PROTOTYPING PHASE 2 INTERMEDIATE REPORT

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

MediaConch

Revision: [draft, final]

Authors:

Jérôme Martinez (MediaArea.net)
Dave Rice (MediaArea.net)
Ashley Blewer (MediaArea.net)
Natalie Cadranel (MediaArea.net)

Dissemination Level		
P	Public	X

1 INTRODUCTION

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

Following the same approach used last year, during the Second Prototyping Phase the plan for releases is as follows:

- Frequent releases: monthly;
- Intermediate releases: end of July 2016 and end of October 2016.

The intermediate release shall contain two parts:

- A functionally stable release, if possible even more organised release compared with the respective predecessor versions
- A report which
    - Describes
        - More in detail the respective release;
        - The time line along with the current position (on time, delayed, ahead)
        - How suppliers managed to provide the required functionality (so far);
        - What is still missing compared to the original specifications and which is the plan to implement it.
    - Provides basic information to be used by PREFORMA WP8 in their deliverables to be submitted to the EC, reporting the work done by both suppliers and PREFORMA consortium members during the prototyping phase.

2

PROTOTYPING PHASE 2 - INTERMEDIATE REPORT

1. Details
Type of Organisation: SARL
Registered Name of Organisation: MediaArea.net
Registered Address: Chemin du Vernay
Town/ City: Curienne
Postcode: 73190
County:
Country: France
Report Author: Jérôme Mertinez
Telephone Number: +33 (0)6.12.65.98.37
E-mail Address: jerome@mediaarea.net
Project Name: MediaConch
Report Type: Prototyping Phase 2 Intermediate Report
Total Contract Price [euro]: 700 000 €
Start Date: April 14, 2015
End Date: December 31, 2016
Sub-contractors: Dave Rice, Ashley Blewer, Natalie Cadranel

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

MediaArea continues to help grow a community via the IETF (Internet Engineering Task Force) CELLAR (Codec Encoding for LossLess Archiving and Realtime transmission) working group, where work is being done to support the standardization of Matroska and FFV1. The working group's mailing list has been very active with members of the core Matroska community and the FFV1 community working on further developing the latest specification for standardization along with the MediaArea team. The listserv of the CELLAR working group now includes 89 members, include 36 active participants and over 750 emails.

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

MediaConch has continued to establish presence in a variety of communities that have interest in the developing software by presenting or being represented at many different conferences.

Progress Highlights Since the October, 2015 Release

Since the October, 2015 release, the project has made significant progress towards prior objectives and more recent goals arising from our user experience research. These include: enhancements based on interviews, community feedback, conference presentations, hosting a symposium, community outreach and collaboration with preservationists and developers within the IETF CELLAR working group.

- discuss original specs w/ anticipated intermediate specs w/ actual specs
- improved build environment with detailed instructions on how to download builds for Linux, Windows, and Mac OSX
- Updated website design so users can find installation links, directions, and daily builds more easily
- GH Issue Tracker links directly from Help/FAQ page in the GUI
- A Roadmap and Timeline were created, here: https://mediaarea.net/MediaConch/documentation/Roadmap.html
- We have responded to feedback from user interviews and have opened tickets for each unresolved issue mentioned in the initial feedback along with others in GitHub- https://github.com/MediaArea/MediaConch/issues

2. Testing

The MediaArea team helped moved the official Matroska collection of test files from a static zip file hosted on sourceforge into a new GitHub repository at https://github.com/Matroska-Org/matroska-test-files. In addition MediaArea is re-processing a collection of a hundred thousand original Matroska files at archive.org under new more verbose parsing options of MediaInfoLib in order to generate more comprehensive test results in the application of the developing implementation checker.

The original work to analyze the massive corpus of the Internet Archive is being redone, since the original analysis parsed files too selectively to cover many types of Matroska implementation checks. The new analysis is slower but more comprehensive and shall allow MediaArea and the CELLAR working group to identify implementation errors and relate them to specific samples, specification language, and the creation software.

The development work in CELLAR has also clarified many obligations, constraints, and expectation for Matroska validity so the need to comprehensive test sets has greatly expanded.

Testing with the policy checker has expanded with collaborations with Artefactual and the Tate Museum. Through testing and collaboration with these organizations and other users, we've determined that the policy checker had to be significantly expanded to cover and/or inter-rule logic as well as allowing the expression of policy for values between two provided files. For instance Archivematica could have a policy that states that an original file and created derivative should share specific significant characteristics or the Tate may use complex and conditional policies to test if media adheres to the requirements of various display hardware.

In the No Time to Wait symposium, the MediaArea team curated a set of presentations that focused on testing FFV1, included Peter Bubestinger's work to test performance and resillience of FFV1 under its massive combinations of options. Additionally Kieran Kuhnya presented on his work to create a fuzz about 22 million FFV1 files in order to isolate decoding errors and file tickets and research that has helped refine the FFV1 decoder and clarify some requirements with the specification language.

3. Dissemination and community building

Bi-monthly Newsletters

https://mediaarea.net/MediaConch/newsletters.html

No Time To Wait! Symposium

MediaArea organized a 3-day symposium held during IETF96 with over 75 people RSVPed and over 60 attendees.

Website: https://mediaarea.net/MediaConch/notimetowait.html

Documents Repository: https://github.com/preforma/notimetowait

Twitter Storify: https://storify.com/ablwr/no-time-to-wait

Report (by Ashley Blewer):
https://mediaarea.net/MediaConch/2016/07/26/No-Time-To-Wait-Preservation-FFV1-Matroska-Symposium/
Report (by Sound&Vision):
http://www.beeldengeluid.nl/en/blogs/research-amp-development-en/201607/tools-trade

IETF

Project leads Jerome Martinez and Dave Rice, along with Matroska developer Steve Lhomme, presented at IETF's CELLAR Working Group on the status and work performed on the FFV1, EBML, and Matroska specifications.

iPRES

MediaArea will give a talk on the standardization process for FFV1 and Matroska. MediaArea will also participate in a PREFORMA workshop, demonstrating the MediaConch tool and how it can be applied to digital preservation workflows.

Programme: http://www.ipres2016.ch/frontend/index.php?folder_id=353

Artefactual

Archivematica roadmap: https://wiki.archivematica.org/Development_roadmap:_Archivematica

Archivematica integration documentation:
https://wiki.archivematica.org/Requirements/MediaConch_integration

Tate Britain via PERCELES Project

Ashley Blewer and Dave Rice gave a workshop at Tate focused on MediaInfo, including demonstrations and workshop projects focused around MediaConch and creation/application of local policies in addition to implementation and conformance.

Mailing list note:
http://us10.campaign-archive1.com/?u=58df5a080c65b0a214c4f4e96&id=86739767e2.

Digitalmeetsculture writeup:
http://www.digitalmeetsculture.net/article/preforma-mediaconch-workshop-at-tate/

4. Open Source approach

By rewriting specifications into Markdown and posting on Github, more people are able to contribute to the specifications, including archivists without development or prior Github experience. Work on this integration has grown since the last report and the community is thriving more than before, particularly after outreach efforts to the respective development listservs and

MediaArea is able to work closely with the FFV1 and Matroska format designers to ensure standardization in line with the formats vision. MediaArea is also working closely with other preservationists, collecting user needs and feedback, to apply to the forthcoming standardized specifications. Supplementally, MediaArea has interviewed users of MediaConch to gather feedback and implement improved features within the software through filing Issues on the code's Github page.

The standardization work is being done through IETF, specifically the CELLAR working group. Within this group, with a mailing list of 87 members and growing, the standardization conversation can happen between designers, developers, and archivists.

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

5. Standardisation efforts

MediaArea has participated and collaboration with several communities to make progress at standardization efforts for FFV1, MKV, and EBML (the foundational format of MKV). The team participation in the first in person meeting of the CELLAR working group at IETF96 hosted in July 2016 in Berlin. The meeting included presentations by Steve Lhomme and Jerome Martinez on the state of Matroksa and FFV1's specification work. The meeting was productive and received helpful feedback from many members of the IETF community.

Surrounding the IETF meeting, MediaArea co-hosted a symposium in Berlin, called No Time to Wait, that focused on the standardization and use of FFV1 and Matroska in archives. The symposium was a great success and extremely productive at speeding along the standardization efforts of FFV1 and Matroska. Since the meeting, the MediaArea team has posted videos from the symposium's presentations and have been engaging with new participants in the CELLAR working group.

The No Time to Wait symposium provided good opportunities to learn from and collaborate with related standardization efforts such as Google's work to standardize a Matroska branch, webm, and the Library of Congress's work in defining AS-07.

6. Impact assessment, sustainability, future use and exploitation

Extending into communities via Archivematica

MediaArea's partnership with Artefactual will include MediaConch being integrated into Archivematica workflows. Archivematica is a popular framework for OAIS-compliant digital preservation with a robust user community.

Extending into other formats

Functionally, MediaConch can already be expanded to support file formats beyond Matroska, FFV1, and LPCM (and PDF/TIFF via integration of work from the other suppliers). There is potential for MediaConch to become the conformance checking software for *any* audiovisual formats after the completion of this project, not just limited to Matroska and FFV1. MediaConch's local policy creation feature can already be extended out to any format that MediaInfo supports.

Highlights from symposium feedback:

Attendees of the No Time To Wait! Symposium from July 18th - 20th felt it was a unique oportunity for archivists and open source software developers to collaborate and share feedback. Archivists appreciated the opportunity to engage more closely with Matroska and FFV1 and welcomed the opportunity to interact with developers about their needs and workflow specifications. Direct access to the specification authors helped the international archival community better understand the tools and have meaningful two-way interactions.

Feedback:

- "Group discussions, hearing where other archives were, hearing about all the different aspects of standardization, seeing the faces behind the mailing list discussions"

- "Detailed presentations on software and formats directly from their developers. Great!"

- "Getting in touch with users (archivists) and developers. Hearing about other use-cases of Matroska/FFV1."

- "I found the data exchange between archives and developers very helpful. The symposium was full of energy and I am very motivated now."

- "Symposium was very useful and I have learned many new things, met many amazing people. I really appreciate that every speaker was talking straight to the point. Overall I am happy that I attended this symposium because we are at the beginning of true digital archivation in the National Film Archive and all what I have learned there will help us to choose the right path."

- "We were really happy to be invited. We are a small institution with a big mission, and we rely heavily on market solutions and vendors, as such to us it is very important to be part of a community that shares the same goals and principles on archiving. I was very much intimidated by the level of smarts at the symposium, but that is a good thing :-) We would also like to remain involved somehow. We are currently communicating about the symposium and the

PREFORMA project on a national level to other archives, and will adopt a pro-active approach in assessing whether we can implement FFv1 and Matroska."

- "I think it was a very positive event to hold and very focused; bringing together a mix of those who had used FFV1 and MKV and were interested in comparing user stories, and others seriously considering the adoption of this combination. Such events help to raise the profile of the FFV1 codec and confidence in the FFV1/ MKV combination; stimulating interest from archives who may not have attended but who do follow archival developments."

7. Gap analysis and next steps

Verbosity settings are present in MediaConch but should be expanded to be more complex. Current verbosity is "high" and "low" and lack nuanced versions in between. There are plans to extend verbosity to 5 levels.

Integration of other conformance checkers has happened. This feature exists and is functional.

In April, MediaArea had the goal of merging the MediaConch GUI and the Web UI into one common UI. For a time, the UI were separate and inluded different features. The best of both, based on user feedback, were integrated into a common UI for consistency and for ease of updating.

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

According to work plan and timeline, the fixer component should already have been integrated into the software. We are behind schedule regarding the metadata fixer, opting instead to prioritize the "performance optimization" of the application, increase implementation checker work, and an overhaul of the policy checking elements of MediaConch.