KOST Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen

Ein Gemeinschaftsunternehmen von Schweizer Archiven

TIFF
Preservation-Prozess in der Praxis

19. Tagung des Arbeitskreises AUdS,
Wien,
11.03.2015

Disposition

- Einleitung
- Der Preservation-Prozess anhand von TIFF
- Erkenntnisse

Die KOST

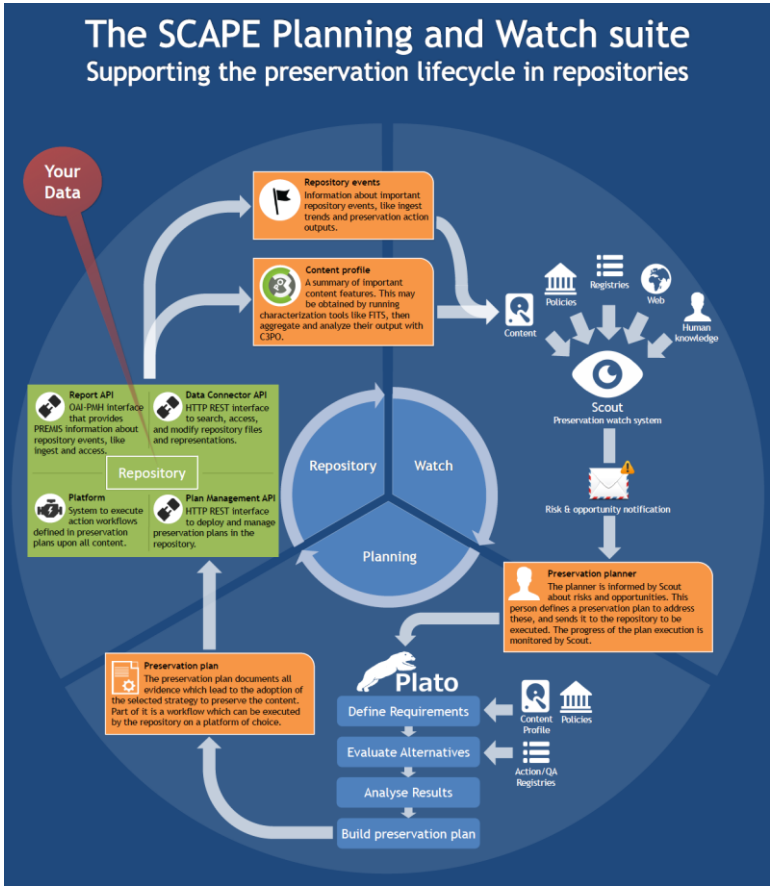
Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen

- Gemeinschaftsunternehmen des Schweizerischen Bundesarchivs, des Landesarchivs des Fürstentums Liechtensteins, 24 kantonalen und 5 kommunalen Archiven der Schweiz.
- Sie hat den Auftrag, ihre Träger bei der Archivierung elektronischer Unterlagen zu unterstützen
  - Standards und Richtlinien
  - Tools
  - Dienstleistungen zur Lösung konkreter Probleme
  - Studien und Veranstaltungen

Preservation-Prozess

In den letzten Jahren wurden diverse Prozesse für die Digital Preservation entwickelt.

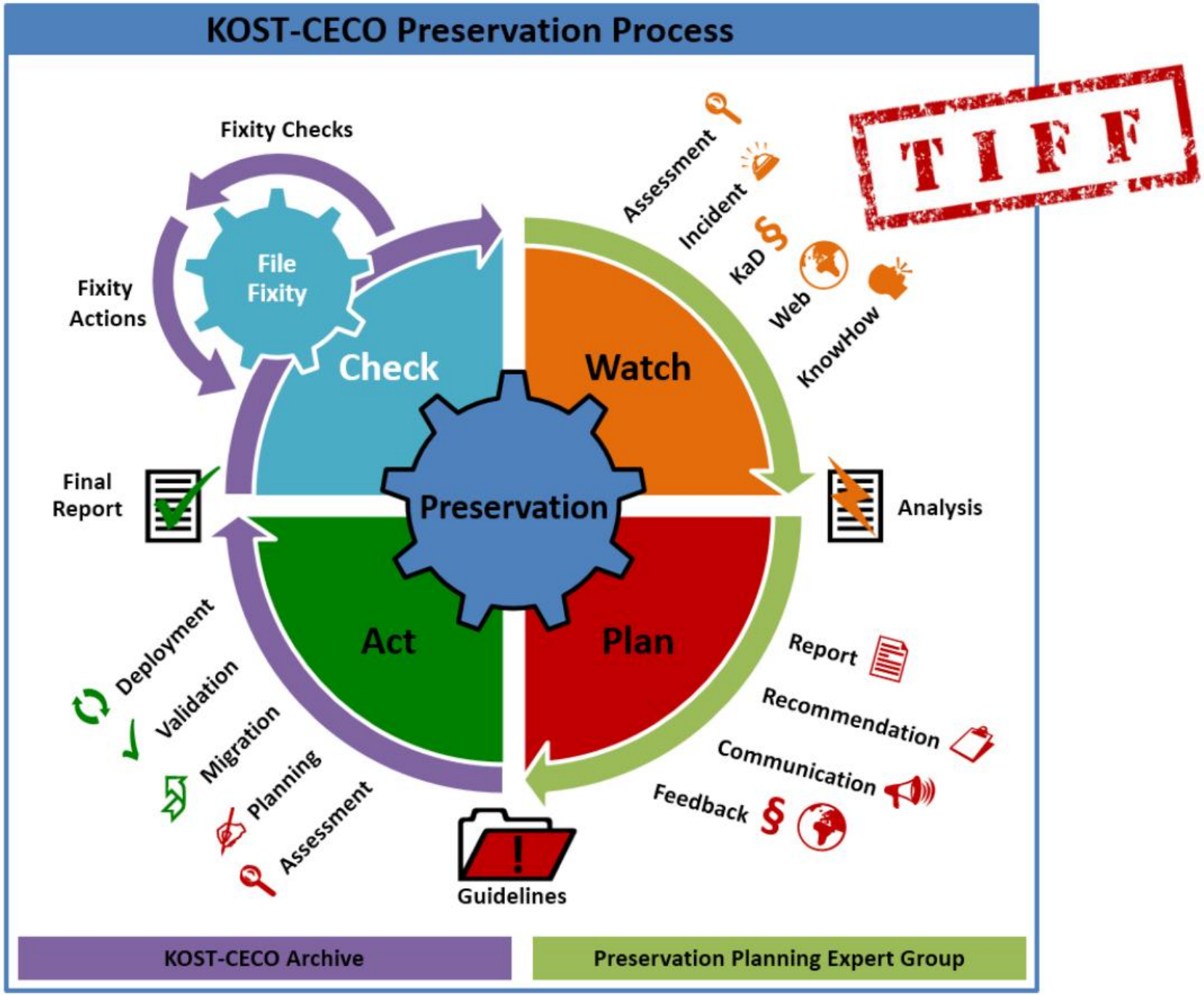
- Grösstenteils von grossen Institutionen und Forschungsprojekten
- Der Schwerpunkt liegt bei der Automatisierung der Bestandserhaltung
  - Z.B. SCAPE mit Scout, Plato, diversen Repository-APIs und C3PO



Preservation-Prozess der KOST

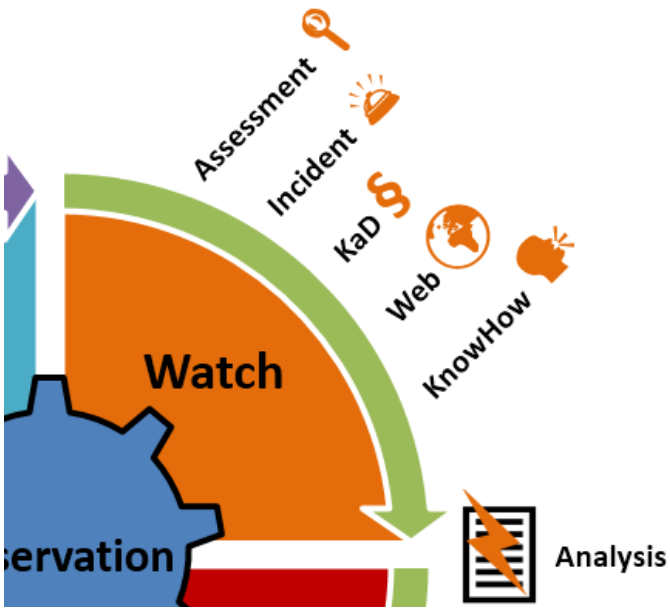
Warum braucht die KOST einen eigenen Prozess?

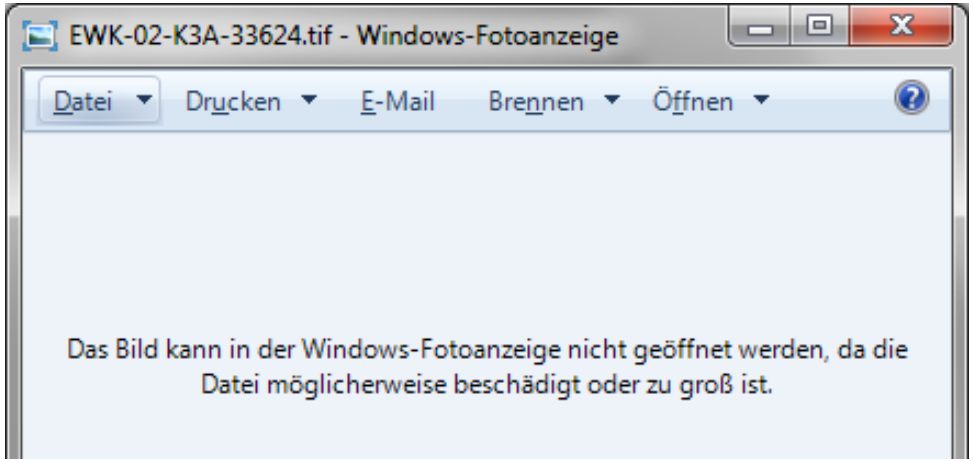
- Die umfassenden Preservation-Lifecycle-Implementationen wie zum Beispiel SCAPE sind u.E. zu komplex.
- Die KOST will ihre Preservation-Aktivitäten aber dennoch strukturiert organisieren und abwickeln.
- Deshalb entwickeln wir einen pragmatischen und flexiblen Preservation-Prozess, der die Verantwortung zwischen der Preservation-Planning-Expertengruppe (PPEG) und den einzelnen Trägerarchiven der KOST aufteilt.
  - Die PPEG besteht aus Experten von 8 Archiven und der Geschäftsstelle


KOST-CECO Preservation Process
Fixity Checks
File Fixity
Fixity Actions
Check
Watch
Assessment
Incident
KaD
Web
KnowHow
TIFF
Final Report
Preservation
Analysis
Act
Plan
Deployment
Validation
Migration
Planning
Assessment
Report
Recommendation
Communication
Feedback
Guidelines
KOST-CECO Archive
Preservation Planning Expert Group

Watch – Incident 🚨 und KaD §

Ein Mitgliedsarchiv konnte ein aus einem Digitalisierungsprogramm stammendes TIFF (Format aus dem KaD) mit dem Standard-Viewer nicht öffnen.

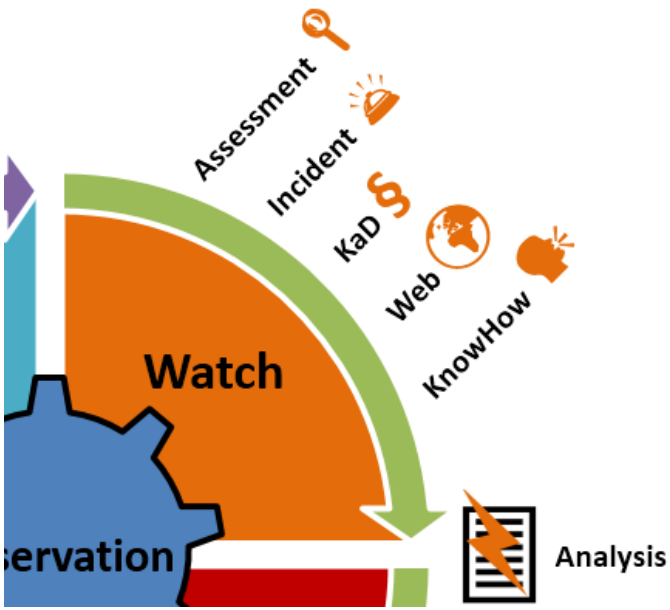




In der Folge wurde die KOST-Geschäftsstelle beigezogen.

Watch – Analysis

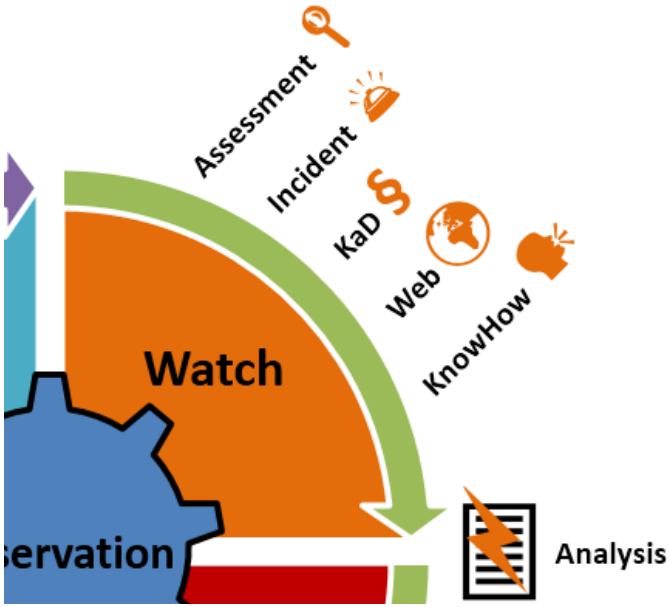
1. Die Komprimierungsart JPEG (6) wurde verwendet
   - Beschrieben in der TIFF-Spezifikation 6.0 von 1992
   - Ersetzt durch ISO JPEG (7) im Supplement 2 von 2002



- JPEG (6) wird nicht mehr unterstützt und sollte nicht verwendet werden

Watch – Analysis

2. Es gibt zu viele Möglichkeiten, ein gültiges TIFF zu erstellen
   - Spezifizierte Baseline TIFFs (Muss für alle Programme): Bilevel, Grayscale, Palette-color & RGB Full Color Images



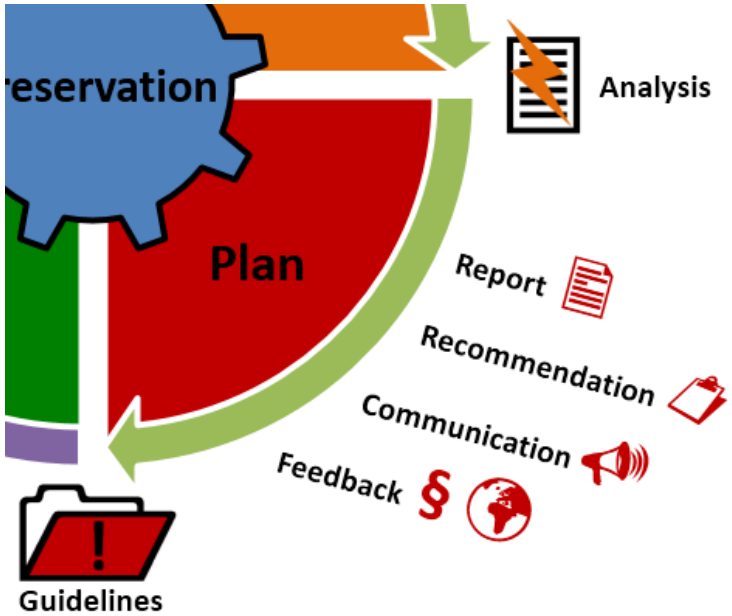
Erläuterung

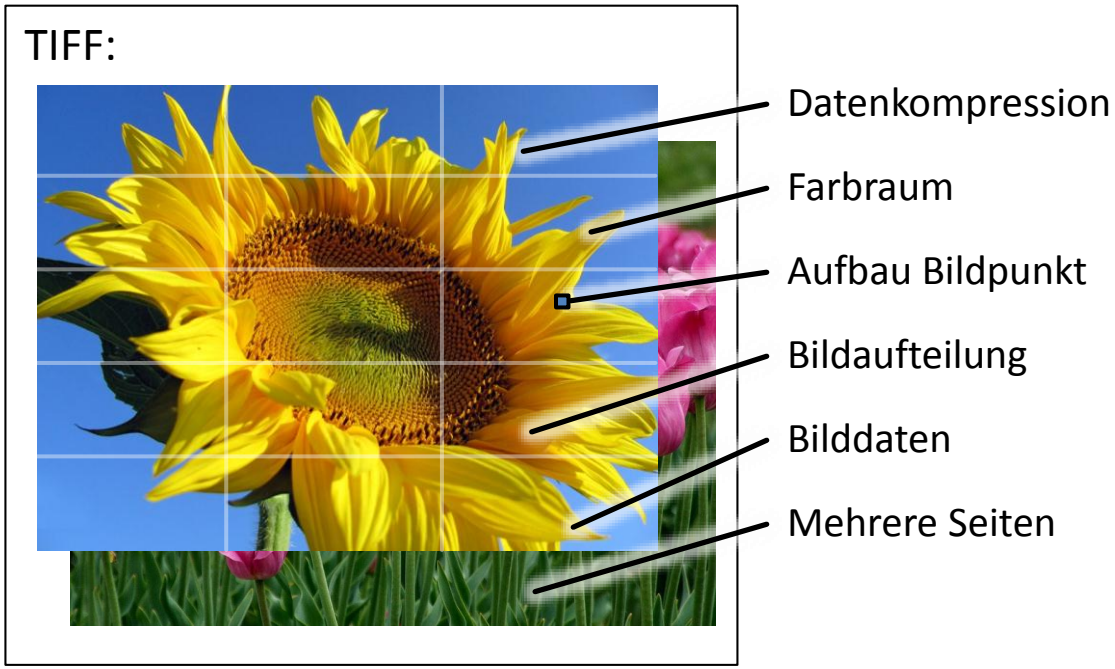
- Datenkompression
  - mind. 29 verschiedene Möglichkeiten
  - 9 in der Spezifikation und TN beschrieben
  - 3 als Baseline definiert
- Farbraum
  - mind. 14 verschiedene Möglichkeiten
  - 8 in der Spezifikation beschrieben
  - 4 als Baseline definiert

Plan – Report

Die Analysis wurde zu einem Report finalisiert

- Was ist ein TIFF und wie ist es aufgebaut?
- Probleme in der Praxis



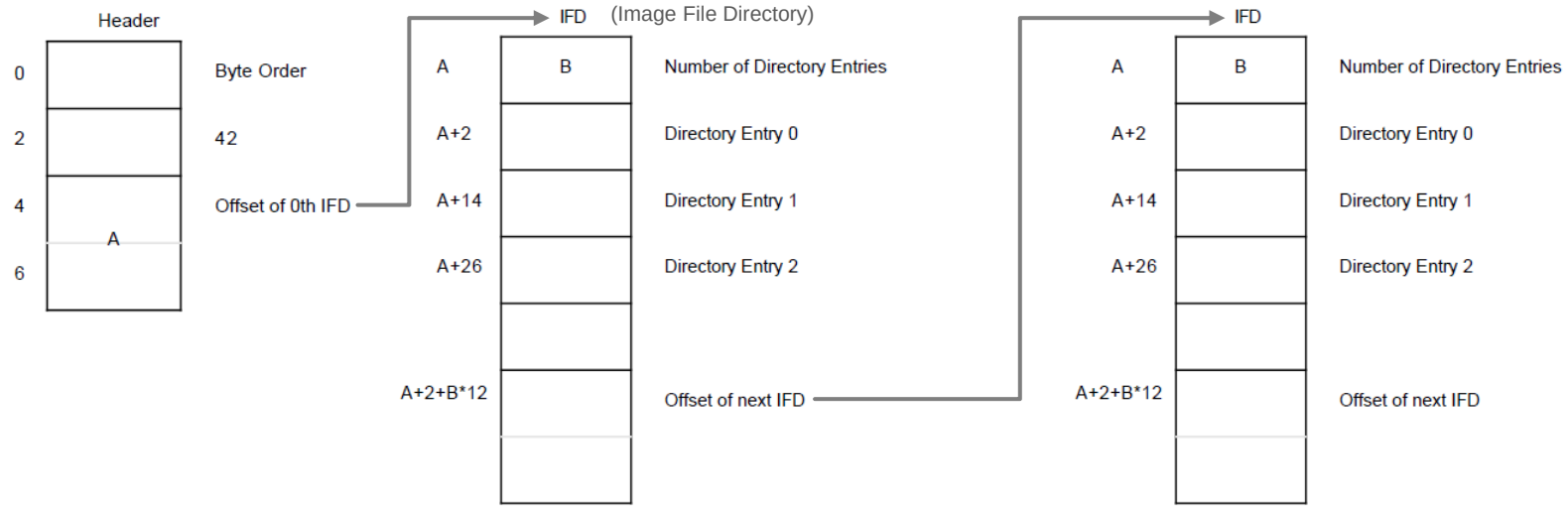


Plan – Report

Die Analysis wurde zu einem Report finalisiert

- Was ist ein TIFF und wie ist es aufgebaut?
- Probleme in der Praxis

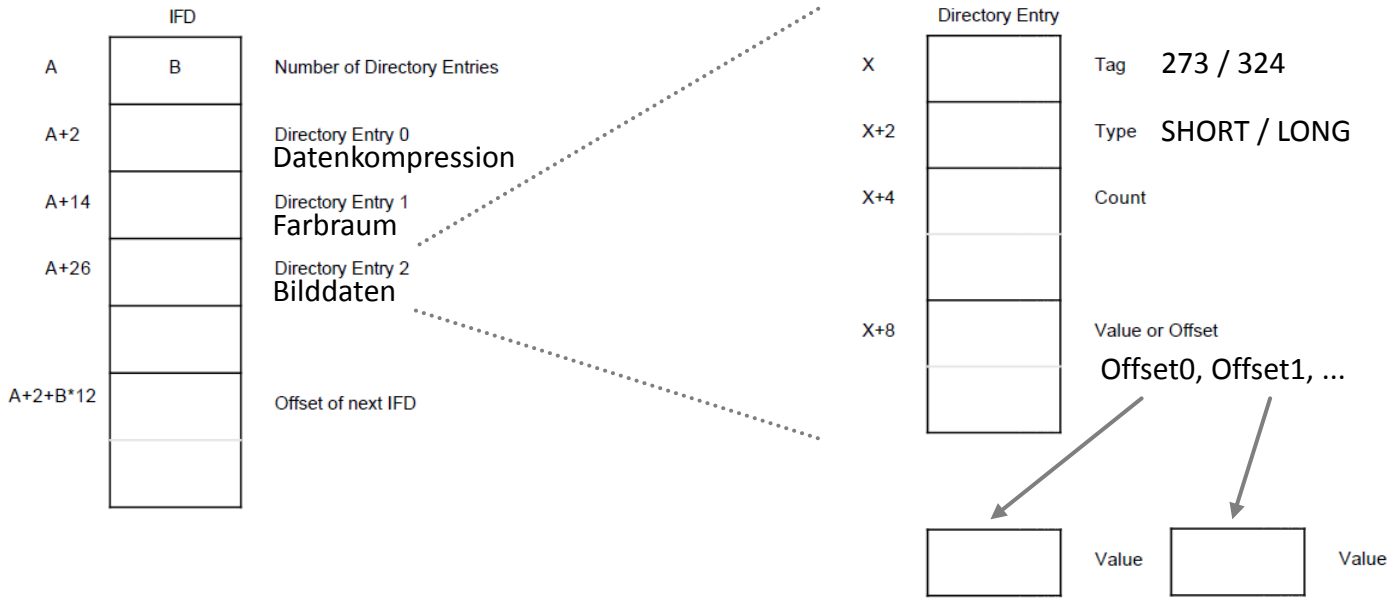
TIFF:



Plan – Report

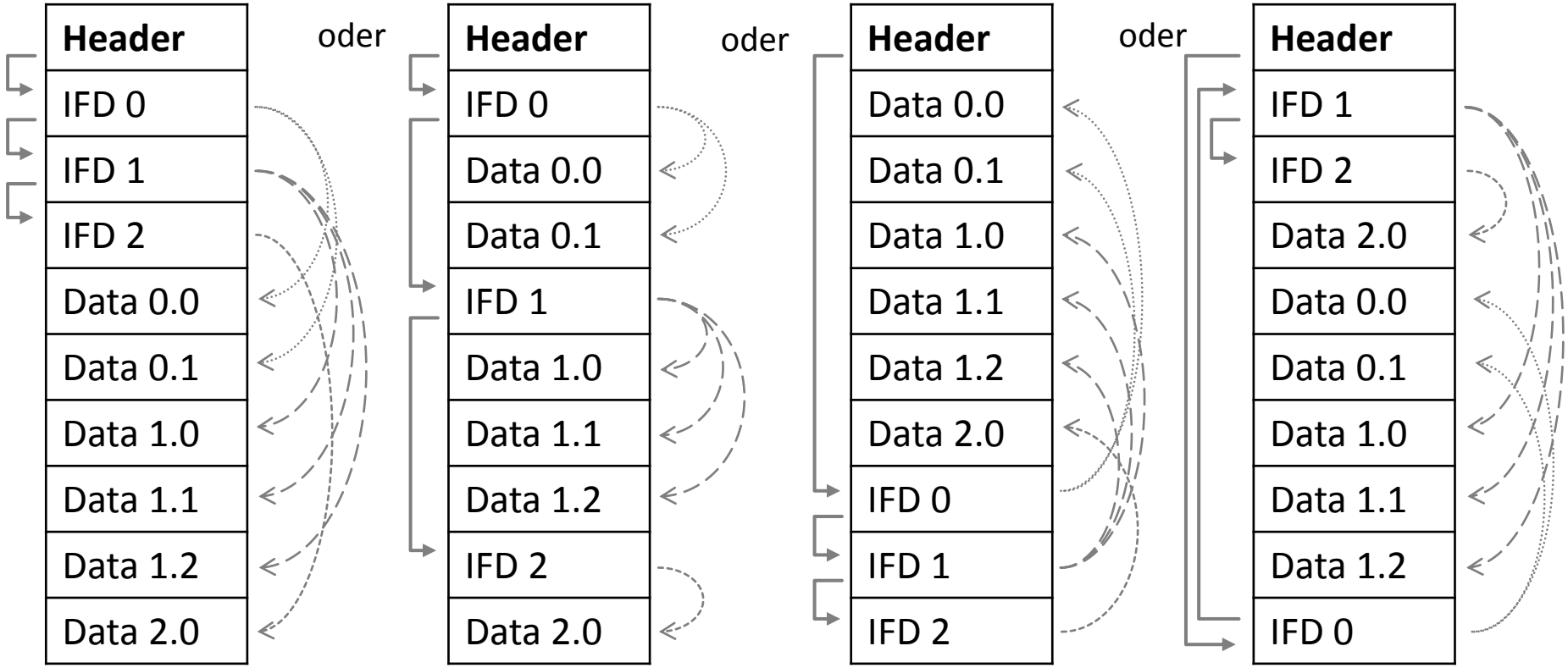
Die Analysis wurde zu einem Report finalisiert

- Was ist ein TIFF und wie ist es aufgebaut?
- Probleme in der Praxis



Plan – Report

Die Analysis wurde zu einem Report finalisiert

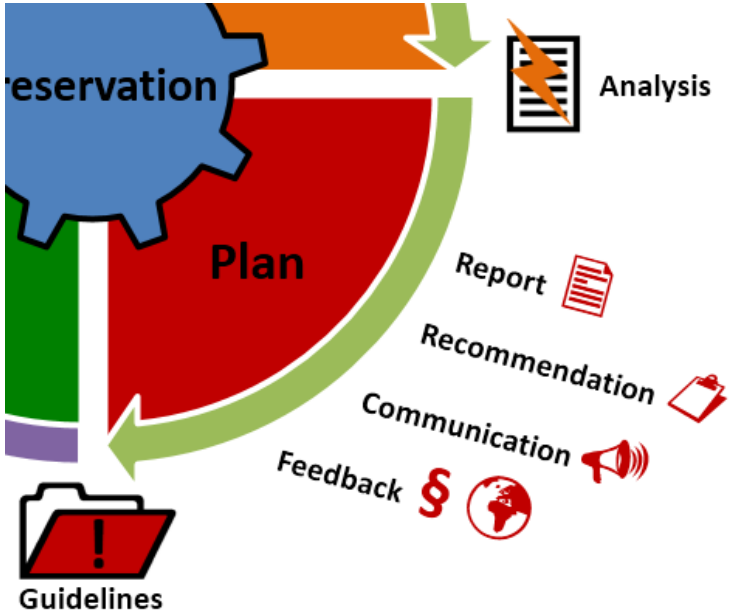


Plan – Recommendation

Die Recommendation enthält in diesem Fall

- TIFF-Präzisierung
- Konvertierungspriorisierung
- Konvertierungsanleitung

KOST-VAL



Datenkompression	259
uncompressed	(1)
CCITT 1D	(2)
CCITT Group 3	(3)
CCITT Group 4	(4)
LZW	(5)
JPEG	(6)
ISO JPEG	(7)
Deflate	(8)
PackBits	(32773)
Andere	

Farbraum	262
white is zero	(0)
black is zero	(1)
RGB	(2)
palette color	(3)
Transparency mask	(4)
CMYK	(5)
YCbCr	(6)
CIELab	(8)
Andere	

BitsPerSample	258
01	
04	
08	
16	
32	
Andere	

Bildaufbau
Streifen
Kacheln

Dateigrösse
≤ 1GB
> 1GB

Metadaten
Fehler in Metadaten

Legende:	Empfohlen	Nicht empfohlen	Kritisch

Empfehlung:

Sollte eine TIFF-Datei ein oder mehrere kritische Eigenschaften aufweisen, ist eine rasche Konvertierung vorzunehmen.

TIFF-Dateien mit mehreren nicht empfohlenen Eigenschaften sollten auch möglichst rasch konvertiert werden.

Enthält die TIFF-Datei nur eine nicht empfohlene Eigenschaft, liegt es im Ermessen des Archivs, ob und wann diese TIFF-Datei konvertiert werden sollte oder nicht.

TIFF-Präzisierung:			
Bezeichnung & Tag-Nr. (Subtag)	TIFF Baseline V6 (Part1)	Erweiterung	Einschränkung
Datenkompression 259			
uncompressed (1)	X		
CCITT 1D (2)	X		
CCITT Group 3 (3)		X	
CCITT Group 4 (4)		X	
LZW (5)		X	
PackBits (32773)	X		
Farbraum 262			
white is zero (0)	X		
black is zero (1)	X		
RGB (2)	X		
palette color (3)	X		
BitsPerSample 258			
01		X	
04	X		
08	X		
16		X	
Anzahl Seiten	Singelpage	Multipage	
Bildaufbau	Streifen		
Dateigrösse			≤ 1GB

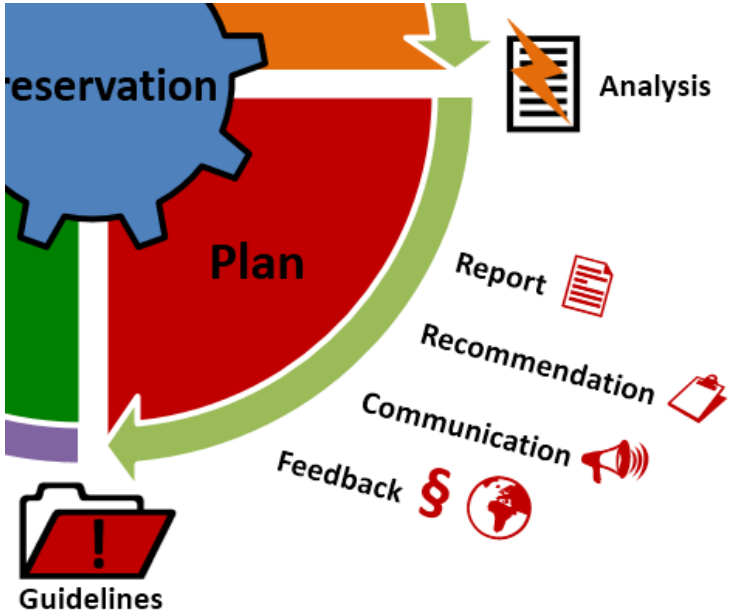
TIFF_Konvertierung_Kompression_Farbraum_V1.pdf
Konvertierungsanleitung im Bereich Bildaufbau

TIFF_Konvertierung_OutOfSequence_TypeMismatch_V1.pdf
Konvertierungsanleitung im Bereich Metadaten

Plan – Communication

Die TIFF-Erkenntnisse wurden der KOST mitgeteilt

- Workshop
- Newsletter

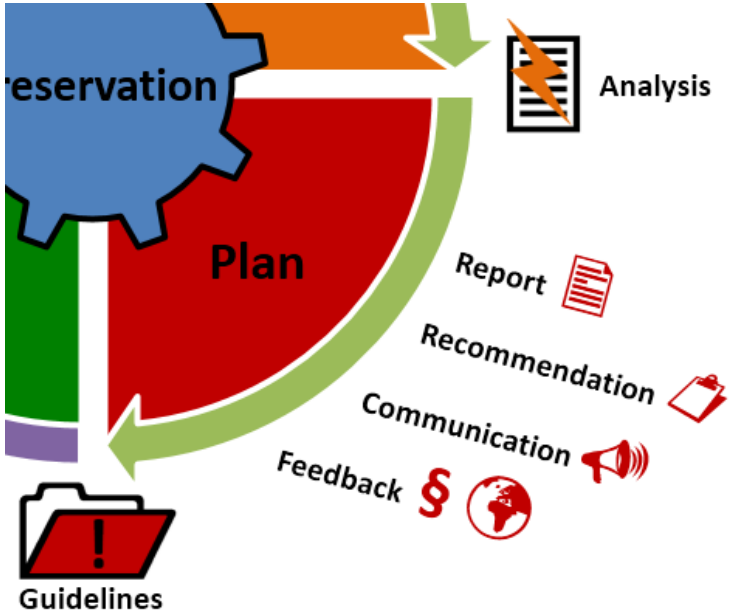


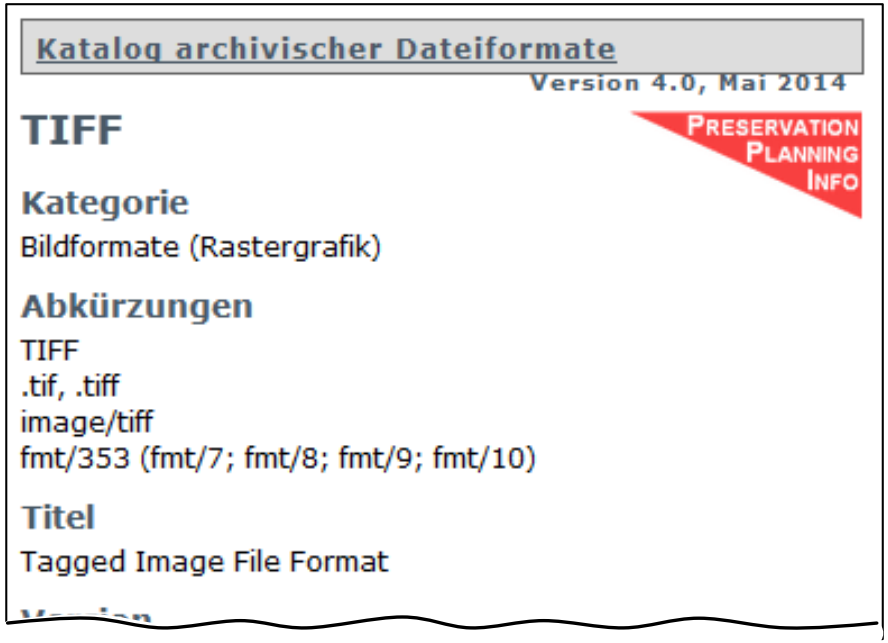
Plan – Feedback § und Guidelines

Die Erkenntnisse wurden auf der Homepage publiziert und flossen zurück in den KaD:

http://kost-ceco.ch/cms/index.php?preservation_tiff_de

http://www.kost-ceco.ch/wiki/whelp/KaD/pages/TIFF.html





Katalog archivischer Dateiformate

Version 4.0, Mai 2014

TIFF

PRESERVATION PLANNING INFO

Kategorie

Bildformate (Rastergrafik)

Abkürzungen

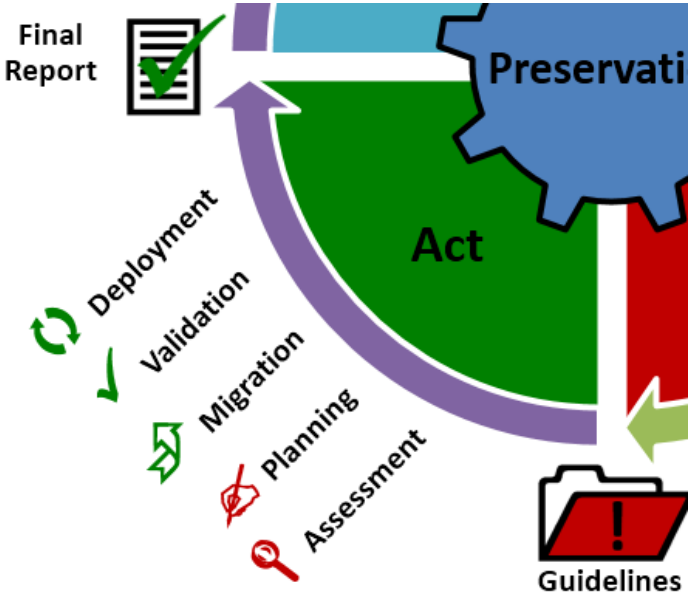
TIFF
.tif, .tiff
image/tiff
fmt/353 (fmt/7; fmt/8; fmt/9; fmt/10)

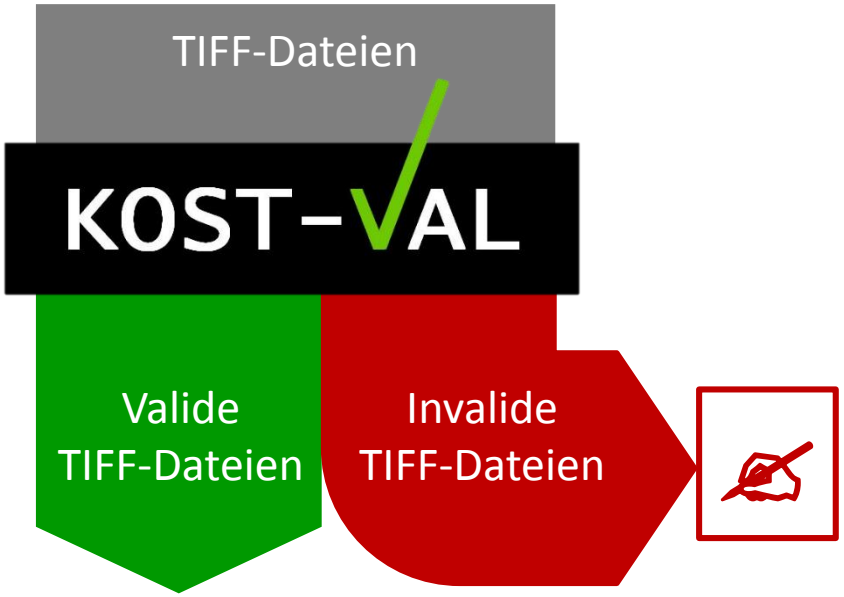
Titel

Tagged Image File Format

Act – Assessment und Planning

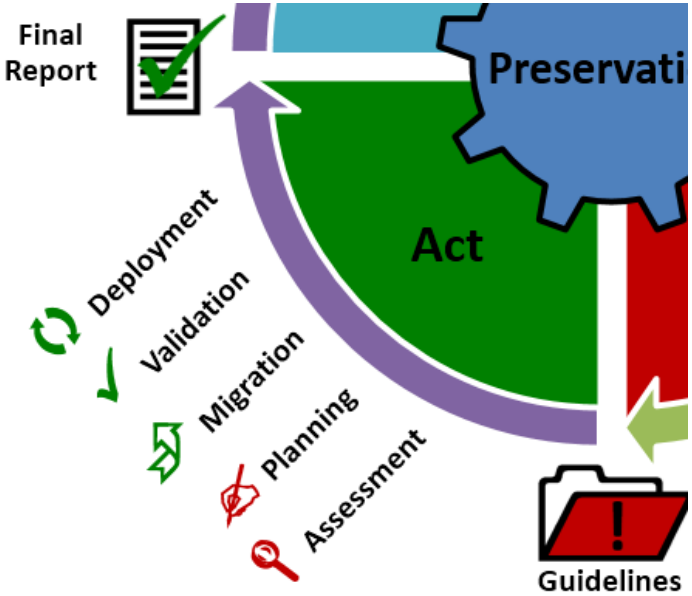
Die einzelnen KOST-Archive können danach mit den vorgeschlagenen Mitteln ein Assessment durchführen und ihr inviduelles Planning ausarbeiten.

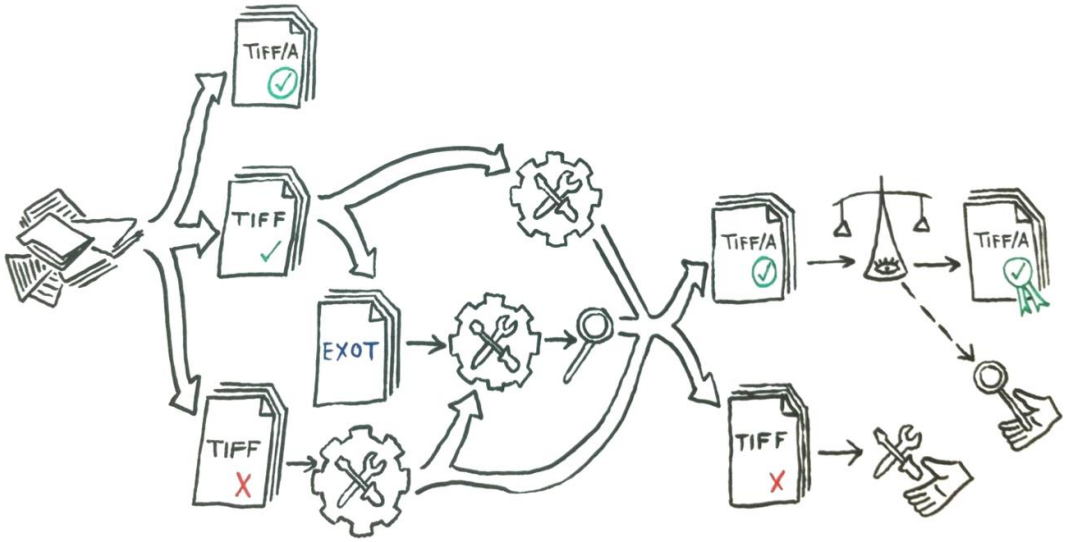




Act – Migration und Validation ✓

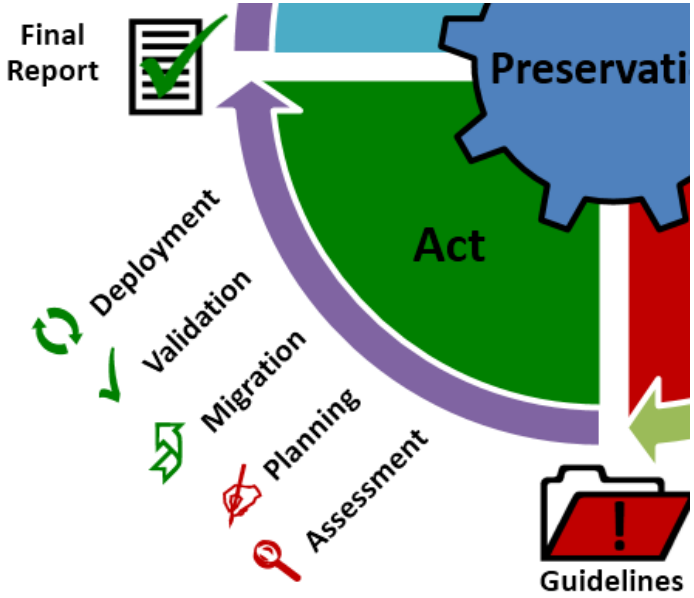
Die Preservation Actions werden durch die KOST-Archive selbständig durchgeführt.
Die PPEG steht bei Fragen und Problemen zur Verfügung und zeigt Möglichkeiten auf.





Act – Deployment und Final Report

Abgeschlossen werden die Preservation Actions durch das Zurückspielen in den Archival Storage zusammen mit der Dokumentation und Aktualisierung im AIS.



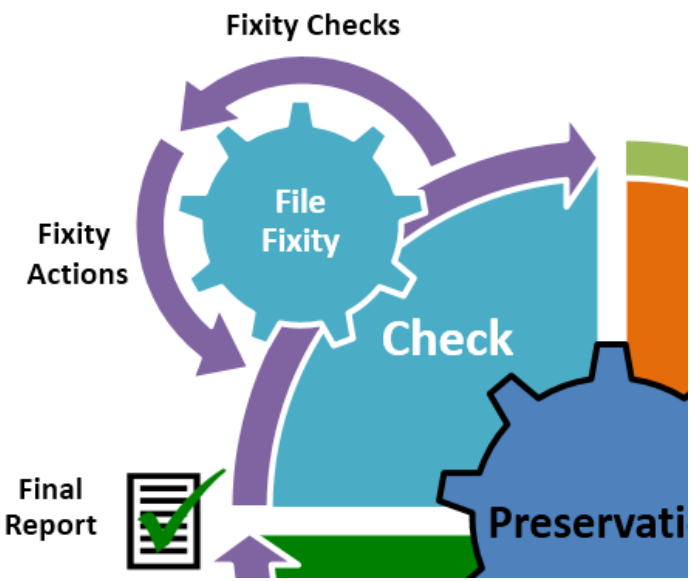
TIFF =

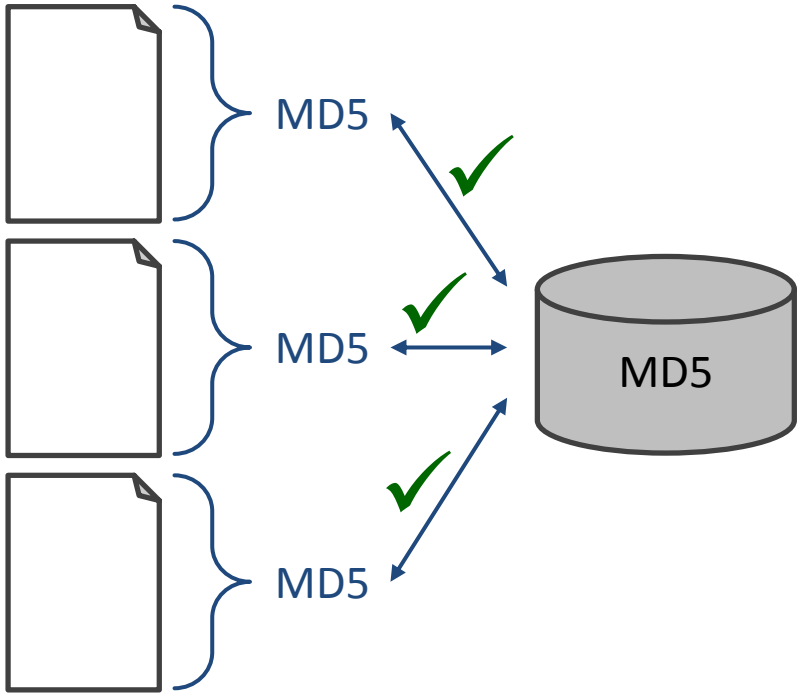
☒ ~~Thousands of Incompatible File Formats~~

☑ Tagged Image File Format

Check – File Fixity: Check und Action

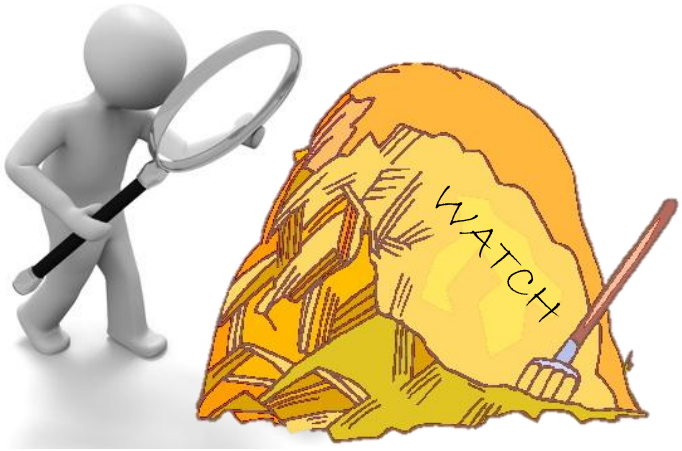
Unabhängig vom restlichen Preservation-Prozess werden in den einzelnen Archiven die Archivalien anhand ihrer Checksummen überprüft und ggf. mit einer unbeschädigten Kopie ersetzt.





Erkenntnisse aus der KOST-Praxis

Die Schwierigkeit liegt im Schritt „Preservation Watch“: die sprichwörtliche Suche nach der Nadel im Heuhaufen.



- Auslöser bei einem Dateiformat aus dem KaD
- Quelle: PPEG, KOST, Dritte, Web

Erkenntnisse aus der KOST-Praxis

Die Auslöser erscheinen zunächst meist als banale kleine Probleme, deren Relevanz für die Preservation erst erkannt werden muss:

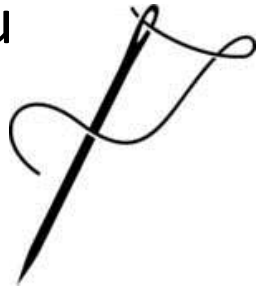
- Probleme beim Öffnen
- Erstes Druckergebnis entspricht nicht dem Bildschirminhalt
- Hardwarebug-Meldungen aus dem Web

Erkenntnisse aus der KOST-Praxis

- Wie jeder Prozess kann auch der KOST-Preservation-Prozess nicht immer eingehalten werden.
- Es gibt Fehler, welche irreversibel sind und wo einzig die Quelle sensibilisiert werden kann, damit nicht mehr solcher Dateien erstellt werden.
- Die Ressourcen sind eine weitere Schwierigkeit. Es beansprucht oftmals Monate, die Analyse zu erstellen.

FAZIT

Wenn bei Ihnen ähnliche Probleme mit einem zentralen Archivformat existieren oder auftauchen, nehmen Sie sich die Zeit, sie zu analysieren oder wenigstens die KOST zu informieren.
Wir interessieren uns für jede potentielle Nadel.



KOST Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen

Ein Gemeinschaftsunternehmen von Schweizer Archiven

Claire Röthlisberger-Jourdan
claire.roethlisberger@kost.admin.ch