings	Checker	Status	Score
filename_code_num1.tif	TIFF/EP	2	1	Policy checker	Failed	75%
filename_code_num2.tif	TIFF/EP	0	0	Policy checker	Passed	100%
filename_code_num3.tif	TIFF/EP	0	0	Policy checker	Failed	55%
filename_code_num4.tif	TIFF/EP	0	2	Policy checker	Passed with warnings	100%
filename_code_num5.tif	TIFF/EP	0	0	Policy checker	Passed	100%
filename_code_num6.tif	TIFF/EP	0	0	Policy checker	Passed	100%

At the bottom right is a 'Back' button.

Check Files > Single File Report

New configuration

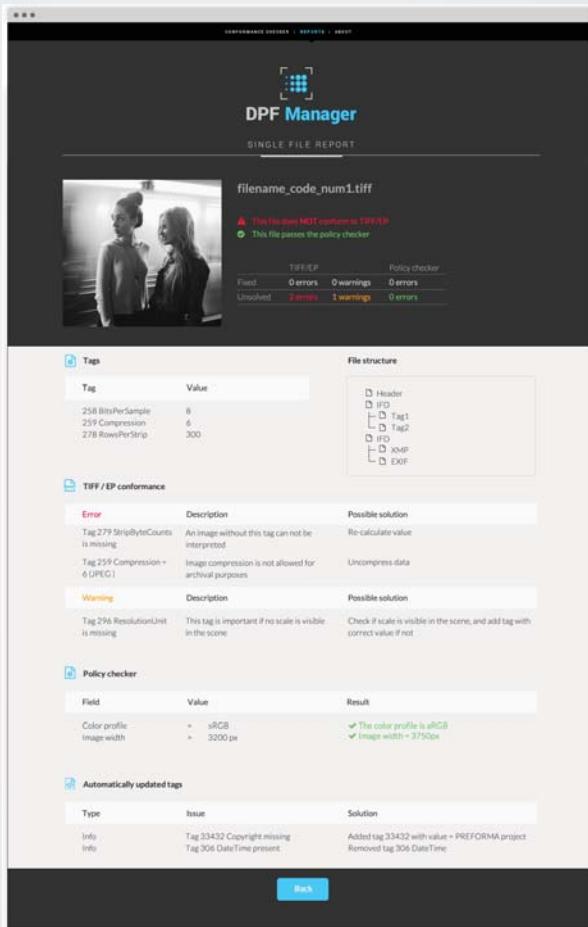
- Implementation checker
- Policy checker
- Reporter
- Metadata fixer
- Periodical checks
- Summary

Check files

- Multiple file report
- Single file report**

View historical data

- List of previous reports



The screenshot shows the DPF Manager interface for a single file report. At the top, it displays the file name "filename_code_num1.tiff". Below this, there are two status indicators: a red triangle pointing down labeled "File does NOT conform TIFF/EP" and a green circle labeled "This file passes the policy checker". A summary table shows error counts: Fixed 0 errors, Unresolved 2 errors, 0 warnings, 0 errors. The main content area is divided into several sections:

- Tags:** Shows tag values like 256 BitsPerSample (8), 259 Compression (6), and 278 RowsStrip (300).
- File structure:** A tree view showing the file structure: Header → IFD → Tag1 → Tag2; IFD → XMP; EXIF.
- TIFF / EP conformance:** A table of errors and warnings:

Error	Description	Possible solution
Tag 279 StripByteCounts is missing	An image without this tag can not be interpreted.	Re-calculate value
Tag 259 Compression + 6 (JPEG)	Image compression is not allowed for archival purposes	Uncompress data

Warning	Description	Possible solution
Tag 294 ResolutionUnit is missing	This tag is important if no scale is visible in the scene.	Check if scale is visible in the scene, and add tag with correct value if not
- Policy checker:** A table showing field values and results:

Field	Value	Result
Color profile ->	sRGB	✓ The color profile is sRGB
Image width ->	3200 px	✓ Image width = 3750px
- Automatically updated tags:** A table showing info tags and their issues:

Type	Issue	Solution
Info	Tag 33432 Copyright missing	Added tag 33432 with value = PREFORMA project
Info	Tag 306 DateTime present	Removed tag 306 DateTime

Report History

New configuration

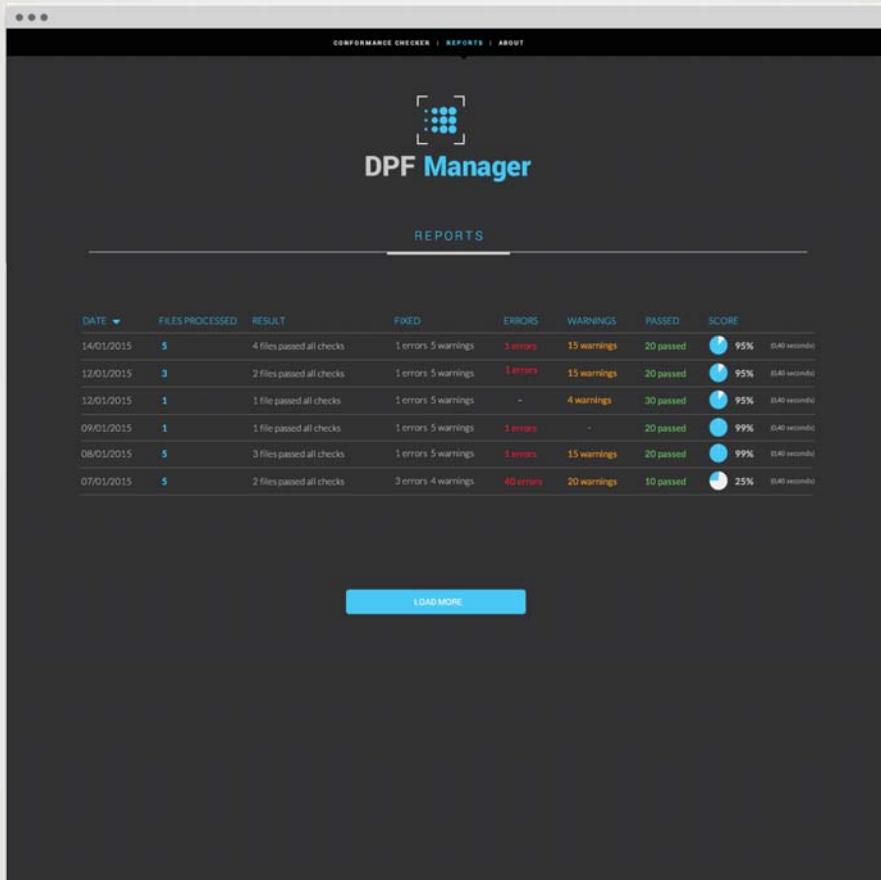
- Implementation checker
- Policy checker
- Reporter
- Metadata fixer
- Periodical checks
- Summary

Check files

- Multiple file report
- Single file report

View historical data

- List of previous reports



The screenshot shows the DPF Manager interface for reports. At the top, it displays the word "REPORTS". The main content area is a table showing a history of processed files:

DATE	FILES PROCESSED	RESULT	FIXED	ERRORS	WARNINGS	PASSED	SCORE	TIME
14/01/2015	5	4 files passed all checks	1 errors 5 warnings	3 errors	15 warnings	20 passed	95%	0:40 seconds
12/01/2015	3	2 files passed all checks	1 errors 5 warnings	3 errors	15 warnings	20 passed	95%	0:40 seconds
12/01/2015	1	1 file passed all checks	1 errors 5 warnings	-	4 warnings	30 passed	95%	0:40 seconds
09/01/2015	1	1 file passed all checks	1 errors 5 warnings	3 errors	-	20 passed	99%	0:40 seconds
08/01/2015	5	3 files passed all checks	1 errors 5 warnings	3 errors	15 warnings	20 passed	99%	0:40 seconds
07/01/2015	5	2 files passed all checks	3 errors 4 warnings	40 errors	20 warnings	10 passed	25%	0:40 seconds

At the bottom, there is a blue button labeled "LOAD MORE".

Use Scenarios

1

Standalone



Graphic Interface

```
C:\>dpfmanager.exe -  
files=archive/*.tiff -  
config=config.xml  
-reports_folder=/tmp/reports -  
fixed_files_folder=/tmp/files
```

Command line

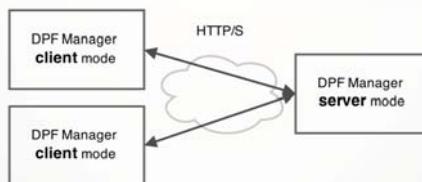
2

Client - Server



```
client:\>dpfmanager.exe  
-server=80.45.32.45  
-port=80
```

Client



```
server:\>dpfmanager.exe  
-mode=server
```

Server

3

Web application



Browser Client



```
server:\>dpfmanager.exe  
-mode=server
```

Server

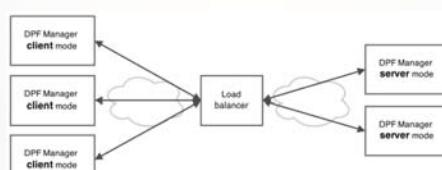
4

High availability



```
client:\>dpfmanager.exe  
-server=80.45.32.45  
-port=80
```

Client



```
server:\>dpfmanager.exe  
-mode=server
```

Server

Use Scenarios

Integration with OAIS

Document lifecycle
Conformance checking at:

- Creation
- Transfer
- Migration
- Digitalization

Reference framework
QA, AIP generation



DPF Manager current version

VERSION 2.0 – Released on April 4th 2016

- **Conformance checker**
 - Baseline version 6.0
 - TIFF/EP
 - TIFF/IT
 - TIFF/IT P1
 - TIFF/IT P2
- **Policy checker**
 - Create Policy rules from GUI
 - Extend developing your own policies
- **Metadata fixer**
 - Add, edit or remove metadata
 - Apply auto-fixes in TIFF files
 - Extend developing your own auto-fixes
- **Interfaces**
 - Command-line interface
 - Graphical user interface
- **Modular architecture**
 - Develop your own modules for DPF Manager



DPF Manager

What's Next?

In the next VERSIONS...

<https://github.com/EasyinnovaSL/DPFManager/milestones>

- **Conformance checker**
 - TI/A (TIFF for Archival)
- **Policy checker**
 - New policies to apply
- **Metadata fixer**
 - New metadata
 - New auto-fixes in TIFF files
- **Interfaces**
 - Server interface
 - Web interface
 - Schedule checks
- **Conformance checkers interoperability**



DPF Manager

Open Source Project



Open source licenses



MPL v2+



GPL v3+

Open source project

<http://dpfmanager.org/>



<https://github.com/EasyinnovaSL/DPFManager>

<http://preforma-project.eu/dpf-manager.html>

INDEX

- Initial Presentation
- DPF Manager
- TI/A Initiative**
 - File Format Preservation
 - TIFF Format
 - TI/A Standard initiative
 - Timeline
- Community



Dr. PETER FORNARO

Managing Director at University Basel
peter.fornaro@unibas.ch



UNIVERSITY OF BASEL

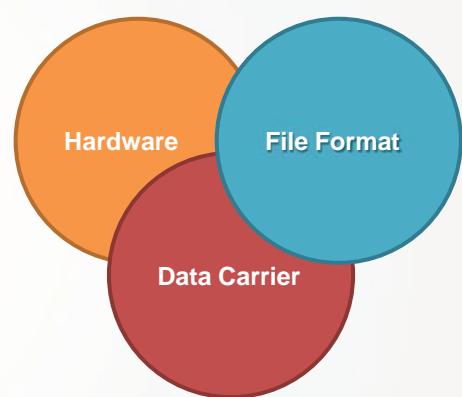
www.unibas.ch



File Format Preservation



- Even if bit-stream preservation was successful (hardware, data-carrier), the format of the files must be decoded to render the content properly
- A format is good for archiving if it is widely used, well documented, open and allows storage of appropriate quality of the content



Migration not necessary

TIFF Format

It is there!

- Tiff was the major file (digital master) for archival purposes
 - TIFF is a final rendered image
 - TIFF is 16bit
 - TIFF is lossless
 - TIFF is professional
 - TIFF is multichannel and multilayer
- Many archives and museums store TIFF files
 - because the others do it
 - because it is of large data volume
 - because it is widely used and well documented

88% Memory Institutions



Standardization

- TTIF-Files will be checked for "correctness".
- TIFF is a complex standard with a lot of possible features. It is not clear to say what "correctness" means. It can depend on the application.
- However: It is necessary to define and recommend a well-defined set of features that are accepted by most memory institutions.



However: It is necessary to define and

TI/A Standard Initiative

The approach is well-known:
The TIFF for Image for Archival (TIFF for Archives) standard initiative is a group of experts focussing on the creation of a recommendation for the use of the TIFF Format.

TIFF is widely used by memory institutions to preserve their digital assets. This initiative has the goal to create a new ISO Recommendation to define a standard for the qualification for archival purposes. We are looking for experts to participate in the TI/A Standard definition. Why don't you join us?

- A PDF is a full featured digital document, some of them are not wanted in archives. Some features are proprietary and therefore not suitable for long-term archival purposes.
- A PDF/A is an adopted version, fulfilling archival needs. Some features are forbidden some are required.
- We follow a similar approach: TIFF will be precisely specified based on a feedback of a expert user community and a survey on „hot“ data.

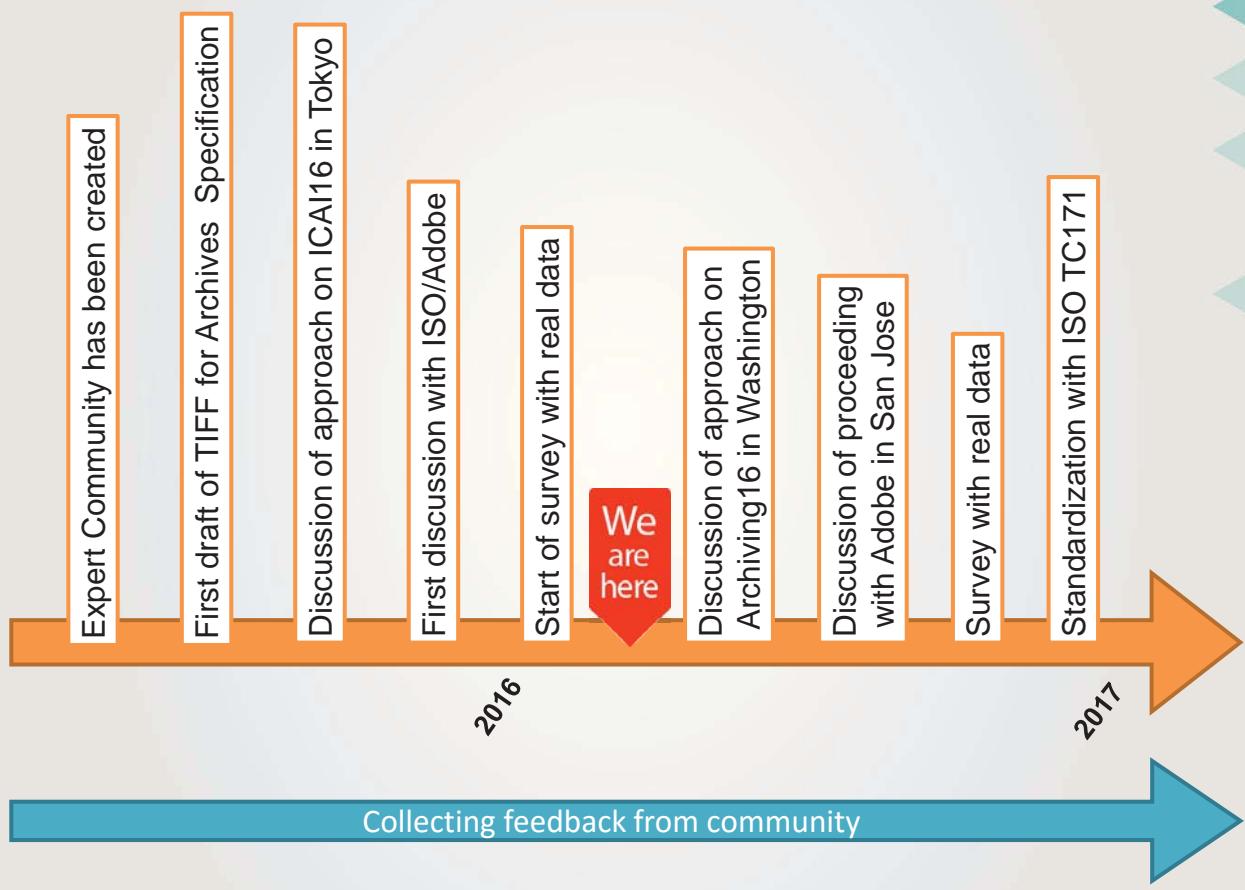
Recommendation of TIFF for Archives is in progress

Well specified set of mandatory, forbidden and optional tags

The versatility of the TIFF format has made it very attractive for memory institutions for long term archival of their digital images. However, since the TIFF format offers such a wide range of possibilities, it is important to have a well specified set of tags to ensure that the information is always available and can be read by different systems. This is why the standardization process has started in 2016.

The limitations of the baseline TIFF are too severe for many applications in digital archiving. It is important that, besides crucial technical metadata such as ICC color profiles (in case of color images) also important descriptive metadata is stored within the image file. Having descriptive metadata available (such as content description, iconography, copyright and ownership information etc.) is crucial for every archive. Having this information in the same file as the image data guarantees that this information will always be associated with the image.

Timeline



INDEX

- Initial Presentation
 - DPF Manager
 - TI/A Standard Initiative
 - # Community

 1. Registered Early Adopters
 2. TIFF Experts Involved
 3. Join us!



Prof. JOSEP LLUÍS de la ROSA

Full Professor at University of Girona
pepluis@eia.udg.edu

Registered Early Adopters



Oregon State
UNIVERSITY



Aquaforest FUNDACIÓANTONITÀPIES



More than **45.000 images** checked

Feedback from more than **2.500 reports**



DPF Manager
@DPFManager

TWEETS
353

FOLLOWING
355

FOLLOWERS
114

TIFF Experts Involved



[Home](#) [Why TI/A?](#) [The Initiative](#) [Get Involved](#) [About us](#) [Twitter](#) [Open](#)

TI/A Standard Initiative

The TI/A (Tagged Image for Archival) standard initiative is a group of experts focussing on the creation of a recommendation for the use of the TIFF Format, TI/A. TIFF is widely used by memory institutions to preserve their digital

69 experts in TI/A intranet from **16 countries**

support of **61 memory institutions** to create the TI/A standard

[Join us](#)



TI/A Initiative
@TI_A_Standard

TWEETS
643

FOLLOWING
735

FOLLOWERS
318

Join us!

visit
www.dpfmanager.org

Early adopters



Install now
For Windows, Mac OS X and Linux

Version 2.0 - Release (04-04-2016)

visit
www.ti-a.org

Tiff experts



Get involved!

visit
<http://www.dpfmanager.org/community.html>
<https://github.com/EasyInnovaSL/DPFManager>

Developers



New pull request

Thanks for your attention!
FEEL FREE TO ASK US



www.easyinnova.com



MediaConch

Implementation and policy checking
on FFV1, Matroska, LPCM, and more



Jérôme Martinez
MediaArea.net SARL

Open Source Preservation Workshop - April 2016



What is MediaConch?

MediaConch is a conformance checker

- Implementation checker
- Policy checker
- Reporter
- Fixer

What is MediaConch?

Implementation and Policy reporter

The screenshot shows the MediaConch web application. At the top, there's a navigation bar with links for 'Checker', 'Result', 'Policies', 'Display', and 'Help'. Below the navigation is a table titled 'Close all results' which lists five media files along with their analysis results. The columns are 'Files', 'Implementation', 'Policy', 'MediaInfo', 'MediaTrace', and 'Status'. Each row contains a file name, its validation status (Valid or Invalid), policy results (e.g., 'Is Matroska'), and a trash bin icon for deletion.

Files	Implementation	Policy	MediaInfo	MediaTrace	Status
Coconut.mp4	✓ Valid	⌚⌚ ✗ Is Matroska	⌚⌚	⌚⌚	Analyzed ⚡
Exampleä.mp4	✓ Valid	⌚⌚ ✗ Is Matroska	⌚⌚	⌚⌚	Analyzed ⚡
ffv1_3 - Copie (2).m...	✓ Valid	⌚⌚ ✓ Is Matroska	⌚⌚	⌚⌚	Analyzed ⚡
ffv1_3 - Copie.mkv	✓ Valid	⌚⌚ ✓ Is Matroska	⌚⌚	⌚⌚	Analyzed ⚡
ffv1_3.mkv	✓ Valid	⌚⌚ ✓ Is Matroska	⌚⌚	⌚⌚	Analyzed ⚡

What is MediaConch?

Example of report

This screenshot displays a detailed report for an MKV/FFV1 file. The report is structured with blue header sections and white content sections. It starts with a general summary and then breaks down specific context and track details.

MKV/FFV1 (reformatted QT/v210)
Policy check for Matroska-wrapped FFV1 and LPCM file, transcoded from QuickTime-wrapped v210 and LPCM file.

General Format equals Matroska

Context (field): Format

Context (value): Matroska

tracktype	tracktypeorder	trackid	actual	outcome	reason
General			Matroska	✓ pass	

Video Width is equal to 720 (pixels)

Context (field): Width

Context (value): 720

tracktype	tracktypeorder	trackid	actual	outcome	reason
Video	1	640	X fail		is not equal

What is MediaConch?

General information about your files

Key	Value
↳ C:/Programmation/PreFormaMediaInfo/SampleTestFiles/FFV1/ffv1_3.mkv	
↳ General	
↳ UniqueID	88323790047680325859674626238128084708
↳ Format	Matroska
↳ Format_Version	4
↳ FileSize	126167
↳ Duration	1.000
↳ OverallBitRate	1009336
↳ FrameRate	25.000
↳ FrameCount	25
↳ StreamSize	2511
↳ Video	
↳ StreamOrder	0
↳ ID	1
↳ UniqueID	1
↳ Format	FFV1
↳ Format_Version	3.4
↳ CodecID	V_MS/VFW/FOURCC / FFV1
↳ Duration	1.000
↳ BitRate	989250
↳ Width	320

What is MediaConch?

Inspect your files

Offset	Key	Value
0x00000000	↳ EBML (47 bytes)	
0x0000002f	↳ Segment (126120 bytes)	
0x0000002f	↳ Header (12 bytes)	
0x0000003b	↳ SeekHead (66 bytes)	
0x0000007d	↳ Void (157 bytes)	
0x0000011a	↳ Info (81 bytes)	
0x0000016b	↳ Tracks (167 bytes)	
0x0000016b	↳ Header (12 bytes)	
0x00000177	↳ TrackEntry (155 bytes)	
0x00000177	↳ Header (9 bytes)	
0x00000180	↳ TrackNumber - 1 (3 bytes)	
0x00000183	↳ TrackUID - 1 (4 bytes)	
0x00000187	↳ FlagLacing - 0 (3 bytes)	
0x0000018a	↳ Language - eng (7 bytes)	
0x00000191	↳ TrackType - 1 (3 bytes)	
0x00000194	↳ DefaultDuration - 40000000 (8 bytes)	
0x0000019c	↳ CodecID - V_MS/VFW/FOURCC (17 bytes)	
0x000001ad	↳ Video (16 bytes)	
0x000001bd	↳ CodecPrivate - Copy of vids (85 bytes)	
0x000001bd	↳ Header (3 bytes)	
0x000001c0	↳ Size	81 (0x51)

What is MediaConch?

Policy editor

To create policies, users may select fields from MediaInfo's stream types (General, Video, Audio, etc.) and apply validators (Equal, Not Equal, Greater Than, etc.) to the associated field values. For example, a policy item may state that the "Width" field from the file's Video stream equal 640 pixels. When adding the field's value, however, be sure to leave off any associated strings, such as "pixels" "Kbps" "fps", etc.

Rule name *

Quicktime wrapper
Editor type : Editor Free text

Track type *

General

Field *

Format_Profile

Occurrence *

*

Validator *

Is equal (==)

Value

QuickTime

CAVPP Preservation Video Files rules

Quicktime wrapper		
10 bit uncompressed 4:2:2 video		
Total Bitrate		
Frame Rate 29.970		
Aspect Ratio 4/3		
Is Interlaced		
Audio is 48kHz		
Audio is stereo		
Audio is 24 bit		

Add a new rule

MediaConch interfaces

- Graphical interface
- Web interface
- Command line
- Server (REST API)
- (Work in progress) a library (.dll/.so/.dylib)

MediaConch output formats

- XML (native format)
- Text
- HTML
- (Work in progress) PDF
- Tweakable! (with XSL)

Open source

- GPLv3+ and MPLv2+
- Relies on MediaInfo (metadata extraction tool)
- Use well-known open source libraries: Qt, sqlite, libevent, libxml2, libxslt, libexslt...

Supported formats

- Priorities for the implementation checker
 - Matroska
 - FFV1
 - PCM
- Can accept any format supported by MediaInfo for the policy checker
 - MXF + JP2k
 - QuickTime/MOV
 - Audio files (WAV, BWF, AIFF...)
 - ...

Supported formats

Can be expanded

- By plugins
 - Support of PDF checker: VeraPDF plugin
 - Support of TIFF checker: DPF Manager plugin
 - You use another checker? Let us know
- By internal development
 - More tests on your preferred format is possible
 - It depends on you!

Versatile

Several input formats are accepted

- FFV1 from MOV or AVI
- Matroska with other video formats
- (Work in progress) Extraction of a PDF or TIFF attachment from a Matroska container and analyze with a plugin (e.g. VeraPDF and DPF Manager)
- ...

Versatile

Input can be from:

- Files (local/network)
- FTP/FTPS/SFTP
- HTTP/HTTPS
- Amazon S3

Versatile

Binaries are provided for:

- Windows
- Mac
 - Homebrew users: "brew install mediaconch", that's all!
- Linux (Ubuntu, Debian, Fedora, OpenSUSE...)
 - Ubuntu 16.04 and Debian Testing/9 users:
"apt-get install mediaconch", that's all!
(it is in the official distros repository)
- Embedded devices? Doable
 - (we tested it on a Raspberry Pi 
- Can be ported on other distros (BSD...)

Standardization

- Matroska is widely used but not (yet) standardized
- FFV1 is gaining increasing usage in preservation contexts but is not (yet) standardized

CELLAR: IETF workgroup

- Open standards group
- Goal to IETF-standardize Matroska/FFV1/FLAC
- A lot of progress, especially with Matroska/EBML specs
- <https://datatracker.ietf.org/wg/cellar/charter/>

Worldwide

- 2 project leaders
 - Jérôme Martinez (Digital Media Analysis Specialist, France)
 - Dave Rice (Archivist, USA)
- Presentations worldwide
 - IASA, France
 - FIAT/IFTA, Austria
 - FOSDEM, Belgium
 - AMIA, USA
 - Code4Lib, USA
 - JTS, Singapore
 - (3-6 October 2016) IPRES, Switzerland
 - (25-29 September 2016) IASA, USA

Matroska research corpus

- We analyze all Matroska files from archive.org
 - Interface with some statistics of Matroska elements usage
(e.g. files with CRC-32 elements...)
- <https://mediaarea.net/MediaConchCorpus/>

What's next?

Still under development but already usable
(PREFORMA prototyping phase up to end 2016)

- Better handling of huge collections
- Better user interface
- Statistics
- Standardize Matroska and FFV1
- More conformance tests
- Integration in Archivematica
- Fixer

And after PREFORMA sponsorship?

It depends on you!

- This is open source
- Driven by user requests
- Everyone can develop or sponsor a development
- Potential features:
 - Integration of test of your preferred format (MXF? doable. JP2k? doable. WAV? doable...)
 - Integration of other checkers (BWF MetaEdit? QCTools?)
 - Better integration in your workflow
 - ...

Example (Online)

The screenshot shows the MediaConch web application interface. At the top, there is a navigation bar with links for Home, Checker, Policies, Display, Help, and a user session indicator. Below the navigation bar, there is a search bar with placeholder text "Search: []".

The main content area has a title "Check files" and three tabs: "Check by file upload" (selected), "Check online files", and "Check server files". Below the tabs, there are dropdown menus for "Policy" (set to "General Conformance") and "Display" (set to "MediaConch Html"). A large orange button labeled "Check files" is positioned below these controls.

Below the controls, there is a section titled "Results" with a table. The table has columns: Files, Implementation, Policy, MediaInfo, MediaTrace, Status, and a trash bin icon. The table lists several media files:

Files	Implementation	Policy	MediaInfo	MediaTrace	Status	Action
ffv1_0.mov			⊕ ⊕	⊕ ⊕	Analyzed	trash
ffv1_test_pixfmt-yuva444p_coder-1_...			⊕ ⊕	⊕ ⊕	Analyzed	trash
ffv1_test_pixfmt-bgr0_coder-0_level...			⊕ ⊕	⊕ ⊕	Analyzed	trash
ffv1_test_pixfmt-gbrp14le_coder-1_...			⊕ ⊕	⊕ ⊕	Analyzed	trash
ffv1_0.mkv			⊕ ⊕	⊕ ⊕	Analyzed	trash
ffv1_test_pixfmt-yuv422p_coder-0_le...			⊕ ⊕	⊕ ⊕	Analyzed	trash
ffv1_test_pixfmt-yuv420p_coder-0_le...					In queue	trash

Example (Command line)

```
EBML Element=0x45A3, Allowed EBML Parent Element=0x67C8, EBML Parent Element=0x67C8, Outcome: ✓ pass
EBML Element=0x4487, Allowed EBML Parent Element=0x67C8, EBML Parent Element=0x67C8, Outcome: ✓ pass
EBML Element=0x7373, Allowed EBML Parent Element=0x1254C367, EBML Parent Element=0x1254C367, Outcome: ✓ pass
EBML Element=0x63C0, Allowed EBML Parent Element=0x7373, EBML Parent Element=0x7373, Outcome: ✓ pass
EBML Element=0x63C5, Allowed EBML Parent Element=0x63C0, EBML Parent Element=0x63C0, Outcome: ✓ pass
EBML Element=0x67C8, Allowed EBML Parent Element=0x7373, EBML Parent Element=0x7373, Outcome: ✓ pass
EBML Element=0x45A3, Allowed EBML Parent Element=0x67C8, EBML Parent Element=0x67C8, Outcome: ✓ pass
EBML Element=0x4487, Allowed EBML Parent Element=0x67C8, EBML Parent Element=0x67C8, Outcome: ✓ pass
EBML Element=0x1F43B675, Allowed EBML Parent Element=0x1853B067, EBML Parent Element=0x1853B067, Outcome: ✓ pass
EBML Element=0xE7, Allowed EBML Parent Element=0x1F43B675, EBML Parent Element=0x1F43B675, Outcome: ✓ pass
EBML Element=0xA3, Allowed EBML Parent Element=0x1F43B675, EBML Parent Element=0x1F43B675, Outcome: ✓ pass
EBML Element=0xA3, Allowed EBML Parent Element=0x1F43B675, EBML Parent Element=0x1F43B675, Outcome: ✓ pass
EBML Element=0xA3, Allowed EBML Parent Element=0x1F43B675, EBML Parent Element=0x1F43B675, Outcome: ✓ pass
EBML Element=0xA3, Allowed EBML Parent Element=0x1F43B675, EBML Parent Element=0x1F43B675, Outcome: ✓ pass
EBML Element=0xA3, Allowed EBML Parent Element=0x1F43B675, EBML Parent Element=0x1F43B675, Outcome: ✓ pass
EBML Element=0xA3, Allowed EBML Parent Element=0x1F43B675, EBML Parent Element=0x1F43B675, Outcome: ✓ pass
EBML Element=0xC53B866B, Allowed EBML Parent Element=0x1853B067, EBML Parent Element=0x1853B067, Outcome: ✓ pass
EBML Element=0xB8, Allowed EBML Parent Element=0xC53B866B, EBML Parent Element=0xC53B866B, Outcome: ✓ pass
EBML Element=0xB3, Allowed EBML Parent Element=0xB8, EBML Parent Element=0xB8, Outcome: ✓ pass
EBML Element=0xB7, Allowed EBML Parent Element=0xB8, EBML Parent Element=0xB8, Outcome: ✓ pass
EBML Element=0xF7, Allowed EBML Parent Element=0xB7, EBML Parent Element=0xB7, Outcome: ✓ pass
EBML Element=0xF1, Allowed EBML Parent Element=0xB7, EBML Parent Element=0xB7, Outcome: ✓ pass
EBML Element=0xF0, Allowed EBML Parent Element=0xB7, EBML Parent Element=0xB7, Outcome: ✓ pass
EBML Element=0xB8, Allowed EBML Parent Element=0x1C53B866B, EBML Parent Element=0x1C53B866B, Outcome: ✓ pass
EBML Element=0xB3, Allowed EBML Parent Element=0xB8, EBML Parent Element=0xB8, Outcome: ✓ pass
EBML Element=0xB7, Allowed EBML Parent Element=0xB8, EBML Parent Element=0xB8, Outcome: ✓ pass
EBML Element=0xF7, Allowed EBML Parent Element=0xB7, EBML Parent Element=0xB7, Outcome: ✓ pass
EBML Element=0xF1, Allowed EBML Parent Element=0xB7, EBML Parent Element=0xB7, Outcome: ✓ pass
EBML Element=0xF0, Allowed EBML Parent Element=0xB7, EBML Parent Element=0xB7, Outcome: ✓ pass
EBML ELEMENT-NONMULTIPLES
Non-Repeating Element=0x42B6, Parent Element=0x1A45DFA3, Outcome: ✓ pass
Non-Repeating Element=0x42F7, Parent Element=0x1A45DFA3, Outcome: ✓ pass
Non-Repeating Element=0x42F2, Parent Element=0x1A45DFA3, Outcome: ✓ pass
Non-Repeating Element=0x42F3, Parent Element=0x1A45DFA3, Outcome: ✓ pass
Non-Repeating Element=0x42B2, Parent Element=0x1A45DFA3, Outcome: ✓ pass
Non-Repeating Element=0x42B7, Parent Element=0x1A45DFA3, Outcome: ✓ pass
Non-Repeating Element=0x42B8, Parent Element=0x1A45DFA3, Outcome: ✓ pass
Non-Repeating Element=0x53AB, Parent Element=0x4DBB, Outcome: ✓ pass
Non-Repeating Element=0x53AC, Parent Element=0x4DBB, Outcome: ✓ pass
```

Example (Plugins)

Check files

Check by file upload Check online files Check server files

Policy Policy Set Example Display Choose a display Verbosity Default level Check files

Results

Show	10	entries	Search:			
Files	Implementation	Policy	MediaInfo	MediaTrace	Status	
ffv1_test_pixfmt-yuv444p10le_coder-...	✓ Valid	⌚ ⓘ ✓ Policy Set Example	⌚ ⓘ	⌚ ⓘ	Analyzed	
ffv1_test_pixfmt-yuva422p_coder-1_1...	✓ Valid	⌚ ⓘ ✓ Policy Set Example	⌚ ⓘ	⌚ ⓘ	Analyzed	
ffv1_test_pixfmt-yuva444p_coder-1_1...	✓ Valid	⌚ ⓘ ✓ Policy Set Example	⌚ ⓘ	⌚ ⓘ	Analyzed	
veraPDF test suite 6-1-10-t01-pass-...	✓ Valid	⌚ ⓘ N/A	⌚ ⓘ	⌚ ⓘ	Analyzed	
train1.tif	✗ Not valid	⌚ ⓘ	⌚ ⓘ	⌚ ⓘ	Analyzed	

Showing 11 to 15 of 15 entries Previous 1 2 Next

Example (Plugins)



Implementation report

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<cliReport xmlns="http://www.verapdf.org/ValidationProfile">
    <itemDetails size="41437">
        <name>veraPDF test suite 6-1-10-t01-pass-a.pdf</name>
    </itemDetails>
    <validationResult flavour="PDFA_1_B" totalAssertions="476" isCompliant="true">
        <assertions/>
    </validationResult>
</cliReport>
```

Download implementation report Close

Files	Implementation	Policy	MediaInfo	MediaTrace	Status
ffv1_test_pixfmt-yuv444p10le_coder-...	✓ Valid	✓ Policy Set Example	✓	✓	Analyzed
ffv1_test_pixfmt-yuva422p_coder-1_1...	✓ Valid	✓ Policy Set Example	✓	✓	Analyzed
ffv1_test_pixfmt-yuva444p_coder-1_1...	✓ Valid	✓ Policy Set Example	✓	✓	Analyzed
veraPDF test suite 6-1-10-t01-pass-...	✓ Valid	N/A	✓	✓	Analyzed
train1.tif	✗ Not valid	N/A	✓	✓	Analyzed

Example (Plugins)



Implementation report

```
<?xml version="1.0" encoding="UTF-8"?>
<globalreport>
    <individualreports>
        <report>
            <file_info>
                <name>train1.tif</name>
                <fullpath>train1.tif</fullpath>
            </file_info>
            <tiff_structure>
                <ifdTree>
                    <ifdNode>
                        <number>0</number>
                        <isimg>yes</isimg>
                        <imagetype check_ifd0="typ">Main image</imagetype>
                        <image_representation>strips[1]</image_representation>
                        <photometric>0</photometric>
                        <hasSubIfd>no</hasSubIfd>
                        <hasExif>yes</hasExif>
                        <hasXMP>yes</hasXMP>
                        <hasIPTC>yes</hasIPTC>
                        <tags>
                            <tag>
                                <name>NewSubfileType</name>
                                <id>254</id>
                                <value>0</value>
                            </tag>
                        </tags>
                    </ifdNode>
                </ifdTree>
            </tiff_structure>
        </report>
    </individualreports>
</globalreport>
```



Stay in touch

MediaArea: <https://mediaarea.net>, @MediaArea_net

MediaConch: <https://mediaarea.net/MediaConch>, @MediaConch

Jérôme Martinez: jerome@mediaarea.net

Slides: <https://mediaarea.net/Events>

License: CC BY