

PREservation FORMAts for culture information/e-archives

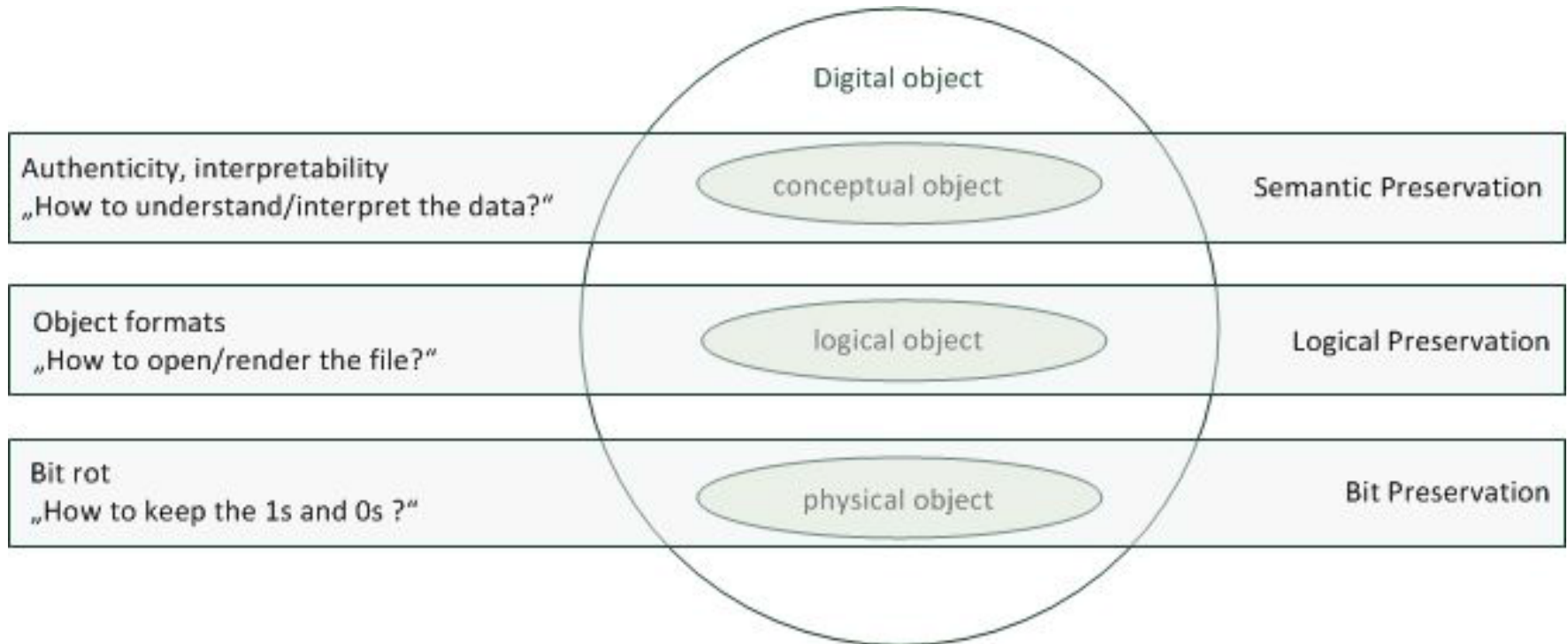
*Künftige Daten- und
Archivierungsstandards für die (Be)-
Wahrung des Kulturerbe*

Peter Pharow, Stefan Rohde-Enslin

PREFORMA - Präsentation

NESTOR – Tag

14. Juni 2016



Nach dem Konzept von
Ken Thibodeau

- Anforderungen an Werkzeuge (Tools)
 - Möglichst vollständige Validierung standardisierter Dateiformate im Bereich der Langzeit-Archivierung
 - Offenes Konzept, ohne Kontrolle bzw. Einflussnahme durch andere Institutionen
- Verschiedene Bereiche, Schritte, Phasen
 - Transfer
 - Import
 - Migration
 - Scanning
 - Kreation

- ❑ PREFORMA ist ein **Pre-Commercial Procurement** Projekt, das von der Europäischen Kommission im Rahmen des FP7-ICT Programms gefördert wird.
- ❑ **Start:** 1. Januar 2014
- ❑ **Laufzeit:** 48 Monate (Ende 31. Dezember 2017)
- ❑ **Website:** www.preforma-project.eu
- ❑ **Kontaktpersonen**
 - Projektkoordinator: Borje Justrell, Riksarkivet, borje.justrell@riksarkivet.se
 - Technische Koordinatorin: Antonella Fresa, Promoter Srl, fresa@promoter.it

- ❑ RIKSARKIVET, Schweden **Projektkoordinator und Archiv**
- ❑ PROMOTER SRL , Italien **Technischer Koordinator und Kommunikation**

- ❑ **Technische Partner**
 - PACKED EXPERTISECENTRUM DIGITAAL ERFGOED VZW, Belgien
 - **Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V., Deutschland (→ IDMT)**
 - HOGSKOLAN I SKOVDE (Universität Skovde), Schweden
 - UNIVERSITA DEGLI STUDI DI PADOVA, Italien

- ❑ **Kulturerbe-Einrichtungen**
 - STICHTING NEDERLANDS INSTITUUT VOOR BEELD EN GELUID, Niederlande
 - Koninklijk Instituut voor het Kunstpatrimonium, Belgien
 - GREEK FILM CENTRE AE, Griechenland
 - LOCAL GOVERNMENT MANAGEMENT AGENCY-AN GHNIOMHAIREACHT BAINISTIOCHTA RIALTAIS AITIUIL, Irland
 - **STIFTUNG PREUSSISCHER KULTURBESITZ, Deutschland**
 - AYUNTAMIENTO DE GIRONA, Spanien
 - Eesti Vabariigi Kultuuriministeerium, Estland
 - KUNGLIGA BIBLIOTEKET, Schweden

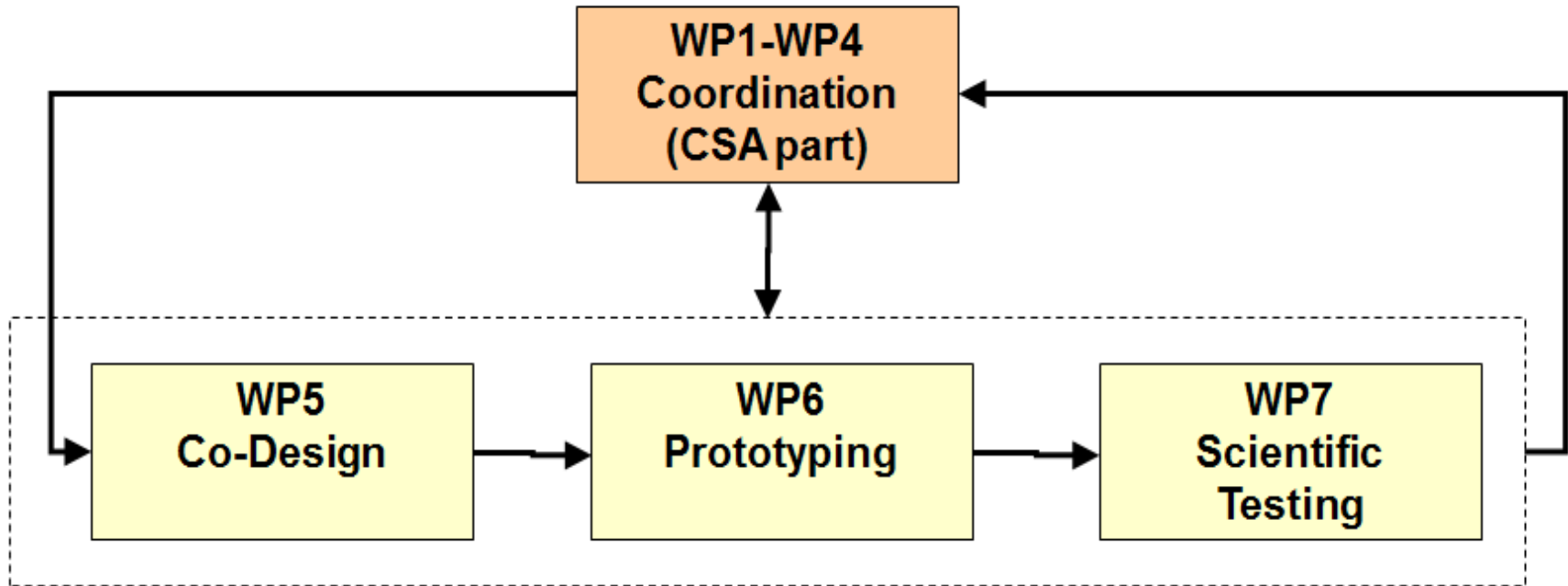
- ❑ Pre-Commercial Procurement (PCP), bzw. vorkommerzielle Auftragsvergabe, ist ein dem Wettbewerb ähnliches Beschaffungs-verfahren
- ❑ Die öffentliche Hand darf gemeinsam mit innovativen Unternehmen oder Partnern Entwicklungsprojekte starten
- ❑ Zur Erarbeitung innovativer Lösungen für spezielle Herausforderungen und Bedürfnisse des öffentlichen Sektors
- ❑ Um das Risiko zu minimieren, werden diese innovativen Lösungen mittels eines in Phasen ablaufenden Prozesses zur Beschaffung von Entwicklungsverträgen erarbeitet.
- ❑ PCP wird innerhalb der Europäischen Union zunehmend eingesetzt.
- ❑ **National? Regional? Wer ist denn schon dabei? Wer kennt diese Möglichkeiten? Wer kennt den national AK PCP/PPI?**

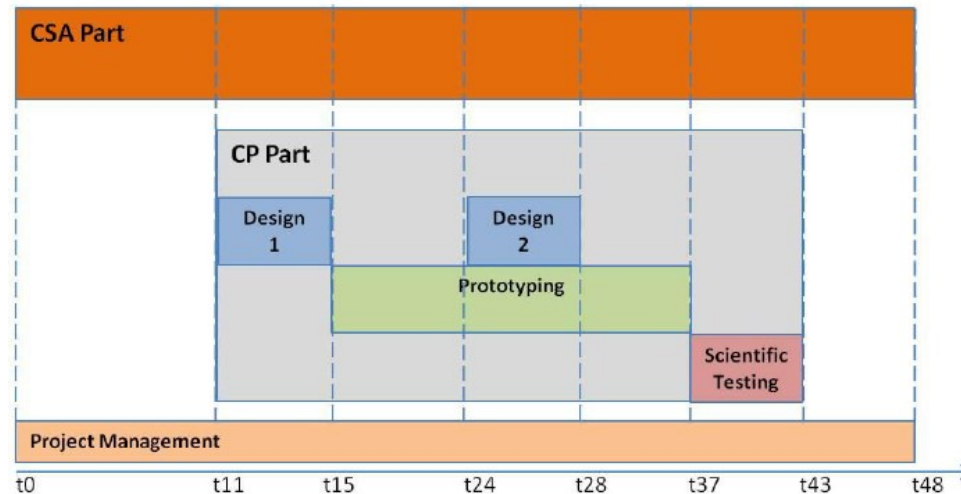
- ❑ Kulturerbe-Einrichtungen, Bibliotheken, Archive und Museen stehen vor der Herausforderung, **immer mehr elektronische Inhalte** langfristig speichern und aufbewahren zu müssen.
- ❑ Sicherheit und Authentizität der Daten bei gleichzeitiger Offenheit (Open Source, Open Standard) müssen garantiert werden.
- ❑ Elektronische Inhalte werden dafür in speziellen **Dateiformaten** (für Dokumente, Bilder, Audio, Video usw.) gespeichert, die üblicherweise mittels Software verschiedener Anbieter erstellt wurden.
- ❑ **Erhöhung der Attraktivität für Kunden (wer sind diese Kunden denn heute und morgen?)**

- ❑ Auch wenn diese Dateien in Standardformaten vorliegen, kann die **Implementierung von Standards** nicht garantiert werden:
 - Die Software, die für die Implementierung von Standards zur Erstellung der elektronischen Dateien verwendet wird, untersteht weder der Kontrolle der Institutionen, die die Dateien erstellen, noch der Kontrolle durch die Kulturerbe-Einrichtungen.
 - Die Kulturerbe-Einrichtungen führen zwar Konformitätstests zum elektronischen Transfer durch, diese sind aber nicht absolut zuverlässig. Je nachdem, welche Software für den Test eingesetzt wird, kann es zu unterschiedlichen Ergebnissen kommen.

- ❑ Dies führt zu Problemen bei **langfristiger Speicherung und Aufbewahrung**.

- ❑ **Datenobjekte, die einen unkontrollierten Erstellungs- und Bearbeitungsprozess durchlaufen, können das gesamte Archivierungsvorhaben gefährden.**





- ❑ **Gesamtbudget der Beschaffung:** 2.805.000 EUR
- ❑ **Ausschreibungen:** beendet, sechs Konsortien ausgewählt
- ❑ **Design Phase** (4 Monate): November 2014 – Februar 2015 (DRI Leitung)
- ❑ **Prototype Phase** (22 Monate): März 2015 – Dezember 2016
 - Erster Prototyp: März 2015 – Oktober 2015
 - Re-design: November 2015 – Februar 2016
 - Zweiter Prototyp: März 2016 – Dezember 2016
- ❑ **Testphase** (6 Monate): Januar 2017 – Juni 2017

- ❑ **Herausforderung:** Implementierung verschiedenster qualitativ hochwertiger, standardisierter Dateiformate für die langfristige Speicherung und Aufbewahrung von elektronischen Inhalten.
- ❑ **Hauptziel:** Kulturerbe-Einrichtungen bekommen die volle Kontrolle über den Prozess der Durchführung von Konformitätstests bei der Aufnahme von Dateien in Datenarchive.
- ❑ **Hauptziel des PCP im Rahmen von PREFORMA:** Entwicklung und Bereitstellung einer Open-Source-lizenzierten Referenzimplementierung für verschiedene Dateiformat-Standards, die von jeder Kulturerbe-Einrichtung (oder anderen Organisation) zur Durchführung von Konformitätstests bezüglich spezieller Standards genutzt werden kann.

- ❑ **Koordinations- und Unterstützungsprojekt (CSA):** Die Teilnehmer definieren die Anforderungen, erstellen die Ausschreibung, entwickeln Test-Datensätze, analysieren die eingegangenen Angebote, wählen die Angebote aus und begleiten den gesamten Prozess vom Design über die Entwicklung bis hin zum Einsatz bei den Partnern (den Kulturerbe-Einrichtungen).
- ❑ **Entwicklungsprojekt (CP):** Die Gewinner der Ausschreibung (je zwei Konsortien pro Medientyp) entwickeln die Modelle der Software und stellen sich der Begutachtung. Ein Konsortium pro Medientyp wird beim ersten Review ausgewählt und entwickelt die eigentliche Software bis zur Einsatzreife.
- ❑ **Unterstützer:** Alle weiteren Interessenten (Bibliotheken, Museen, Firmen) nutzen die Test-Datensätze und die Spezifikationen zur Erstellung eigener Module außerhalb des Projektes.

- ❑ Entwicklung eines **Open Source Conformance Checkers**, welcher:
 - prüft, ob eine Datei den Standardspezifikationen entspricht,
 - prüft, ob eine Datei den Akzeptanzkriterien der jeweiligen Archiveinrichtung entspricht (auch inhaltlich, politisch, strafrechtlich),
 - Prüfberichte erstellt, die sowohl von Menschen als auch von Softwareanwendungen interpretiert werden können,
 - einfache Reparaturmaßnahmen durchführt.

- ❑ Etablierung eines Ökosystems, in das die **Open-Source-Referenzimplementierung** eingebettet ist, welches:
 - nützliches Feedback für diejenigen liefert, die die Software kontrollieren,
 - die Verbesserung der Standardspezifikation vorantreibt,
 - die Entwicklung neuer Business Cases für das Management der langfristigen Speicherung und Aufbewahrung von Dateien vorantreibt.

- ❑ PREFORMA basiert auf einem **Open-Source-Ansatz**, der das Ziel verfolgt, eine nachhaltige Forschungs- und Entwicklungs-Community zu etablieren, die ein breites Spektrum an Mitwirkenden und Nutzern aus unterschiedlichen Stakeholder-Gruppen umfasst.
- ❑ Der Open-Source-Ansatz stellt die nachhaltige Verfügbarkeit der Software sicher – unabhängig von den an PREFORMA beteiligten Kulturerbe-Einrichtungen und Softwareentwicklern/-anbietern.
- ❑ Lizenzen
 - Jegliche **Software**, die im Rahmen von PREFORMA entwickelt wird, wird unter den beiden Open-Source-Lizenzen „GPLv3 oder höher“ oder „MPLv2 oder höher“ bereitgestellt.
 - Alle **digitalen Assets**, die im Rahmen von PREFORMA entwickelt werden, werden unter der Open-Access-Lizenz „Creative Commons CC-BY v4.0“ und in offenen Dateiformaten (SOU 2009:86) bereitgestellt.

- ❑ **Kulturerbe-Einrichtungen** und Organisationen die diese Archive betreuen oder koordinieren - die mit digitalem Kulturerbe arbeiten oder dies planen (auch für Events wie Hack-A-Thons).
- ❑ **Entwickler**, die Code für die PREFORMA-Tools bereitstellen, sowie Entwickler, die Referenzimplementierungen in Produktionssoftware umsetzen (auch außerhalb des Projektes).
- ❑ **Forschungseinrichtungen**, die technisches Know-how und Expertenwissen für die genannten kulturellen Stakeholder bereitstellen.

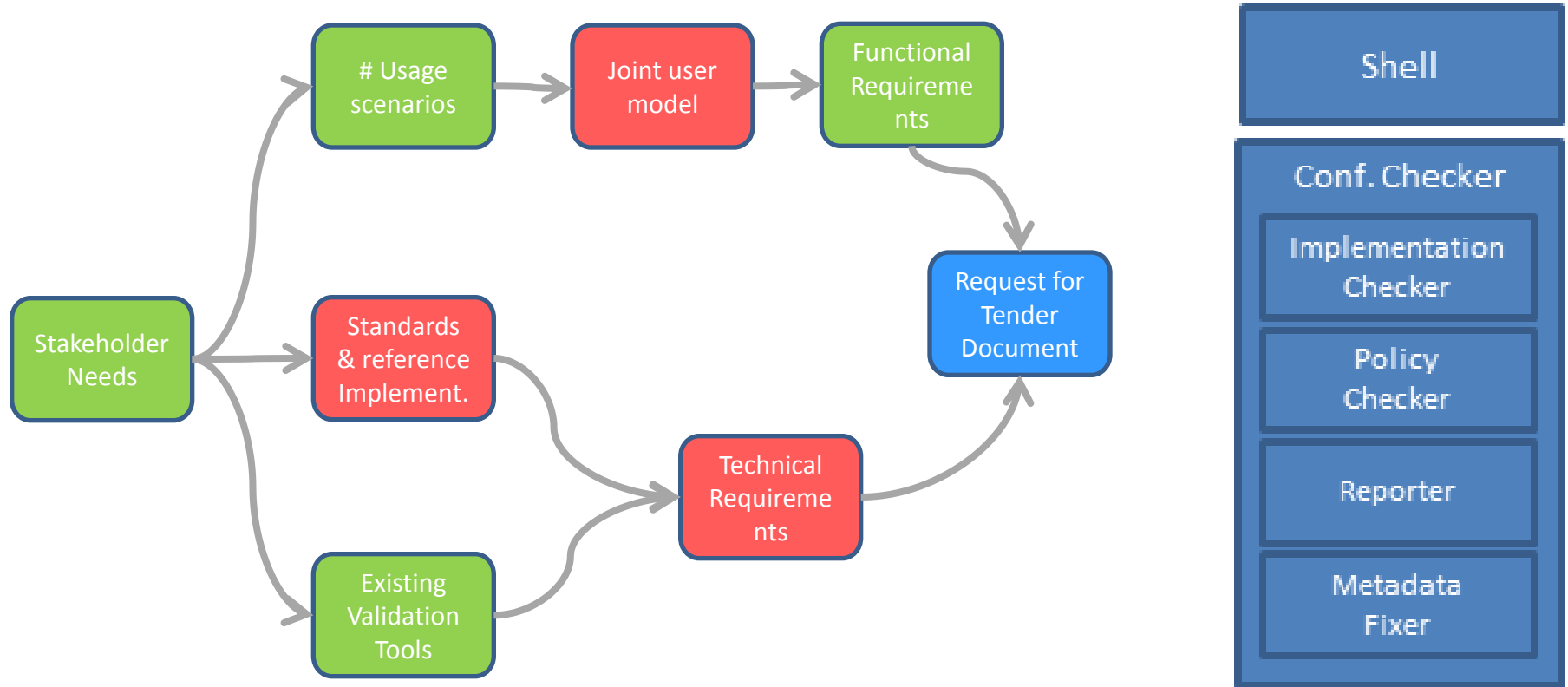
- ❑ **Standardisierungsorganisationen**, die die technischen Spezifikationen der von PREFORMA abgedeckten Speicherformate (Text, Image, AV) definieren.
- ❑ **Förderorganisationen**, wie z. B. Kultusministerien oder nationale/regionale Verwaltungen, die Digitalisierungsprogramme durchführen und verwalten, und die die Nutzung der PREFORMA-Tools im Prozess der Digitalisierung unterstützen können.
- ❑ Weitere **Projekte** (EU, national, regional) in den Bereichen „Digital Culture“, „eInfrastructure“ und in vielen Bereichen der Politik.

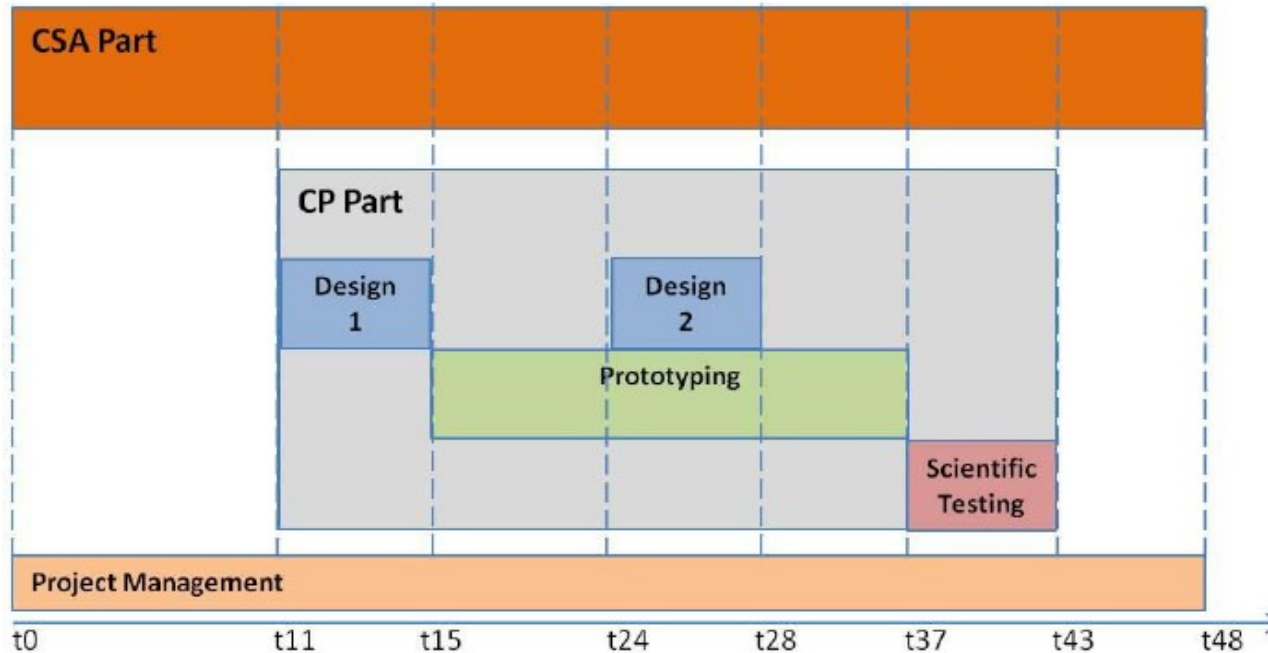
- ❑ Folgende Arten von Medien wurden von der Ausschreibung abgedeckt: Textdateien, Bilddateien und Audio/Video-Dateien (Container).
- ❑ Bereiche, die von den im Rahmen von PREFORMA entwickelten Werkzeugen abgedeckt werden, sind:
 - speicherrelevante Metadaten (für Textdateien, Bücher, Bilder und Audio/Video-Aufzeichnungen),
 - besonders anspruchsvolle bzw. schwierige Arten von digitalen Objekten (insbesondere Audio/Video-Aufzeichnungen),
 - Integration digitaler Speicheranforderungen in existierende Systeme, um deren Widerstandsfähigkeit gegenüber technologischen Veränderungen zu erhöhen.

□ Erwartet wurden folgende Ergebnisse:

1. ein Set von Modulen (im Einzelnen: „Abstract Conformance Checker“, „Scorer“, „Reporter“, und „Abstract Fixes Suggester/Correcter“) für jeden Medientyp,
2. die technische Referenz dieser Module sowie deren APIs zur vollständigen Dokumentation für die zukünftige Integration der Module in die Workflows der Gedächtnisinstitutionen,
3. eine Web-Applikation, die das User Interface zu den vier Modulen darstellt.

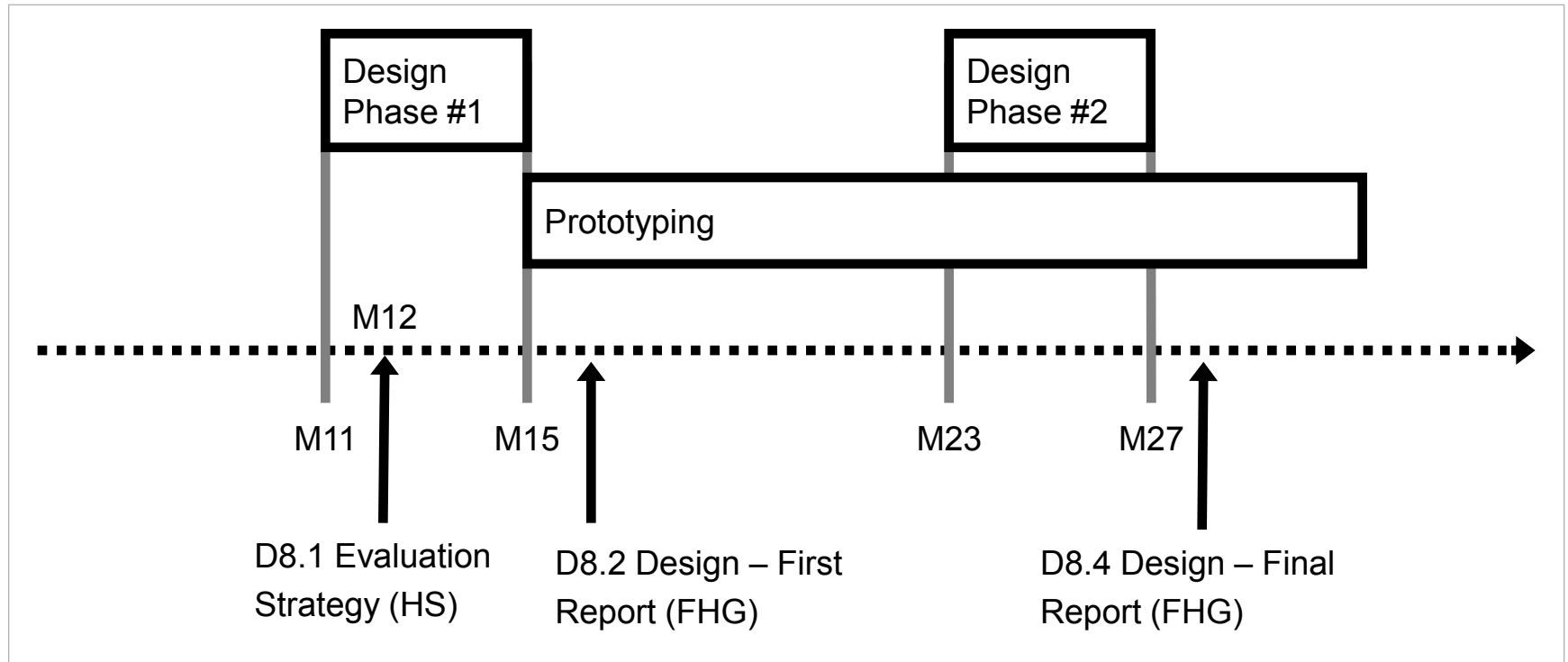
Conformance Checker





- ❑ **Designphase** (4 Monate): November 2014 – Februar 2015
(Leitung durch das Fraunhofer IDMT Ilmenau)
- ❑ **Prototyp-Phase** (22 Monate): März 2015 – Dezember 2016
 - Erster Prototyp: März 2015 – Oktober 2015
 - Re-design: November 2015 – Februar 2016
 - Zweiter Prototyp: März 2016 – Dezember 2016

Überblick Designphasen

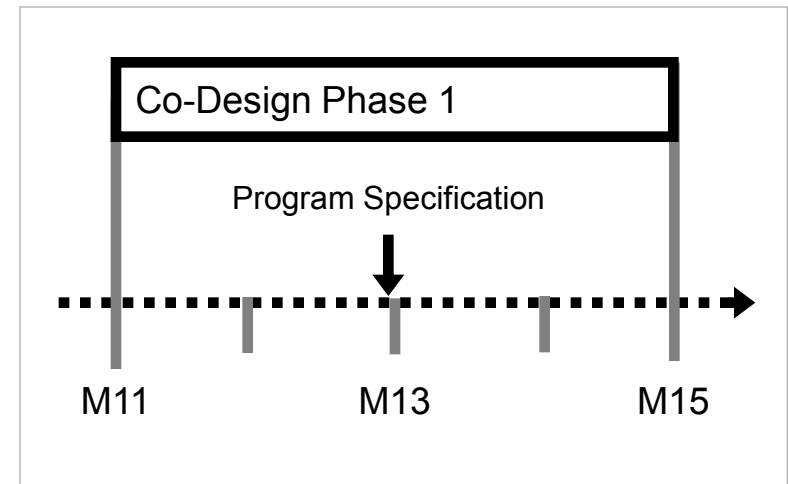


■ Anforderungen an die Bieter:

1. M11 – M12: Funktionale Spezifikation (Analyse der funktionalen und technischen Anforderungen aus dem Tender heraus)

■ Ergebnis: Programmspezifikat.

- Beschreibt die Intentionen
- Über Use Cases, Szenarien
- Hauptteil der Dokumentation



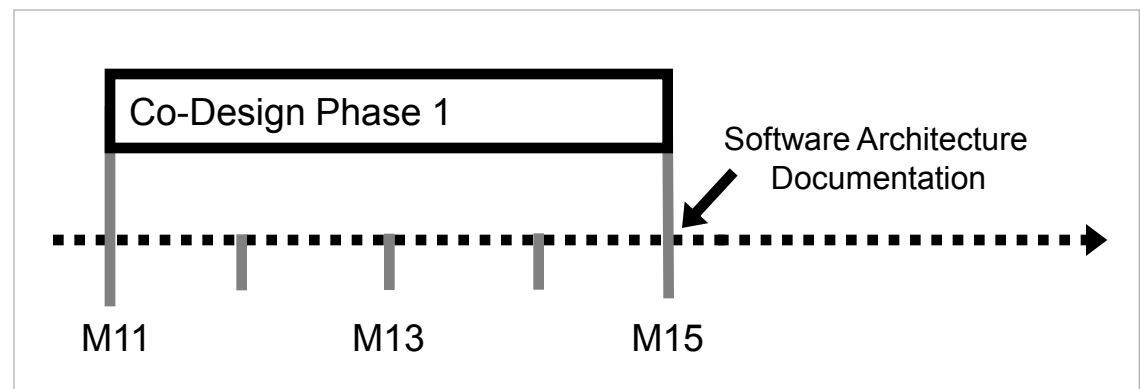
■ Anforderungen an die Bieter:

2. M13 – M14: High-Level-Struktur der Komponenten (Software)

- Ergebnis: Softwarearchitektur Dokumentation
- Beschreibt die Struktur der einzelnen Komponenten und deren Verbindung (Nahtstellen) auch zu den anderen Datenformaten

■ Interoperabilität

- Technisch
- Semantisch
- Policy



Designphase 2



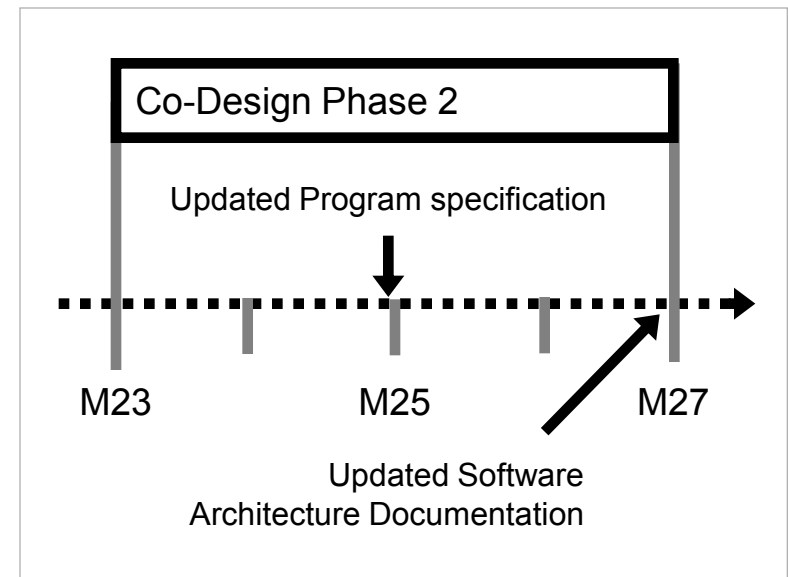
■ Anforderungen die Bieter:

1. M23 – M26: Update und Upgrade des gewählten Designs

→ Unterstützung der Bieter, Review, Dokumentation, Tests, Daten, Nutzer, Anwender, praktische Relevanz

- Ergebnis: Programm-Spezifikation (Update)
- Ergebnis: Update Software-Architektur-Dokumentation

■ Präsentation der Ergebnisse, Workshop



Designphase 2

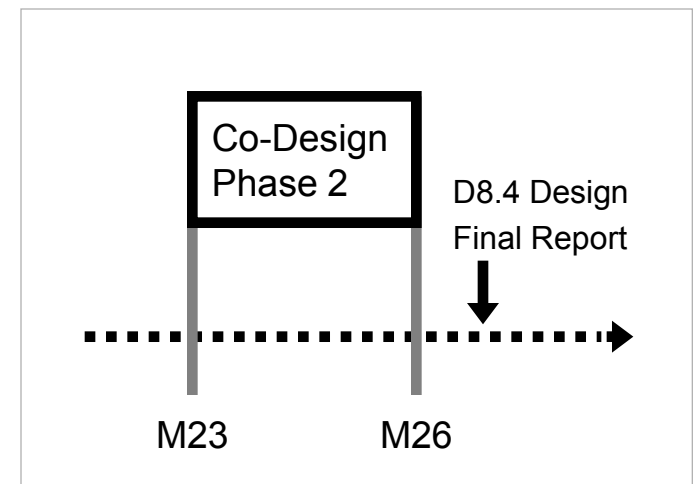


■ Anforderungen an PREFORMA

1. M27: D8.4 Design – Final Report

- Dokumentation der Änderungen durch die Bieter
- Response der Bieter auf die Kommentare und Gutachten
- Funktionalität und nicht-funktionale Aspekte und Anforderungen

→ Software-Qualität nach ISO/IEC 9126



| Content type | Standard specifications |
|--------------|--|
| TEXT | PDF 1.7 (ISO 32000-1:2008) PDF/A-1 (ISO 19005-1:2005) PDF/A-2 (ISO 19005-2:2011) PDF/A-3 (ISO 19005-3:2012) |
| IMAGE | TIFF/EP (ISO 12234-2:2001) TIFF/IT (ISO 12369:2004) |
| AUDIOVISUAL | MKV (http://www.matroska.org/technical/index.html) Ogg (https://xiph.org/ogg/doc/) Lossless JPEG2000 (ISO/IEC 15444-1:2004) Lossless FFV1 http://www.ffmpeg.org/~michael/ffv1.html Lossless Dirac v2.2.3 (http://diracvideo.org/download/specification/dirac-spec-latest.pdf) Linear PCM (IEC 60958-1 ed3.1 Consol. with am1: 2014) |

- Open Planets - VeraPDFa (PDF/A)**
- Preservica - Digital Preservation Validation Framework (PDF/A)

- EasyInnova - DPF Manager (TIFF)**
- LIBIS - LIBIS/Aware checker for TIFF (TIFF)

- UCL - OpenMediaCheck (MKV|JPEG2000|LPCM)
- MediaInfo - PreForma MediaInfo (MKV|FFV1|LPCM)**

- Passende Evaluationskriterien zu finden ist der kritischste Teil des Projektes, da diese Kriterien von Fall zu Fall auch quantitative Werte generieren müssen.
- Die Angebote der Bieter trafen den Kern der Ausschreibungen sehr gut, Nacharbeiten waren meist einfach umzusetzen und haben das Projektziel nicht gefährdet.
- Die Bewertung qualitativer Kriterien wie die „Eleganz“ der Architektur kann nur durch die Kombination qualitativer und quantitativer Kriterien abgebildet werden.
- PREFORMA legt Wert darauf, dass die Bieter wissenschaftliche Arbeit leisten und Ergebnisse zur Verfügung stellen.

Formal vs. Informell



- Ansatz der letzten Begutachtung (informelle Evaluation):
 - Wesentliche Kommentare existierender Quellen
 - Keine „Eliminierung“ von Bietern, dafür stärkere Bewertung
- Ergebnisse:
 - Drei hoch motivierte Bieter, die wissen, was sie wie tun
 - Nur geringe Abweichungen vom Originalkonzept
 - Orientierung auf Standards! Orientierung auf Interoperabilität!
- Anpassungsbedarf:
 - Weitere Gutachter und Experten, die schnell auf Anfragen reagieren können und den Bietern damit helfen
 - Letzter formaler Review: Ende der Prototypentwicklung



Besuche bei den Bietern



- Begründung:
 - Es ist nicht nur das Angebot, das zählt
 - Es ist auch das Team hinter dem Angebot
 - Es ist die Herangehensweise an die Arbeit, die zählt
- Ergebnisse:
 - Alle drei Teams haben überzeugt
 - Jedes Team ist anders (Hacker, Geschäftsmann, Standardisierer)
 - Jedes Team arbeitet anders (lokale Labs bis voll virtuell)
- Wo sich die Bieter gegenseitig unterstützen:
 - Jeder Ansatz hat Vor- und Nachteile (Typ, Modell, Organisation)
 - Sowohl Wissenschaft als auch Interoperabilität und Produkt



Ende der Designphase



- Was wir haben:
 - Jede Menge Zwischenberichte und Software-Versionen zum Test
 - Spezifikationen mit Updates und Upgrades halbjährlich
 - Jedes Ende einer Phase wird durch einen Bericht abgeschlossen
- Wo es immer hapert:
 - Interoperabilität zwischen Conformance Checkern der Bieter
 - Einbeziehung der Ergebnisse des Entwickler-Workshops
- Was wir noch vorhaben:
 - **Sie sind alle herzlich eingeladen am Test teilzunehmen.**
 - **Schicken Sie uns Ihre Testfiles**
 - **Oder laden Sie die entwickelte Software herunter und testen**
 - **Schicken uns einen kurzen Bericht → jedes Feedback hilft!**



Bisherige öffentliche Veranstaltungen



- 1. Informationsveranstaltungen für Interessenten**, in Brüssel (April 2014), Helsinki (Juni 2014) und Mailand (Juli 2014).
- 2. Präsentation des Projektes**, u.a. in Helsinki (Juni 2014), Brüssel (Juli 2014), Girona (Oktober 2014), Limassol (November 2014), usw.
- 3. Kooperationsevents** in Brüssel (Februar 2014, EU; März 2016, EU) und Stockholm (Juli 2014, Schweden) zum Austausch von Erfahrungen mit anderen PCP-Projekten bzw. mit nationalen Projekten, welche mögliche Nutzer der Technologie sind.



Künftige öffentliche Veranstaltungen



1. Eine **Trainings- und Informationsveranstaltung für Open-Source-Unternehmen**, die vom 5. bis 8. April 2016 in Verbindung mit der Demonstration des ersten Prototypen in Stockholm stattfand
2. Ein **Workshop zum Erfahrungsaustausch**, der vom 21. bis 24. November 2016 in Verbindung mit der Demonstration des zweiten Prototypen in Berlin stattfinden wird. Die PREFORMA Partner werden anderen Kulturerbe-Einrichtungen von ihren Erfahrungen und der Zusammenarbeit mit Zulieferern im Rahmen von R&D Dienstleistungsverträgen berichten
3. Eine **Abschlusskonferenz** im Dezember 2017 in Stockholm, auf der die Projektergebnisse präsentiert werden



Besuchen Sie uns im Web!



PREFORMA Website

www.preforma-project.eu

A screenshot of the PREFORMA website homepage. The page has a dark header with navigation links: WEBSITE, PROJECT, PARTNERS, TENDER, ACTIVITIES, OPEN SOURCE PORTAL, COMMUNITY, DOWNLOAD, CONTACTS. Below the header is a banner with the European Union flag, the PREFORMA logo, and the SEVENTH FRAMEWORK PROGRAMME logo. The main content area is divided into several sections: a left sidebar with user information (Logged in as: preforma, Logout), MEDIA PARTNER (DIGITAL CULTURE), CONTACTS, and LATEST NEWS; a central main content area with a "PRESENTATION OF THE PROJECT" section featuring a diagram of document flow and a "PREFORMA, FUTURE MEMORY STANDARDS" section; and a right sidebar titled "PARTNERS" listing various institutions like Riksarkivet, PAKED, PROMOTER, Fraunhofer IDMT, and the University of Skovde.

A screenshot of the PREFORMA website showing an upcoming event. The header includes the PREFORMA logo, navigation links, and logos for the European Union and the Seventh Framework Programme. The main content area features a large graphic with a document icon, a box with an arrow, and a smiling document icon. Below this is a section titled "UPCOMING EVENTS" with a "VIEW ALL" link. The event is "PREFORMA CALL FOR TENDER, INFORMATION DAY" on "BRUSSELS, 4 APRIL 2014". A "READ MORE" link is visible at the bottom of the event description.

PREFORMA Blog

www.digitalmeetsculture.net/projects/preforma/

