

Von der Theorie zur Praxis: Bestandserhaltung digitaler Unterlagen

Workshop des KLA-Ausschusses Digitale Archive

Bundesarchiv Koblenz, 7./8. November 2018

Tagungsdokumentation

Redaktion:
Tim Dahlmanns
Kai Naumann



Bedienungsanleitung

Dies ist ein PDF-Portfolio. Es enthält kurze Mitschriften des Vortrags, die gestellten Fragen und die PowerPoint-Folien.

Sie können die im PDF eingebetteten Lesezeichen benutzen, um an bestimmte Abschnitte zu springen.

Die Konferenzsprache war primär Deutsch, aber immer wieder verliefen Diskussion und Darstellung auf Englisch, was sich in Folien und Diskussionen widerspiegelt.

Liste der Beiträge

in der Reihenfolge des Workshops:

Kai Naumann | Einleitende Bemerkungen

Christian Keitel | Überblick zur digitalen Bestandserhaltung

William Kilbride | The Bit List of Digitally Endangered Species

Tobias Steinke | Organisations- und Rechtsaspekte der Emulation

Martin Wrigley | The Preservation Action Registry

Michelle Lindlar | Bestandserhaltung mit Ex Libris Rosetta aus der Perspektive einer großen Forschungsbibliothek

Yvonne Tunnat | Ob ihr wirklich richtig steht, seht ihr, wenn das Licht angeht – die nestor AG Formaterkennung

Jenny Mitcham (via Skype) | Making WordStar files accessible: An imperfect file migration case study

Siegfried Fößel | Nachhaltige Videoformate

Martin Hoppenheit | Qualitätssicherung bei der Datenbankarchivierung

Andreas Trautheim-Hofmann | Erhaltung von 3D Ingenieurdokumenten und Produktdaten

Michelle Lindlar | Archivierung von Building Information Models (BIM) - Risiken und Möglichkeiten

Wolfgang Kresse | Erhaltung von Geodaten (ISO 19165-1 und ISO/TC 211)

Rainer Jacobs | Extrahieren von Dateien aus Containern

Michael Tobegen | Entwicklung eines DIMAG-Bestandserhaltungsmoduls: Werkstattbericht

Kai Naumann | Einleitung

Das Dunkle Zeitalter 2.0 sei nahe, prophezeite Bryan Bergeron 2001 in einem englischen Sachbuch.¹ Demnach werde die Menschheit erstens den Zugriff auf ihre Backups verlieren und zweitens in älteren Dateiablagen nur noch Kauderwelsch vorfinden, mangels passender Softwareprodukte zur Darstellung. Die heutige Lage ist besser als damals gedacht, denn eine breite Koalition von Fachleuten kümmert sich um das Problem.

Im vergangenen Jahrzehnt sind digitale Bestandserhaltung und digitale Nachhaltigkeit weltweit aus der Phase der Theorie und der Konzeption in die Phase der Umsetzung getreten. Während die Praxis in der deutschen Bibliotheks- und Archivwelt bis 2010 vor allem von Digitalisierung im Sinne von Scannen handelte, so stellt dieses Geschäft inzwischen eine Routinehandlung dar.

Die Herausforderung von heute ist, die großen Mengen erzeugter digitaler Objekte, die in den Archiven und in den Behörden entstanden sind, für eine dauerhafte Verwahrung zu ertüchtigen. Die immensen Mittel, die derzeit in der Wissenschaftsförderung für Forschungsdatenmanagement ausgeschüttet werden, dienen ähnlichen Zielen. Nicht zuletzt sind inzwischen Bücher erschienen, die eine Gesamtschau des Problems bieten, es historisch und gesellschaftlich einbetten und uns den Mut geben, einfach selbst die Verantwortung zu übernehmen.²

Wie wir mit dem Thema „digitale Nachhaltigkeit“ die Wirtschaft, die öffentliche Verwaltung und die gesamte Gesellschaft mitnehmen können, erklärte William Kilbride (Digital Preservation Coalition DPC, Glasgow) mit seinem Vortrag recht eindrucksvoll:

- Wir reden über undeutliche Gefahren für Daten oder Software, aber die Verluste sind echt und betreffen uns alle im Alltagsleben.
- Dies ist eine soziale und technische Herausforderung. Wenn wir nur technische Lösungen anbieten, haben wir keinen Erfolg.
- Digitale Bestandserhaltung handelt nicht von einem „dunklen Zeitalter“. Es geht darum, dem Versprechen des digitalen Zeitalters gerecht zu werden.

Am World Digital Preservation Day, dem 29. November, haben auch Archive die Gelegenheit, dieses Anliegen in ihre Öffentlichkeitsarbeit einzubauen!

Bei der gemeinsamen Aufgabe hilft inzwischen eine größere Zahl von Softwareprodukten, deren Nutzer im Rahmen des Workshops zusammenkamen. Dabei kristallisiert sich heraus, dass die zentralen Systemkomponenten meist kostenpflichtig sind, während sich für viele Peripherieprodukte eine gemeinsame Nutzung anbietet. Ein Beispiel sind Erkenner und Validierer wie DROID, JHOVE oder veraPDF, die aber bei weitem noch nicht ausreichen

¹ Bryan Bergeron: Dark Ages II. When the Digital Data Die, Upper Saddle River 2002.

² Christian Keitel: Zwölf Wege ins Archiv. Umriss einer offenen und praktischen Archivwissenschaft, Stuttgart 2018; Trevor Owens: The Theory and Craft of Digital Preservation, Baltimore 2018; Heather Ryan, Walker Sampson: The No-nonsense Guide to Born-digital Content, London 2018. Zur Archivaliengattung der Dateisammlung auch Kai Naumann, Michael Puchta: Kreative digitale Ablagen und die Archive, München 2017.

und gemeinsam ergänzt werden müssen. Wer weiß, ob das von Martin Wrigley bei dem Workshop vorgestellte PAR eine weitere Lücke füllt?

In der Diskussion wurde dargelegt, dass bei Formaten, die ISO-Standards darstellen, mit dem Auslaufen der mit dem Standard verbundenen Patente (nach 20 Jahren) keine Entgeltforderungen durch die am Standard beteiligten Parteien mehr möglich sind. Wäre diese These untermauert, so könnte dies ebenfalls eine Erleichterung der Bestandserhaltung darstellen.

Anerkannt wurde in der Runde der Bedarf, Archive und Bibliotheken in gemeinsamen Bestrebungen zur digitalen Langzeitarchivierung näher zusammenzurücken. Bei nestor wird der Erfahrungsaustausch bereits gepflegt, doch lässt sich die gegenseitige Beteiligung von Teilnehmern aus der jeweils anderen Community noch ausbauen. Auf operativer Ebene existiert zwar ein bereits guter Erfahrungsaustausch via z.B. gemeinsamer Gremienarbeit in nestor, doch scheint es auf strategischer Ebene oft noch Barrieren in Form von als zu groß erachteten Unterschieden zwischen der digitalen Archivierung in Archiven und Bibliotheken zu geben.

Mehrere Teilnehmer lobten die Zusammensetzung der Runde, die auf 40 Personen beschränkt war. Die Plätze wurden nicht nach der Reihenfolge der Anmeldung vergeben, sondern die Einladungen gingen an die verschiedenen Sparten des Archiv- und Bibliothekswesens und wurden mit Quoten versehen. Die entsprechenden Communities schickten ausgewählte Personen, so dass eine wirklich heterogene, aber auch diskussionsfähige Gruppe entstand.

Viel Freude mit dieser Publikation! Mal sehen, ob es so etwas bald wieder gibt.

Kai Naumann, März 2019

KLA Konferenz der Leiterinnen und
Leiter der Archivverwaltungen
des Bundes und der Länder

Digitale Bestandserhaltung Von der Theorie zur Praxis

Workshop, Koblenz, 7./8. November 2018

My most impressing encounters in digital preservation:

- My own work in the DIMAG project
- Ange Albertini's talk „Funky File Formats“ (2014)
<https://www.youtube.com/watch?v=hdCs6bPM4is>
https://www.youtube.com/watch?v=UxM3q6eIE_Q
(mit deutscher Synchronstimme)
- David H. S. Rosenthal's blogposts on strategy in preservation
<https://blog.dshr.org/2007/05/format-obsolescence-prostate-cancer-of.html>
<https://blog.dshr.org/2015/09/the-prostate-cancer-of-preservation-re.html>

Threats to digital information

- Budget cuts
- Decay
 - of media only
- Obsolescence
 - of media and reading devices
 - of formats
 - file formats
 - file component formats (a font, an embedded picture)
 - structure formats (a records management system being retired)

Threats and strategies to cope with them

- Budget cuts => demonstrating value
- Decay
 - of Media only => transfer to better storage systems
- Obsolescence
 - of media and reading devices => transfer to better storage systems
 - of formats
 - file formats => emulation and/or format migration
 - file component formats
 - structure formats

Special requirements for archives

- Sharing content is not allowed (data protection)
- Designated community very broad
(from private household to specialised academic)
- All kinds of media types (records, photos, sound, moving images, databases, geodata)
- All kinds of source systems (many platforms, solutions, vendors)
- Emphasis von authenticity (when did who what on this file?)

What we have learned

- There is no miracle solution
- It is useless to look for the „best media“
- Rejecting objects that „don't preserve well“ is no option
- We don't need handicraft, but industrial-scale facilities
- We must have reliable metadata for preservation.
 - But: For using our objects, no metadata set can replace software
- Licenses for software are an important issue

What we want to learn

- Do our concepts of digital information meet our practical ends?
- Where are good tools that perform well at identification, emulation, migration?
- How are they integrated into systems like Archivematica, DIMAG, DiPS, Rosetta and Preservica?
- What amount of validation, at which point in the lifecycle?
- Intellectual property rights are an obstacle – how can we arrange with them?
- Can we share knowledge for good practice across system borders?

Christian Keitel | Überblick zur digitalen Bestandserhaltung

Prof. Dr. Christian Keitel (Landesarchiv Baden-Württemberg) gibt in seinem Vortrag einen umfassenden Überblick zur digitalen Bestandserhaltung und den damit verbundenen Herausforderungen für aktuelle und zukünftige Generationen.

Er wirft eingangs die Frage auf, ob aktuelle Herangehensweisen an das Thema digitale Bestandserhaltung kritisch hinterfragt und an archivische Anforderungen angepasst werden müssen. So gelte aktuell die Blockchain-Methode aufgrund des Anscheins einer beweisbaren Unveränderbarkeit als neuer „heiliger Gral“ der digitalen Bestandserhaltung. Die Vertreter von Blockchain kämen in der Regel zwar nicht aus der Archiv-Szene. Diese müsse daher definitiv beratschlagen, ob Blockchain als Lösungsansatz wirklich in Frage kommt, und sich gegenüber den Befürwortern entsprechend äußern.

Verallgemeinernd stellt er die These auf, dass die Dauer der Archivierung die notwendigen Erhaltungsmethoden bedingt. So ist zum Beispiel für den Zeitraum von fünf Jahren die Bitstream Preservation mit Generierung von Hashwerten ausreichend, während bei einem angestrebten Erhaltungszeitraum von bis zu 250 Jahren mit einer unbestimmten Anzahl von Format-Migrationen oder Emulationen sowie der Nachhaltung signifikanter Eigenschaften gerechnet werden müsse.

Im Zuge seines Vortrages geht Keitel konkreter auf den Bereich der „signifikanten Eigenschaften“ ein und stellt hierzu verschiedene internationale Beispiele vor. Er hebt dabei hervor, dass im Zuge der grundsätzlichen Diskussion über signifikante Eigenschaften auch darüber gesprochen werden muss, was keine signifikanten Eigenschaften sind. Die Begrifflichkeit darf nicht zu einem Sammelbecken werden. Signifikante Eigenschaften sind eine Teilmenge aller Eigenschaften eines digitalen Objekts, die als Maßstab („Benchmark“) für Erhaltungsmaßnahmen dienen. Zudem sind ohne fixierte Definitionen an dieser Stelle ohnehin keine Fortschritte zu erwarten. Auch müssen künftige Diskussionen und Definitionen auf vorangegangenen aufbauen können.

All dies funktioniert nur durch einheitliche Metadatenschemata, die rückwärtskompatibel sind. Nur durch die Vergleichbarkeit von früheren Metadaten mit aktuellen Metadaten kann das ganze Problem bewältigt werden.

Zu Verdeutlichung stellt Keitel das Konzept des Performance Models vor.



Das Ziel des Modells ist es, mit Hilfe von Hard- und Software ähnliche Leistungsergebnisse (Performances) über einen langen Zeitraum hinweg sicherzustellen. Die konkreten Hilfsmittel bei Hard- und Software sind dabei auswechselbar. Wichtig ist, dass die signifikanten Eigenschaften bei allen künftigen Performances erfüllt werden. Zum Schluss des Vortrags weist Keitel auf die verwandten Handlungsfelder hin, die sich über die Grenzen der akademischen Fachgemeinden hinweg auftun, und auf die vielen offenen Fragen, die das Tagungsthema mit sich bringt.

Diskussion und Fragen:

1. Wulf Pingel: Welche Rolle und welchen Einfluss haben die eingesetzten Liefersysteme in Bezug auf die digitale Bestandserhaltung?

A: Letztlich bestimmen die eingesetzten Liefersysteme einen gewichtigen Teil der Bestandserhaltung in den Archiven. Die Idee, durch archivistisches "Records Management" alle Liefersysteme zu normieren, benötigt viele Ressourcen und ist daher nicht immer praktikabel. Von 1996 bis 2007 stand dieser Bereich in Deutschland im Mittelpunkt. Die Übernahme von bereits bestehenden digitalen Unterlagen ist erst später in den Fokus des Interesses gerückt.

2. Michelle Lindlar: Erscheint es sinnvoll, die Erhebung signifikanter Eigenschaften bereits bei der ersten Migration in Betracht zu ziehen und die Produzenten der Daten frühzeitig in die Erhebungsprozesse einzubinden?

A: Jede Veränderung sollte grundsätzlich hinterfragt (Haben wir unser Ziel erreicht?) und dokumentiert werden. Die gefragten Eigenschaften sollten bei jeder einzelnen Erscheinungsform erfasst und in den Metadaten nachgehalten werden.

3. Tobias Steinke: Worin besteht die Bedeutung der Bestandserhaltung bei signifikanten Eigenschaften?

A: Die Behandlung des Themas „Signifikante Eigenschaften“ befindet sich noch am Anfang. Es ist wichtig, dass die Community nach vorne denkt und sich eingehend mit der Definition von Standards befasst. Dabei ist jedoch zu betonen, dass einzelne Archive individuelle Spielräume benötigen.

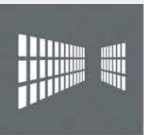
4. Martin Wrigley: Ist Blockchain an sich bereits als archivfähig anzusehen?

A: Einerseits ja, da es unveränderbar ist. Auf der anderen Seite werden wir möglicherweise auch in Zukunft in diesem Bereich ebenso wie in allen anderen Bereichen Veränderungen vornehmen müssen, die dann mit der Unveränderbarkeit der Technologie kollidieren könnten. Auf jeden Fall sollte der Komplex näher untersucht werden.

Digital preservation at a glance

Christian Keitel

7.11.2018



Digital preservation 2018

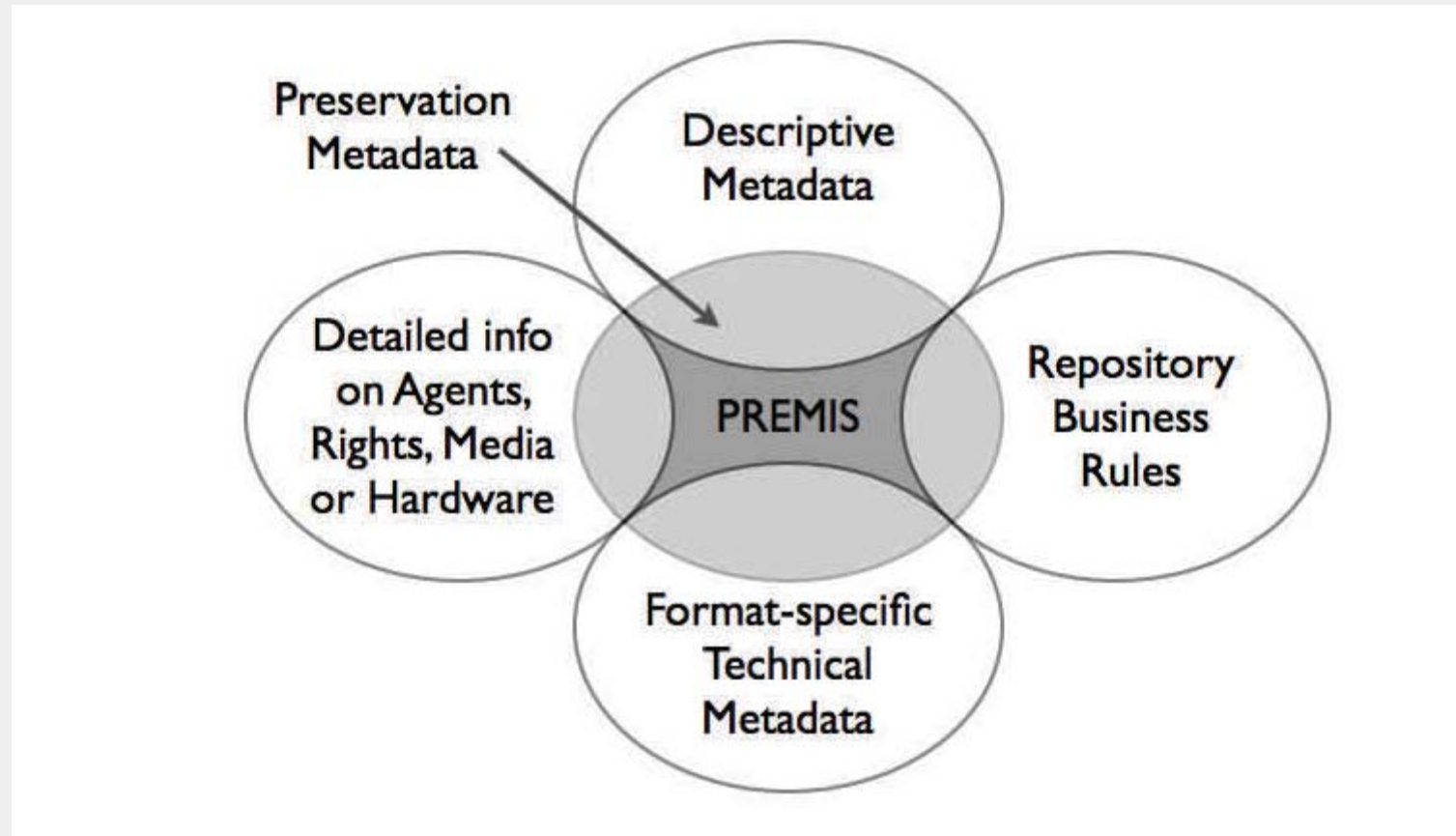
- > 50 years of vague discussions
- > 20 years of intensive discussions
- Growing practical experiences

>>>>

- Numerous little steps
- Growing need to see the bigger picture

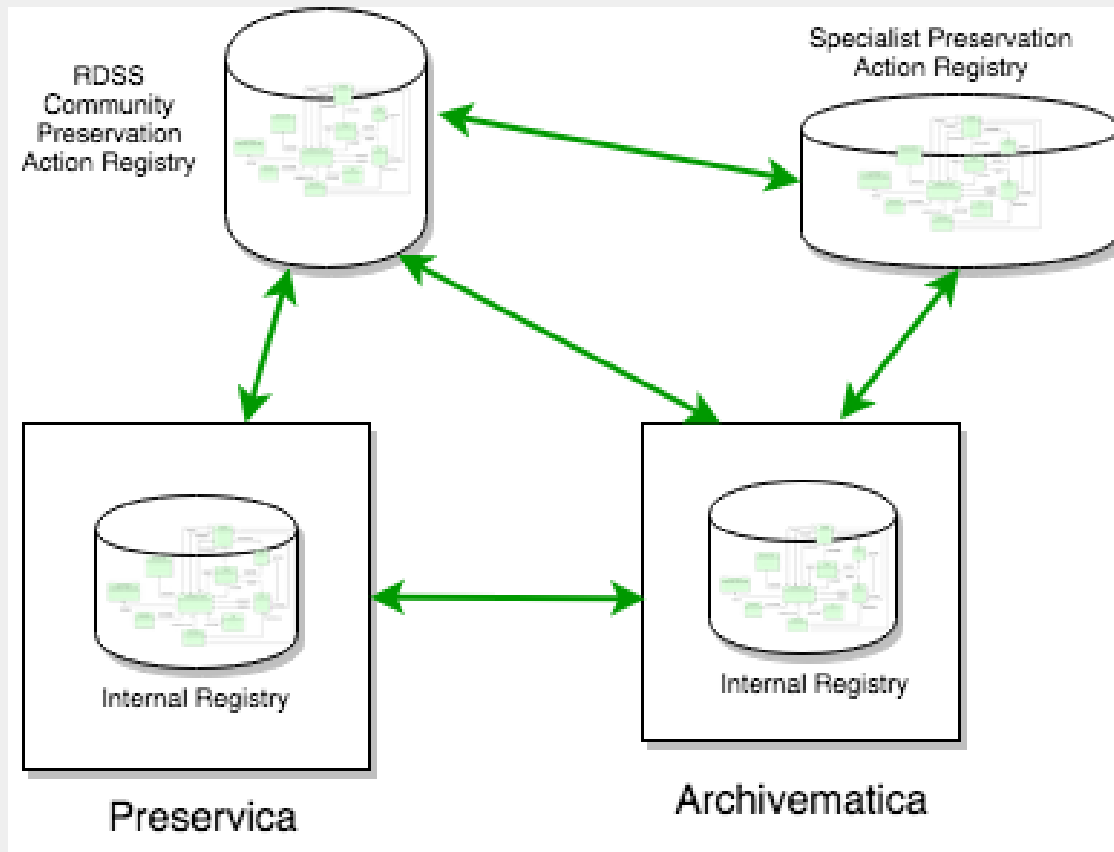


Conceptualisation 1: Related areas



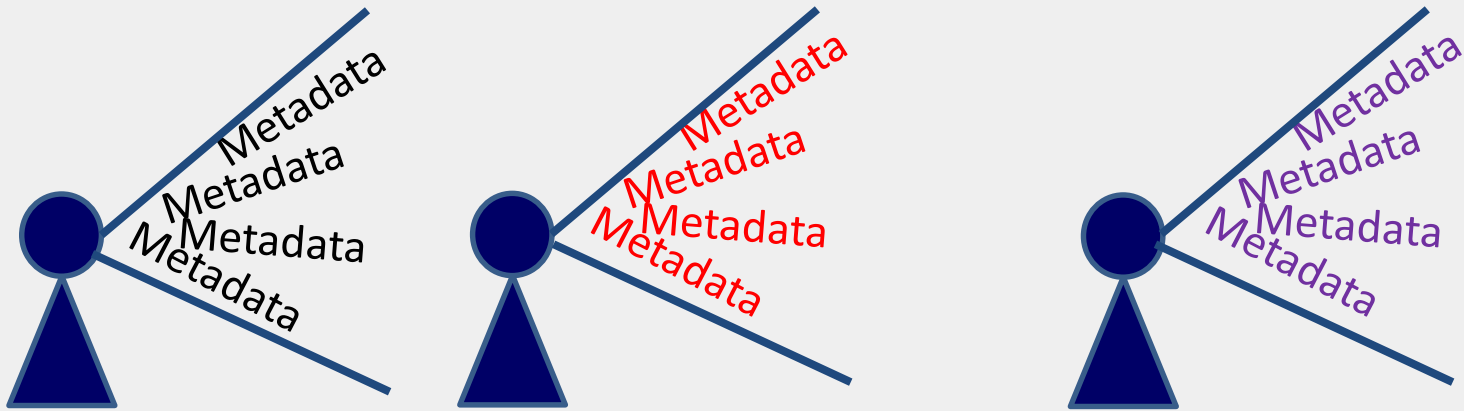
Priscilla Caplan, Understanding Premis

Conceptualisation 2: An ecology of systems



Matthew Addis, Justin Simpson, Jon Tilbury, Jack O'Sullivan, Paul Stokes,
Digital Preservation Interoperability through Preservation Actions Registries

Conceptualisation 3: A Communication Scenario



Are changes necessary?

- OAIIS: To preserve information in spite of „the impacts of changing technologies, including support for new media and data formats, and of a changing Designated Community.“
 - Migration or emulation
 - Changes necessary
- Blockchain???
 - Unchangeable

Duration of archiving

	Preservation method	Validation
• 5 years?	Bit preservation	Hash values
• 25 years?	1 format migration	Comparison
• 250 years?	n format migrations	Significant properties



Acceptable alteration?

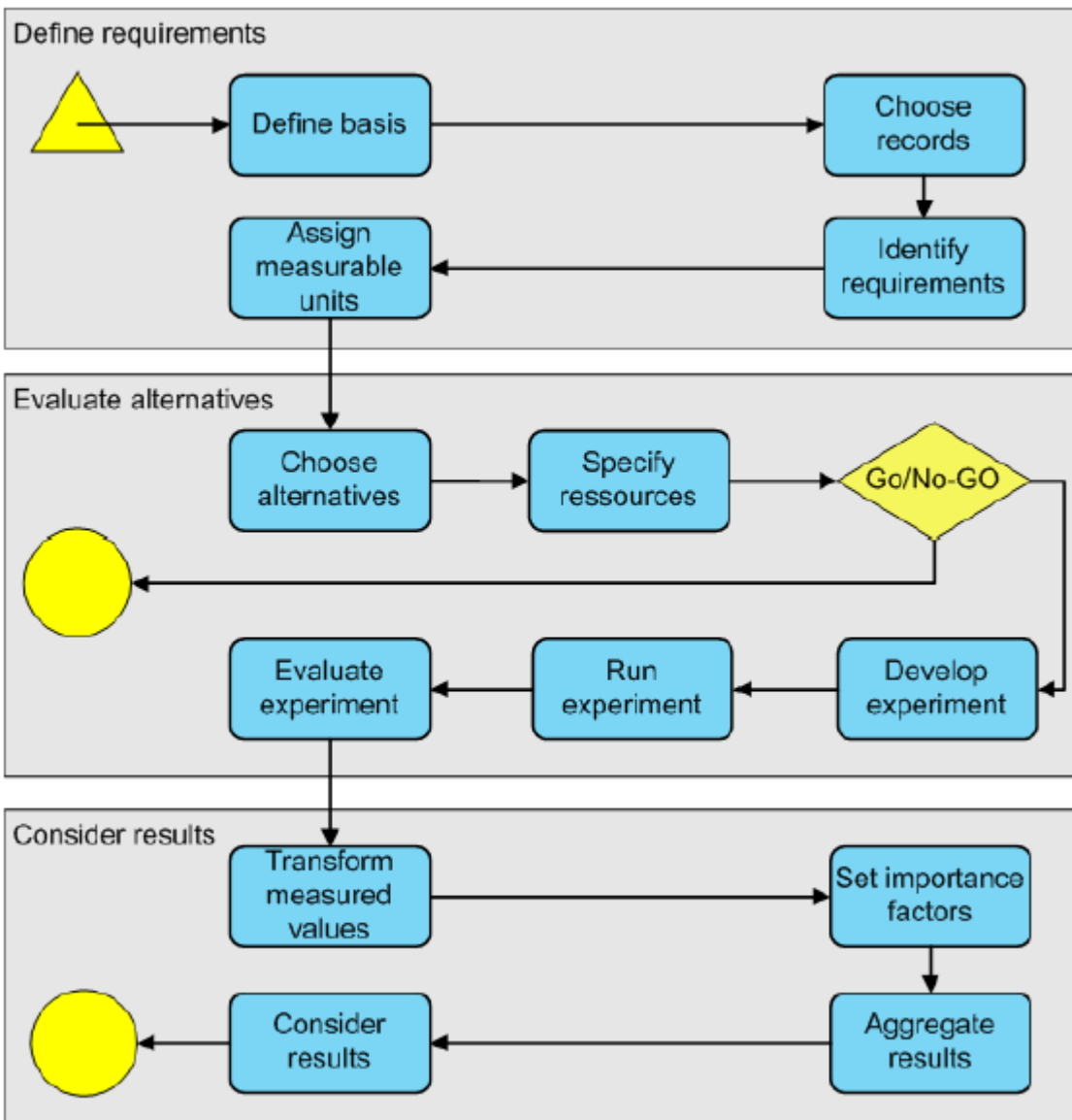
Text·Text·Text·Text·Text·Text·
Text·Text·Text·Text·Text·Text·
Text·~~Text~~·Graphik·Text·Text·Text·
Text·Text·Text·Text·Text·Text·
Text·Text·Text·Text·Text·~~Text~~·
Text·Text·Text·Text·Text·Text·
Text·Text·Text·Text·Text·Text·
Text·Text·Text·Text·Text·Text·
Text·~~Text~~·Tabelle·Text·Text·Text·
Text·Text·Text·Text·Text·Text·
Text·Text·Text·Text·Text·Text·
Text·Text·Text·Text·Text·Text·
Text·Text·Text·Text·Text·Text·
Text·Text·Text·Text·Text·Text·
Text·Text·Text·Text·Text·Text·
Text·Text·Text·Text·Text·Text·
Text·Text·Text·Text·Text·Text·
Text·Text·Text·Text·Text·Text·
Text·Text·Text·Text·Text·Text·

MS-Word

Text Text Text Text Text Text
Text Text Text Text Text Text
Text ~~Text~~ Graphik Text Text Text
Text Text Text Text Text Text
Text Text Text Text Text ~~Text~~
Text Text Text Text Text Text
Text Text Text Text Text Text
Text Text Text Text Text Text
Text ~~Text~~ Tabelle Text Text Text
Text Text Text Text Text Text
Text Text Text Text Text Text
Text Text Text Text Text Text
Text Text Text Text Text Text
Text Text Text Text Text Text
Text Text Text Text Text Text
Text Text Text Text Text Text
Text Text Text Text Text Text

PDF/A





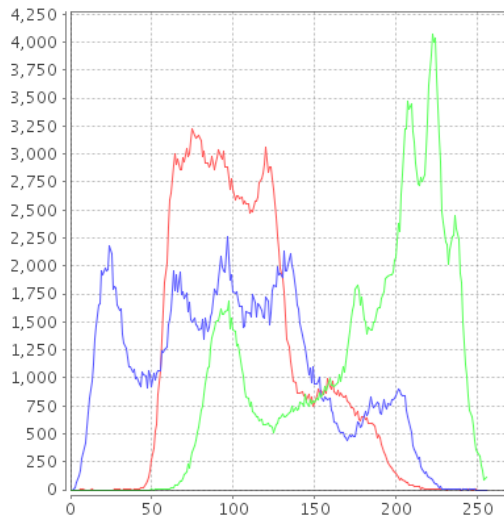
Delos Testbed / PLATO

Comparing Images 1

Image A – differ_webapp6154389266516020945.jpg



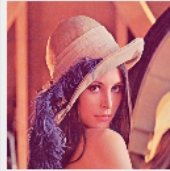
Hash: 6a534138b08aeb48f03b3aabcc25b660



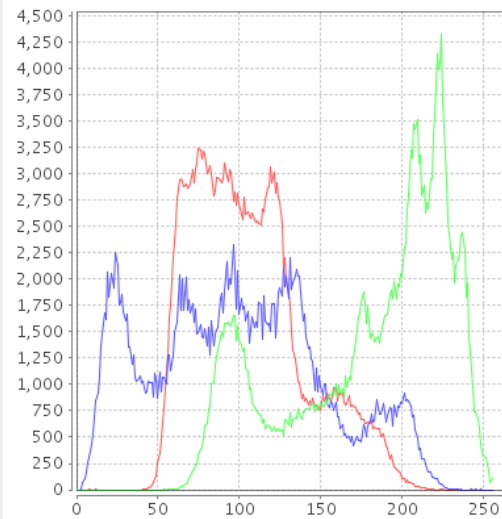
Download as CSV

Setting

Image B – differ_webapp1506169851565439340.png



Hash: 336c37cba1980331ac6279ef7acf4342



Download as CSV

Setting

Image Comparison

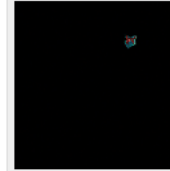
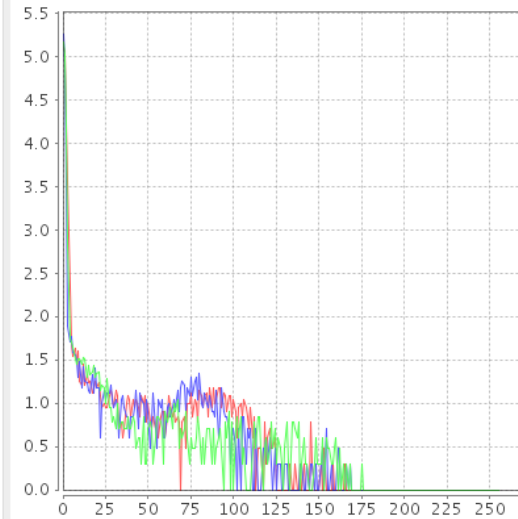


Image hash values are NOT equal.

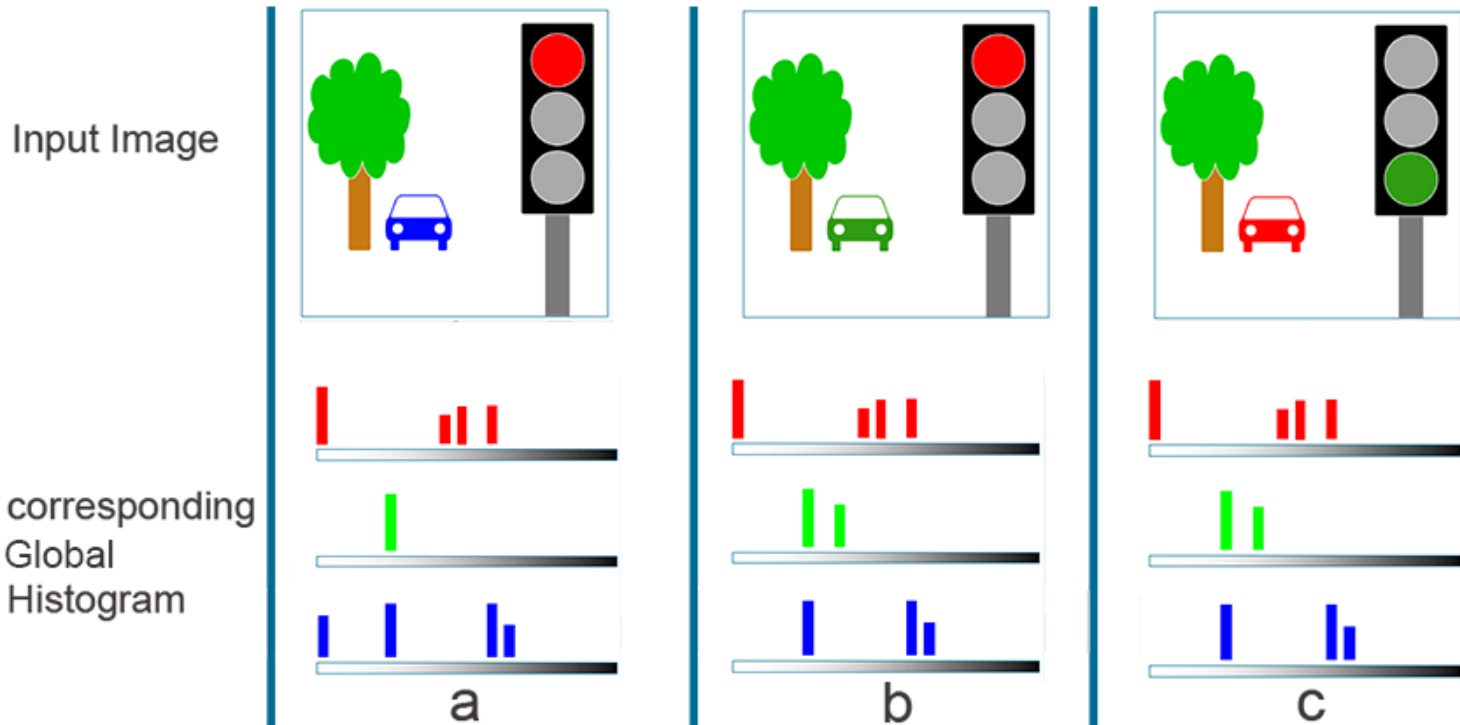


Download as CSV

Setting

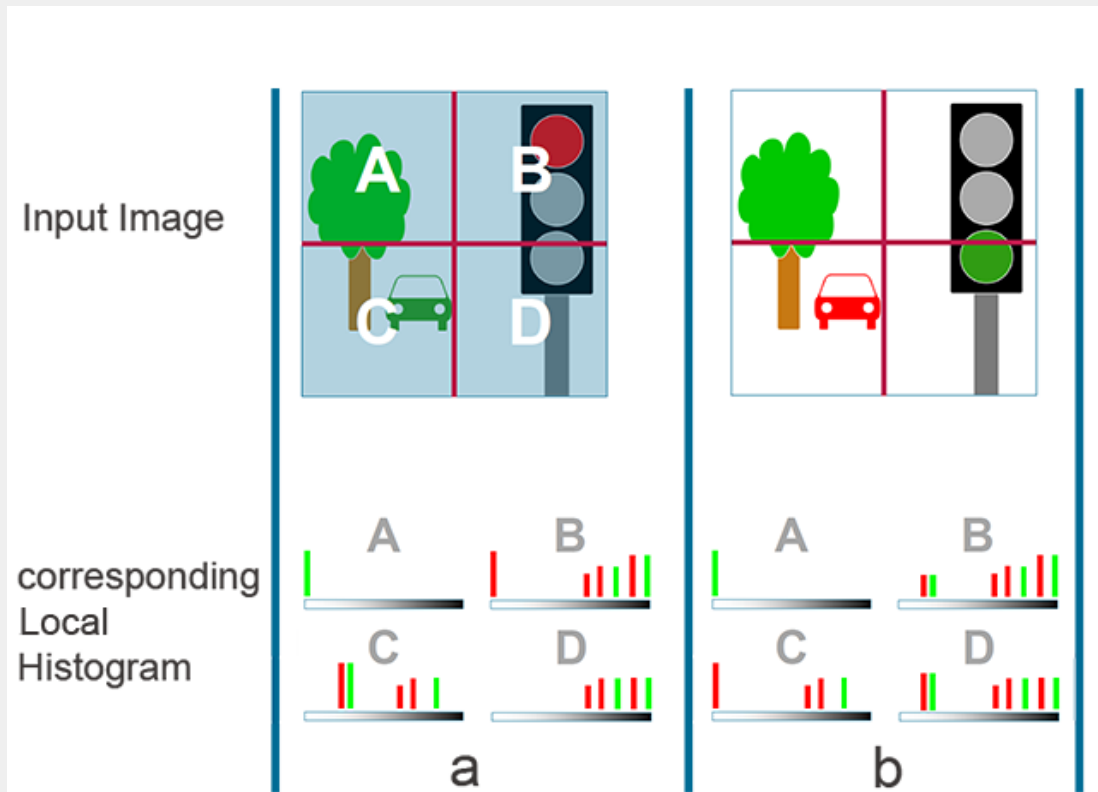
Veronika Krauß, Arefeh Bahrami, Content-based Authentication of images under format migration, 2016

Comparing Images 2



Veronika Krauß, Arefeh Bahrami, Content-based Authentication of images under format migration, 2016

Comparing Images 3



Veronika Krauß, Arefeh Bahrami, Content-based Authentication of images under format migration, 2016

Significant properties: State of the art 1

- **We define significant properties as: those properties of information types that most preservation practitioners consider significant in most contexts.**

R. van Veenendaal, P. C. M. Lucker, C.D. Sijtsma², National Archives of the Netherlands

Significant properties: State of the art 2

...
Database	Content	For the user application: the content of the database displayed onscreen must be preserved. (SQL is used to retrieve the data from the database. The queries that are used must be preserved so as to be able to represent the required content.)
Database	Context	Name (of the organisation producing the database)
Database	Context	Work process (to which the database belongs)
Database	Context	Relationship with other information objects
...

R. van Veenendaal, P. C. M. Lucker, C.D. Sijtsma2,
National Archives of the Netherlands

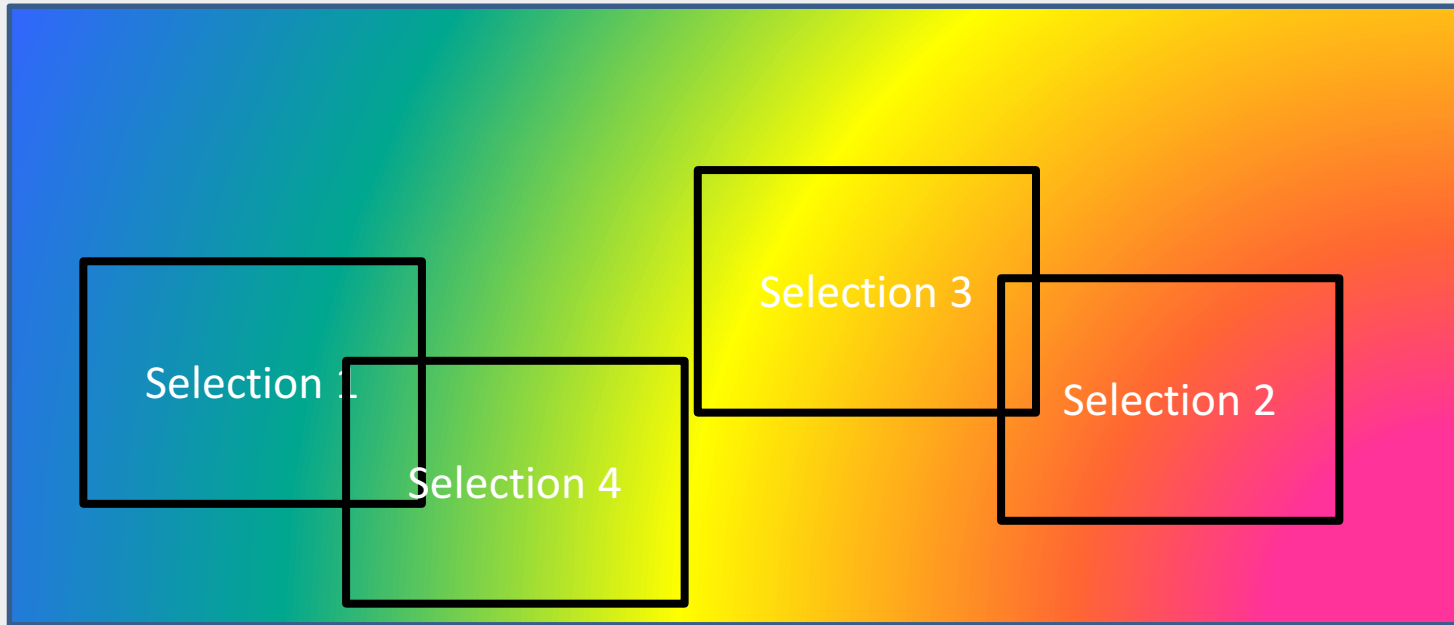


Significant properties: Definitions

- Margaret Hedstrom & Christopher A. Lee: "those properties of digital objects that affect their quality, usability, rendering, and behavior."
- Cedars project (1): "those components of a digital object deemed necessary for its long-term preservation."
- Cedars project (2): "those technical characteristics agreed by the archive ... to be most important for preserving the digital object over time
- InSPECT: "the characteristics of digital objects that must be preserved over time in order to ensure the continued accessibility, usability, and meaning of the objects, and their capacity to be accepted as evidence of what they purport to record."

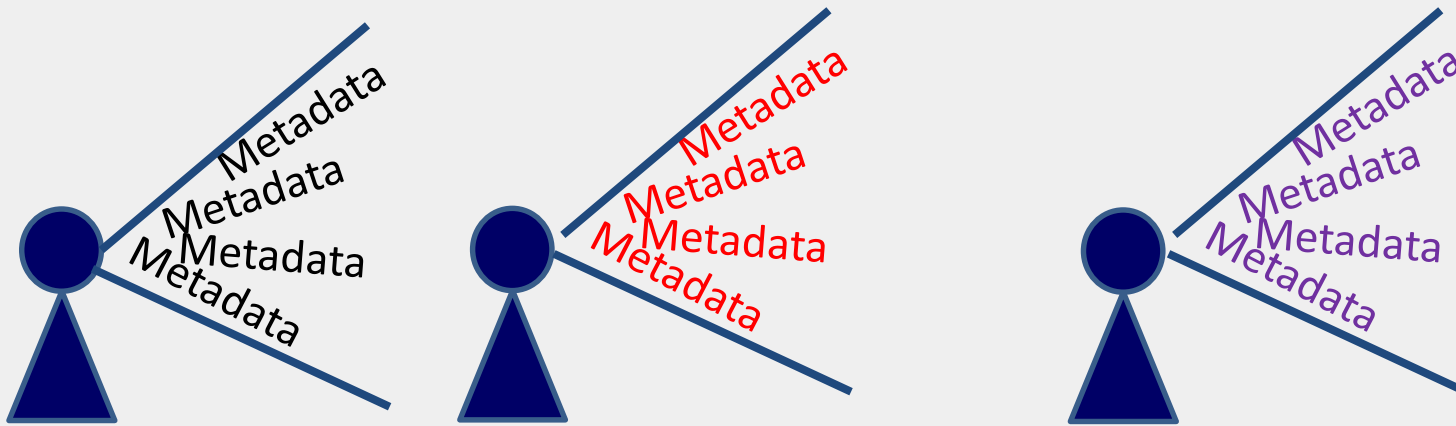


nestor definition

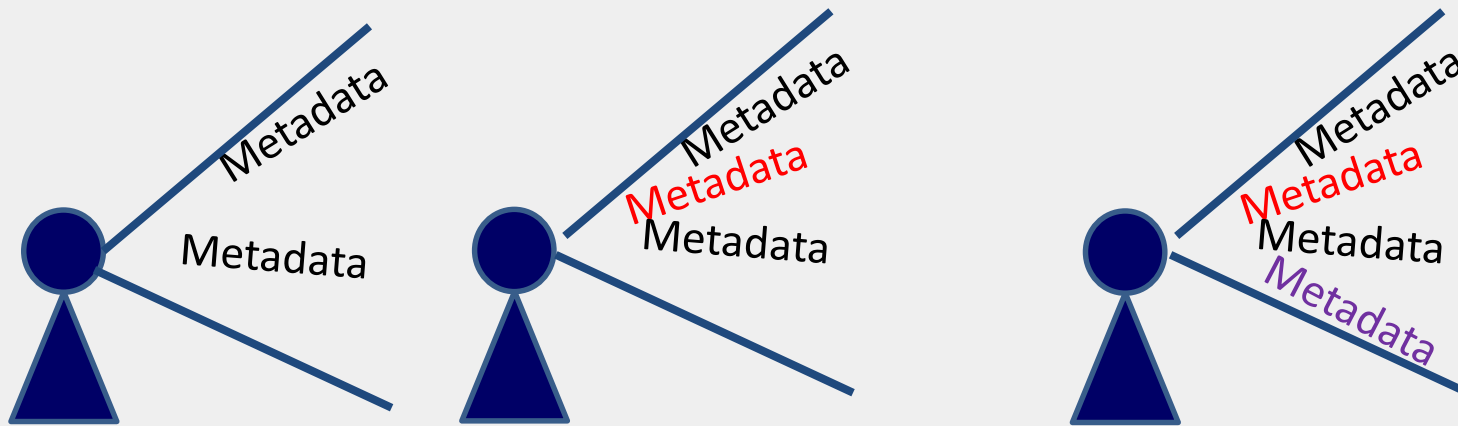


- Significant properties are a subset of all object properties
- The most important properties, which has to be preserved in any case
- The definition of the subset is a subjective assessment

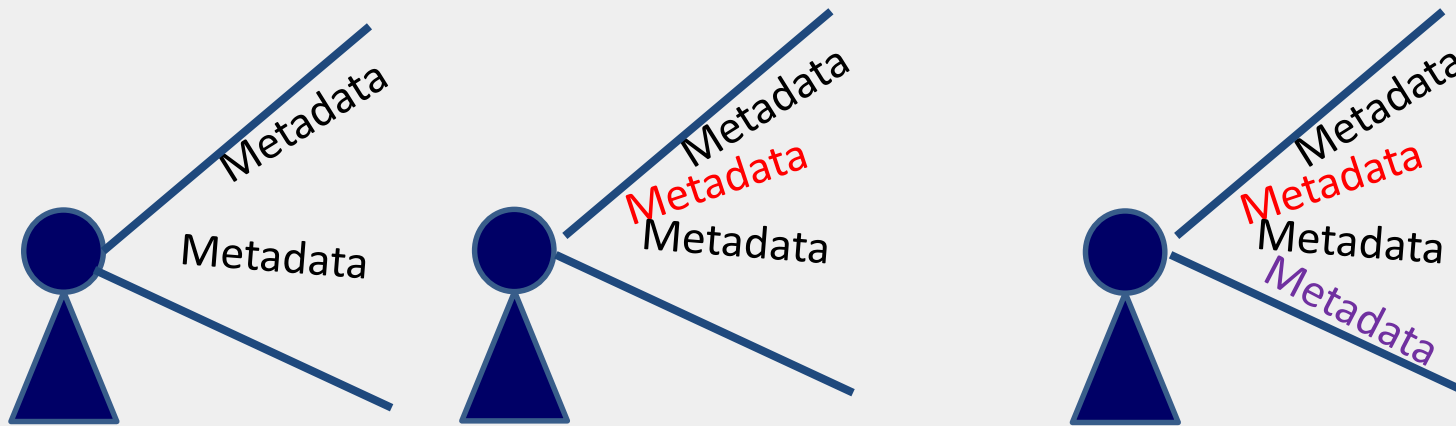
Communication Scenario (archivists) 1

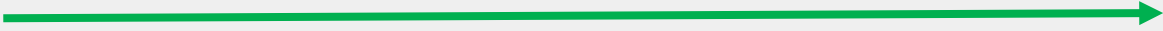
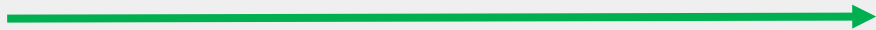
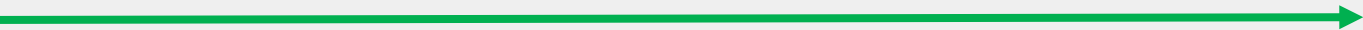


Communication Scenario (archivists) 2



Communication Scenario (archivists) 2



- Persistent Identifiers 
- Assumptions about Knowledge Base 
- Validation 



Proposed assumptions

- Concept

- Performance model



- Significant properties as a benchmark (subset of all properties)

- Target: Ability to recreate similar performances

- Distinct from auxiliary means (data, hard- and software)
- Distinction between must-haves (significant properties) and nice-to haves (all other properties)

Premises

- Subject
 - Information Object (OAIS)
 - Intellectual Entity (PREMIS)
- Means
 - Data, files...
- Human factor
 - Designated community (OAIS)
 - Designated uses (nestor)
- Decisions
 - Means + human factor >>> subject

Related Realms

- Topics
 - Appraisal
 - Access
- Communities
 - Persistent Identifier
 - Audio & Video
- Sciences
 - Archival Science
 - Digital Curation
 - Computer Sciences
- Questions
 - Concepts and terminology
 - Policies
 - Practical doing
 - Programming
 - Costs
 - Lessons Learned

William Kilbride | The Bit List of Digitally Endangered Species

William Kilbride (Digital Preservation Coalition, DPC) stellt in seinem Vortrag die von der DPC erstellte „Global List of Digitally Endangered Species“ (kurz: Bit List; Link: <https://www.dpconline.org/our-work/bit-list>) vor.

Die Bit List gliedert sich in die folgenden fünf Kategorien:

1. Praktisch ausgestorben
2. Kritisch gefährdet
3. Gefährdet
4. Verwundbar
5. Geringes Risiko

Das Projekt, das die DPC von Grund auf konzipiert und durchgeführt hat, erwies sich als Mischung aus Erfolgen, Herausforderungen und Risiken. Sollte es sich etablieren, ist die Veröffentlichung jährlicher Updates geplant. Hinter der Bit List steht eine zwanzigköpfige Jury mit Mitgliedern aus sieben Ländern auf drei Kontinenten. Das Projekt führt laut Kilbride zu drei Kern-Thesen:

1. Wir reden über undeutliche Gefahren für Daten oder Software, aber die Verluste sind echt und betreffen uns alle im Alltagsleben.
2. Diese ist eine soziale und technische Herausforderung. Wenn wir nur technische Lösungen anbieten, haben wir keinen Erfolg.
3. Digitale Bestandserhaltung handelt nicht von einem „dunklen Zeitalter“. Es geht darum, die Versprechungen der digitalen Welt jetzt wirklich einzulösen.

Diskussion und Fragen:

1. Tobias Steinke: Welche spezifischen „digitalen Probleme“ bestehen im Hinblick auf die Bit List?

A: Wichtig sei, die Frage der Verantwortlichkeit für die digitalen Objekte zu klären.

2. Christian Keitel: Maschinenlesbare Inhalte (z.B. Lochkarten) drohen, nicht weiter lesbar sein zu können und sind daher als bedroht anzusehen.
3. Kai Naumann: Es gibt weiterhin die Möglichkeit, die Bit List zu kommentieren und Vorschläge für die Erweiterung zu machen.
4. Martin Wrigley: Ist in allen Fällen eine Provenienz-Bestimmung möglich?

A: Vertrauen und die Beweismittelkette (chain of custody) sind zu beachten. Es stellt sich immer die Frage, woher die Objekte kommen und wer sich darum kümmern

muss. In manchen Fällen gibt es jedoch keine Möglichkeit, die Provenienz einzelner Objekte zu bestimmen.

5. Michael Puchta: Die Datenschutz-Gesetze und ihre Vorschriften finden in der Praxis oft keine Anwendung. Soziokulturelle Gründe bedingen die vorarchivische Löschung.

A: Die Gesellschaft braucht mehr Zuversicht, Selbstvertrauen und Sicherheit im Umgang mit digitalen Inhalten.

ENGLISH

1. Tobias Steinke: What specific "digital problems" exist with regard to the Bit List?

A: It is important to clarify the question of responsibility for the digital objects.

2. Christian Keitel: Machine-readable content (such as punched cards) threatens to be illegible and must therefore be regarded as threatened.

3. Kai Naumann: There is still the possibility to comment on the Bit List and make suggestions for the extension.

4. Martin Wrigley: Is a provenance determination possible in all cases?

A: Trust and the chain of custody are to be considered. It always raises the question of where the objects come from and who has to take care of them. In some cases, however, there is no way to determine the provenance of individual objects.

5. Michael Puchta: Data protection laws and regulations often do not apply in practice. Socio-cultural reasons condition the pre-archival deletion.

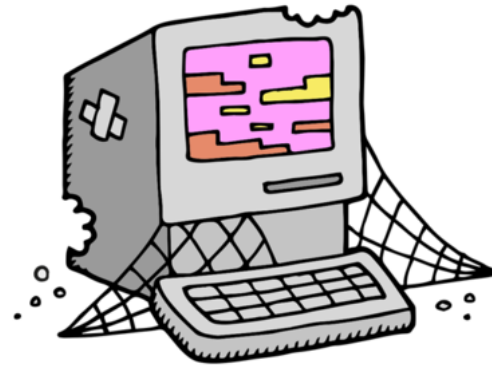
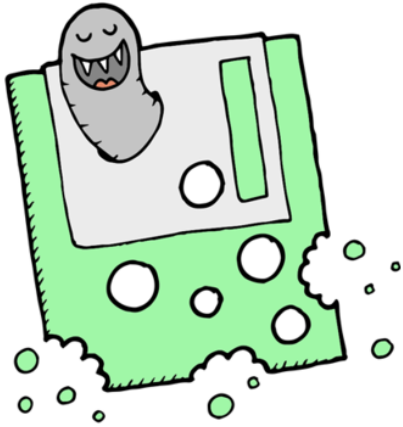
A: Society needs more confidence, self-confidence and security in dealing with digital content.

The BitList:

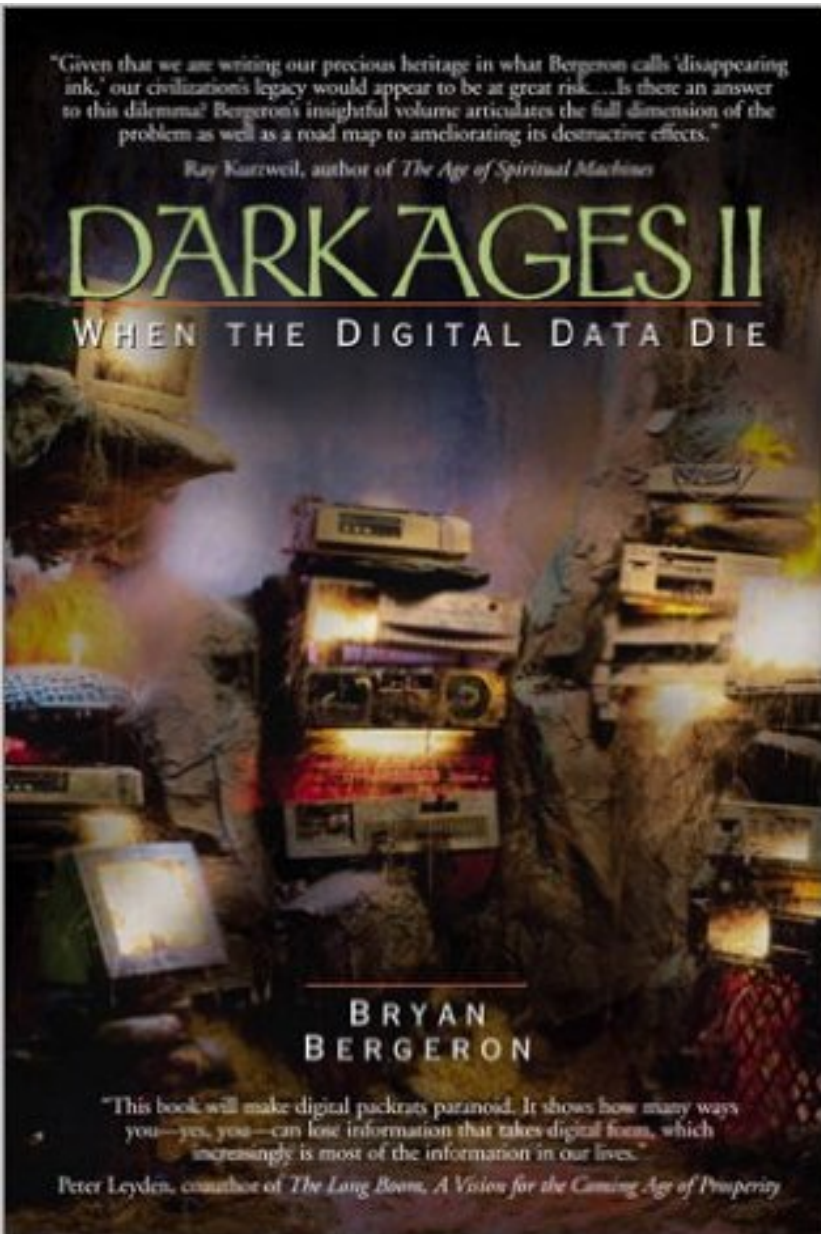
The Global List of Digitally Endangered Species 2018

- Bleak reading
- Agenda Setting?
- The BitList 2017
- Lessons Learned
- The BitList 2018
- Our work

Digital Preservation typically makes bleak reading ...



Your details here:



...very bleak reading: e.g. the Digital Dark Age

Dark Ages II shows why our data is at far greater risk than we've ever imagined—and envisions a frightening future, where so much critical information is lost that civilization itself could collapse.

Tempting?

- *Gets attention*

Realistic?

- *Coming soon: the apocalypse*
- *The disk / tape / cat that never dies*
- *Ignores some really great work...*



REALLY GREAT WORK





REALLY CLEVER PEOPLE





ALMOST 20 YEARS



And still ...



M-DISC | YOUR LIFE. ENGRAVED.™

Home What is M-DISC? How do I use it? Products Store Testimonials FAQ Contact Us

M-DISC Technology News Business Partners

What is M-DISC?

The 1,000 Year M-Disc. Your Life, Engraved In Stone.



M-DISC is an archival-quality storage solution that preserves photos, videos, music and documents for 1,000 years or more. Unlike hard drives, flash drives, and other writable media, that can lose data, M-DISC has been designed to protect your information from degradation and loss for centuries.

FOREVER STORAGE

e's Vint Cerf warns of 'digital Dark

INDY/TECH

'DIGITAL DARK AGE' COULD LEAVE HISTORIANS WITH NO RECORDS OF THE 21ST CENTURY

Resident engineer John Roberts poses as he looks at the millions of newspapers stored on racks at the National Newspaper Archive in Boston Spa. The British Library's brand new National Newspaper Building has officially opened. The newly built storage void



iWonder

Is DNA the best way to store data?

1. Smart molecule
2. CLICKABLE: Perfect data store?
3. Replacing hard drives with DNA
4. Success stories
5. Where next?



theguardian
Winner of the Pulitzer prize 2014

style fashion environment tech travel

 browse all sections

digital dark age

to prevent a



Welttag Digitale Erhaltung

29. November 2018

We want to set the agenda

Three simple initiatives to draw attention to good work and focus on what needs done:

- *World Digital Preservation Day 29th November*
- *Digital Preservation Awards*
- *TheBitList*

In this presentation: TheBitList

The Global List of Digitally Endangered Species

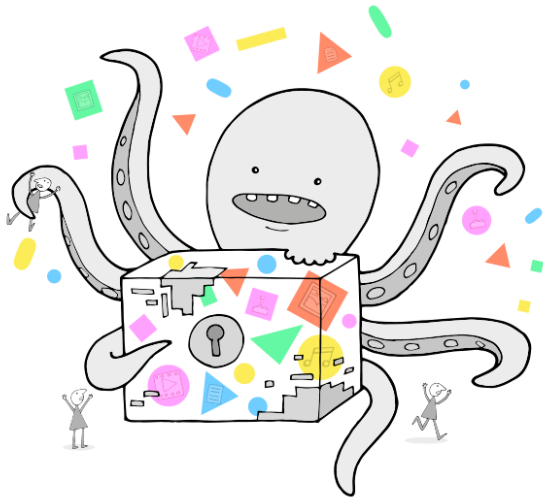
#bitlist - background



- Proposal from DPC advocacy committee
- Modelled on the 'Red List'
- Five categories:
 - practically extinct,
 - critically endangered,
 - endangered,
 - vulnerable,
 - lower risk
- Mix of successes and challenges
- Mix of types, mix of risks
- Created from scratch
- If it works, updated annually
- (invitation!)

The Global List of Digitally Endangered Species 2017

#bitlist - process









- Nominations in September (circa 100 proposals)
- Jury met three times in October (20 members, 3 continents, 7 countries)
- List finalised and reviewed in November
- Published 30th November 2017
- Revision due 29th November 2018

#bitlist - jury



Thomas Bähr - nestor / TIB
Victoria Brown - Scottish Council on Archives
Amber Cushing - University College Dublin
Bradley Daigle - NDSA / AP Trust
Faye Lemay - Libraries and Archives Canada
Charlie McCann - BBC
William Kilbride - DPC
Michelle Lindlar - nestor / TIB
Bo Middleton - University of Leeds
John McDonough - National Archives of Ireland
William Nixon - University of Glasgow
Jaana Pinnick - British Geological Survey
Marcel Ras - NCDD
Barbara Sierman - OPF / KB
Claire Tunstall - Unilever
Guus van Domburg - European Central Bank
Jaye Weatherburn - University of Melbourne
Jane Winters - School of Advanced Studies (Chair)

How the #BitList Works: Categories

	<p>Lower Risk</p> <p>Digital materials are listed as 'Lower Risk' when it does not meet the requirements for other categories but where there is a distinct preservation requirement. Failure or removal of the preservation function would result in re-classification to one of the threatened categories.</p>
	<p>Vulnerable</p> <p>Digital materials are listed as <i>Vulnerable</i> when the technical challenges to preservation are modest but responsibility for care is poorly understood, or where the responsible agencies are not meeting preservation needs. This classification includes <i>Lower Risk</i> materials in the presence of aggravating conditions.</p>
	<p>Endangered</p> <p>Digital materials are listed Endangered when they face material technical challenges to preservation or responsibility for care is poorly understood, or where the responsible agencies are poorly equipped to meet preservation needs. This classification includes <i>Vulnerable</i> materials in the presence of aggravating conditions.</p>
	<p>Critically Endangered</p> <p>Digital materials are listed <i>Critically Endangered</i> when they face material technical challenges to preservation, there are no agencies responsible for them or those agencies are unwilling or unable to meet preservation needs. This classification includes <i>Endangered</i> materials in the presence of aggravating conditions.</p>
	<p>Practically Extinct</p> <p>Digital materials are listed as Practically Extinct when the few known examples are inaccessible by most practical means and methods. This classification includes <i>Critically Endangered</i> materials in the presence of aggravating conditions.</p>
	<p>Of Concern</p> <p>Digital materials are listed as Of Concern when an active member of the digital preservation community has expressed a legitimate concern but the concern has not yet been assessed by the BitList jury. They will be assessed for inclusion in the subsequent year.</p>

#bitlist - details



<http://www.dpconline.org/our-work/bit-list>

#bitlist – Message 1

We talk about obscure challenges to data or software, but loss is real and has consequences for us all, in the real world.

#bitlist – messages 2

This is a socio-technical challenge. If all we offer are technical solutions then we will not succeed.

#bitlist – messages 3

Digital Preservation is no longer about a Digital Dark Age.

It's about coming good on the digital promise.

#bitlist – reflection





Digital materials are listed as *Vulnerable* when the technical challenges to preservation are modest but responsibility for care is poorly understood, or where the responsible agencies are not meeting preservation needs.

This classification includes *Lower Risk* materials in the presence of aggravating conditions.

1 Item 'Vulnerable'

- Materials Posted to Current Web-based Social Media Platforms or Equivalents



Digital materials are listed *Endangered* when they face material technical challenges to preservation or responsibility for care is poorly understood, or where the responsible agencies are poorly equipped to meet preservation needs.

This classification includes *Vulnerable* materials in the presence of aggravating conditions.

8 Items 'Endangered'

- Born Digital Photos and Video Shared on Social Media or Uploaded to Cloud Services
- Corporate Records of Long Duration held on Network Drives, Intranets and Document Systems
- Digital Legal Records and Evidence
- Digital Music Production and Sharing
- Digital Radio Recordings
- Orphaned Digital Works: digital materials where copyright cannot be traced
- Published Research Outputs
- Records of Long Duration from Local Government or Other Government Agencies



Digital materials are listed *Critically Endangered* when they face material technical challenges to preservation, there are no agencies responsible for them or those agencies are unwilling or unable to meet preservation needs.

This classification includes *Endangered* materials in the presence of aggravating conditions.

10 Items ‘Critically Endangered’

- Born Digital Images Held Offline on Portable Storage Devices
- Community Archives and Community Generated Content
- Digital Materials Stored on Magnetic Portable Media
- Digital Materials Stored on Older Portable Media (Non-Magnetic)
- Family or Personal Records
- Gaming
- Media Art
- Politically Sensitive Data
- Smart Phone Apps
- Unpublished Research Outputs



Digital materials are listed as *Practically Extinct* when the few known examples are inaccessible by most practical means and methods.

This classification includes *Critically Endangered* materials in the presence of aggravating conditions.

2 Items 'Practically Extinct'

- Pre-WWW Videotex Data Services and Bulletin Board Services
- Pre-WWW ViewData and Teletext Services



Digital materials are 'Of Concern' when an active member of the digital preservation community has expressed a legitimate concern but the concern has not yet been assessed by the BitList jury.

They will be assessed for inclusion in the subsequent year.

9 Items 'Of Concern'

- PDF
- Websites containing Flash
- Email services
- Data posted to defunct or little-used social media platforms
- GeoMagnetic Data
- Pre-production TV and Movie materials
- Pension, mortgage and insurance records
- Medical records
- Architectural and engineering data

... and nominations are now open

The Digital Preservation Coalition



...a secure digital legacy...

- *Not for profit*
- *Member owned, member run*
- *International, cross-sector, cross-discipline, independent*
- *Vendor, technology neutral*



An Chartlann Náisiúnta
National Archives



Trinity College Library Dublin



EUROPEAN CENTRAL BANK

FIROSYSTEM



BANK OF ENGLAND



University of Salford
MANCHESTER



The Open University
BRITISH LIBRARY



ÁRAINNEACHD EACHDRAIDHEIL ALBA

Historic England

dri
Digital Repository of Ireland
Taislann Dhiigitreach na hÉireann

National Records of Scotland



RESEARCH COUNCILS UK

the national archives

ULCC

NATIONAL LIBRARY of IRELAND

University of Glasgow

UNIVERSITY OF CAMBRIDGE

UKLA



PORTICO

RIBA

ACADEMIC PRESERVATION TRUST

ual: university of the arts london



NDA
Nuclear Decommissioning Authority

ILGC NLW



LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

LSE
THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

UNIVERSITY OF BIRMINGHAM

Chester Beatty LIBRARY

DPN
THE DIGITAL PRESERVATION NETWORK

Unilever

D|C|C

ARA
Archives & Records Association UK & Ireland

The University of Sheffield

UNIVERSITY OF BATH

THE UNIVERSITY of York

UCL

ads
ARCHAEOLOGY | DATA SERVICE



PRONI
Public Record Office of Northern Ireland

Llywodraeth Cymru
Welsh Government

US



Durham University

Lancaster University

مكتبة قطر الوطنية
QATAR NATIONAL LIBRARY

BBC

SCOTTISH COUNCIL ON ARCHIVES



UNIVERSITY OF Hull

ROYAL MUSEUMS GREENWICH

THE UNIVERSITY OF WARWICK

University of Portsmouth



UNITED NATIONS
Mechanism for International Criminal Tribunals

UK Data Archive

PRIFYSGOL ABERYSTWYTH UNIVERSITY

University of Kent

The British Museum

Publications Office of the European Union



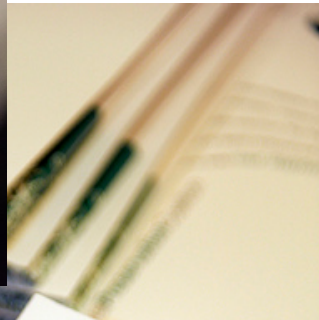
Together we do 6 things ...

- **Advocacy:** raising awareness about digital preservation
- **Community Engagement:** a mutually-supportive, global DP community
- **Workforce Development:** competent and responsive workforces
- **Capacity building:** high quality and sustainable digital preservation
- **Good practice and standards:** making digital preservation achievable
- **Governance:** A stable and trusted platform for collaboration



DPC as a partner and friend!

Join us: <http://www.dpconline.org/join-us>



Tobias Steinke | Organisations- und Rechtsaspekte der Emulation

Tobias Steinke (Deutsche Nationalbibliothek, DNB) eröffnet seinen Beitrag mit einem kurzen Überblick über den gesetzlichen Auftrag an die DNB, gesammelte digitale Publikationen zu erhalten und dauerhaft zur Verfügung zu stellen und geht in der Folge auf den Erhaltungsansatz der Emulation ein.

Hierzu schildert er detailliert das Projekt „Emulation of Multimedia Objects in Libraries“ (EMiL), welches von 2014 bis 2016 als DFG-Projekt zur Entwicklung einer langfristigen Bereitstellungslösung für die multimedialen Bestände von Gedächtnisinstitutionen konzipiert wurde. Projektbearbeiterin war Nathalie Küchler (ehem. Lubetzki). Beteiligt waren hierbei die DNB, die Bayerische Staatsbibliothek, die Staatliche Hochschule für Gestaltung Karlsruhe sowie die Universität Freiburg, die mit bwFLA ein wegweisendes Vorgängerprojekt verantwortet hat. Seit 2017 wird DNB-intern ein Nachfolgeprojekt zur Integration des EMiL-Systems durchgeführt. Der Abschluss des Projekts ist für 2019 geplant.

Das von der Universität Freiburg i.Br. entwickelte EMiL-System dient als Framework für die automatische Bereitstellung in Browsern mit objektbezogen passender Emulation von Medienwerken, die ursprünglich auf optischen Datenträgern und Disketten bereitgestellt wurden. Der Inhalt dieser Datenträger wurde von der DNB ausgelesen und in Disk Images überführt, die im Emulations-Framework aufgerufen werden können.

Im Hinblick auf die rechtlichen Problematiken im Zusammenhang mit EMiL geht Steinke auf die Anzahl, Zugänglichkeit und Plattform der Zugriffsrechner ein. Die Bereitstellung durch die DNB erfolgt ausschließlich über eigene Terminalrechner mit Server-Client-Architektur. Zugang zu den Lesesälen der DNB ist dabei nur für registrierte Benutzerinnen und Benutzer möglich.

Da die erforderlichen Betriebssysteme überwiegend lizenzbeschränkt sind, wurden auf lizenzrechtlicher Ebene mit einzelnen Lizenzinhabern feste Verträge geschlossen. Lizenzanfragen an die Firma Apple wurden dabei schlichtweg nicht beantwortet - egal auf welchem Kanal diese versendet wurden. Beim EMiL-Framework selbst sowie bei den dort eingebetteten Emulatoren handelt es sich um freie LINUX-basierte Software.

Steinke betont, dass gültige Lizenzbedingungen in der Regel nicht die Erfordernisse der Gedächtnisinstitutionen berücksichtigen. So gewährt das deutsche Urheberrecht, im Gegensatz zum US-amerikanischen Urheberrecht, keinen grundsätzlichen Handlungsspielraum für die genehmigungsfreie Nutzung der geschützten Werke.

Zum konkreten Handlungsspielraum deutscher Institutionen im Rahmen des Urheberrechts zeigt Steinke, dass die vorhandenen Schranken nicht ausreichen:

- § 69d Abs. 2 UrhG: Sicherungskopie als Ausnahme von den zustimmungsbedürftigen Handlungen. Die Vorschrift macht keine Aussage über die

Art der künftigen Benutzung und ist daher nicht als Erlaubnis zum Bereitstellen von Software per Emulation anwendbar.

- § 57 UrhG: Unwesentliches Beiwerk. Betriebssysteme fallen nicht unter die anerkannte Definition von unwesentlichem Beiwerk.
- § 60f Abs. 2 UrhG: Recht zur Vervielfältigung im Zuge von Übernahmehandlungen in Archiven. Die Betriebssysteme gehören nicht zum Auftrag der DNB und dürfen daher nicht im Rahmen dieser Schrankenbestimmung in den Bestand der DNB übernommen werden.

Also geht er näher auf die Konsequenzen ein, die sich aus der urheberrechtlichen Situation ergeben. So können Gedächtnisinstitutionen Rechtssicherheit nur durch individuelle Lizenzvereinbarungen erlangen. Zudem besteht, wie im Falle von Apple, immer die Möglichkeit, dass seitens des Lizenzinhabers keine Bereitschaft zur Kooperation gegeben ist. Als dritter und letzter Punkt wird genannt, dass die Divergenz zwischen Urheberrecht und DNB-Gesetz im Zweifelsfall zu Lasten der Erhaltung des Kulturgutes geht.

Herr Steinke schließt seinen Vortrag mit einer Darstellung der Erfahrungen, die die DNB aus den bisherigen Lizenzverhandlungen gewinnen konnte:

1. Die meisten der bislang gestellten Lizenzanfragen waren erfolgreich.
2. Im Falle eines Scheiterns erfolgte selbiges bereits bei der Kontaktaufnahme
3. 75% der von der DNB benötigten Systemumgebungen können in EMiL mittels Einbeziehung freier (Alternativ-)Systeme eingesetzt werden.
4. Der überwiegende Teil der multimedialen Bestände der DNB wird langfristig weiter genutzt werden können.

Diskussion und Fragen:

1. Christian Keitel: Organisatorische und rechtliche Fragen müssen bei der Emulation geklärt werden. Gab es seitens der DNB Überlegungen, dass die sie als Provider für das Freiburger System auftreten würde, damit andere Institutionen die Lizenzen mit-/nachnutzen können?

A: Die rechtlichen Rahmenbedingungen sind ausschließlich auf die DNB zugeschnitten. Die Nutzung für mögliche weitere Partner ist daher nicht möglich. Übergreifende Planungen würden die Situation grundlegend verändern, weshalb keine generische Lösung möglich ist! Ein Auftreten der DNB als Provider für den Zugriff auf einzelne Publikationen erfolgt bislang in kleinem Rahmen für regionale Bibliotheken. Hier ist wichtig, dass alle Teilnehmer die exakt selben Zugriffsrechte haben. Eine Ausweitung auf weitere Institutionen ist unter den genannten Voraussetzungen denkbar.

2. Susanne Paulukat: Wie soll künftig mit Apple Inc. umgegangen werden?

A: Gegenwärtig besteht keinerlei Zugriff auf Apple-Lizenzen, was sich zunächst wohl auch nicht ändern wird.

3. William Kilbride: Die rechtlichen Regelungen stellen hier einen frustrierenden Punkt dar, über den man gemeinsam reden muss. Dieser Dialog muss sowohl national als auch international geführt werden. Hier ist perspektivisch auch der unmittelbare Kontakt mit den produzierenden Institutionen wünschenswert.

A: Das Problem ist international, jedoch kann - wie im Fall von Apple Inc. - meist nur mit den nationalen Dependancen verhandelt werden, wenn überhaupt.

4. Marc Straßenburg: Gibt es Erfahrungen mit notwendigen zusätzlichen Installationen und den damit verbundenen zusätzlichen Lizenzen?

A: Dabei handelt es sich tatsächlich um ein Problem. Für alle lizenzierten Produkte müssen dieselben Schritte unternommen werden.

5. Ulrike Gutzmann: Sind besondere "Digitale Hilfswissenschaften" notwendig, um den Mitarbeiterinnen und Mitarbeitern die Arbeit mit der EMIL-Software zu ermöglichen?

A: Die Software gibt umgebungsbezogenen Hilfestellungen für die Bereitstellung der Inhalte. Eine gesonderte Einarbeitung ist daher nicht zwingend notwendig.

6. Peter Worm: Befinden sich die Archive wegen der gesetzlichen Anbietungspflicht in einer rechtlich besseren Situation als die DNB?

A (von Christoph Schmidt): Anbietungspflicht bezieht sich nur auf die Informationen der anbietungspflichtigen Stellen. Dazu zählen die entsprechenden Lizenzgeber nicht. Es bedarf hier einer Gesetzesinitiative für die Gedächtnisinstitutionen.

7. Marc Straßenburg: Die "Abo-Lösungen" (z.B. von Microsoft) werden in Zukunft aus lizenztechnischer Sicht zu einem sehr großen Problem.

8. Martin Wrigley: Gibt es zeitliche Limitierungen bei der Lizenzvergabe?

A: Es wird versucht, zeitliche Limitierungen zu vermeiden, dies hängt jedoch stark von den einzelnen Firmen ab.

Tobias Steinke, Nathalie Küchler

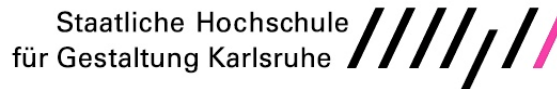
Organisations- und Rechtsaspekte der Emulation

Erhaltungsmaßnahmen der Deutschen Nationalbibliothek

- Gesetzlicher Auftrag zur Erhaltung und dauerhaften Verfügbarmachung der gesammelten Publikationen
- Digitale Publikationen: E-Books, E-Journals, Dissertationen, Webseiten, Multimedia-Publikationen
- Erhaltungsstrategien: Migration, Emulation
- Emulation: Authentische Wiedergabe und Nutzung in emulierter Originalumgebung

EMiL: Emulation of Multimedia Objects in Libraries

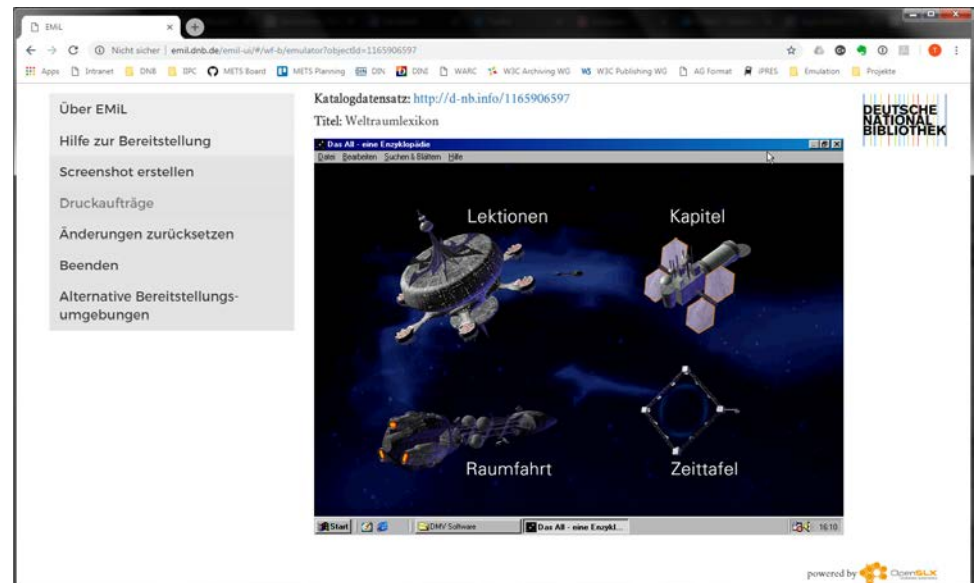
- 2014-2016 DFG-Projekt zur Entwicklung einer langfristigen Bereitstellungslösung für die multimedialen Bestände von Gedächtnisinstitutionen



- 2017-2019 DNB-internes Nachfolgeprojekt zur Integration des EMiL-Systems (inkl. Lizenzprüfung)

EMiL-System

- Entwickelt von der Universität Freiburg, Open SLX
- Framework für die automatisierte Bereitstellung im Browser mit passender Emulation
- Sessionmanagement
- Flexible Umgebungen: Windows, Mac, C64, Amiga, Atari ST, etc.



Einflussfaktoren auf die rechtliche Situation

- Zahl, Zugänglichkeit und Plattform der Zugriffsrechner
 - Die DNB nimmt die Bereitstellung ausschließlich auf ihren eigenen Terminalrechnen vor (Server-Client-Architektur, Windows-basiert)
 - Die DNB-Lesesäle sind nur für registrierte Benutzerinnen und Benutzer zugänglich
- Host-Plattform des Emulationsframeworks
- Bestehende Verträge mit Lizenzinhabern
- (nicht-kommerzielle) Motivation der Systemnutzung

Lizenzrechtliche Rahmensituation

- Sowohl das EMI-L-Framework als auch die eingebetteten Emulatoren sind freie Software (Linux-basiert)
- Die erforderlichen Betriebssysteme sind überwiegend lizenzbeschränkt
 - Lizenzbedingungen berücksichtigen nicht die Erfordernisse in Gedächtnisinstitutionen
 - Im Gegensatz zum US-amerikanischen Urheberrecht (17 U.S. Code § 107 - Fair Use) gewährt das deutsche Urheberrecht keinen generellen Handlungsspielraum für die genehmigungsfreie Nutzung geschützter Werke

Handlungsspielraum des deutschen Urheberrechts

- UrhG § 69d (2) Ausnahmen von den zustimmungsbedürftigen Handlungen
 - Die Erstellung einer Sicherungskopie durch eine Person, die zur Benutzung des Programms berechtigt ist, darf nicht vertraglich untersagt werden, wenn sie für die Sicherung künftiger Benutzung erforderlich ist.

- Die Rechtsvorschrift trifft keine Aussage über zulässige Arten der künftigen Benutzung (z. B. Emulation auf einer anderen als der Ursprungsplattform)

Handlungsspielraum des deutschen Urheberrechts

- UrhG § 57 Unwesentliches Beiwerk
 - Zulässig ist die Vervielfältigung, Verbreitung und öffentliche Wiedergabe von Werken, wenn sie als unwesentliches Beiwerk neben dem eigentlichen Gegenstand der Vervielfältigung, Verbreitung oder öffentlichen Wiedergabe anzusehen sind.

- Betriebssystemen halten der Beiwerk-Schranke nicht stand:
 - Als Beiwerk gilt, ein Werk dem „keine noch so geringfügige inhaltliche Beziehung zum Hauptgegenstand der Verwertung zuzubilligen ist, sondern [das] durch seine Zufälligkeit und Beliebigkeit für diesen ohne jede Bedeutung“ ist (BGH, Urteil vom 17. November 2014, Az. I ZR 177/13).

Handlungsspielraum des deutschen Urheberrechts

- UrhG § 60f (2) Archive, Museen und Bildungseinrichtungen
 - Archive, die auch im öffentlichen Interesse tätig sind, dürfen ein Werk vervielfältigen oder vervielfältigen lassen, um es als Archivgut in ihre Bestände aufzunehmen.

- Die Aufnahme von Bereitstellungssoftware in die DNB-Bestände würde sie dem Erhaltungsauftrag unterordnen. Die DNB besitzt allerdings keinen Sammelauftrag für die Systeme, die zum langfristigen Erhalt ihrer Bestände erforderlich sind.

Konsequenzen der urheberrechtlichen Bedingungen

- Rechtssicherheit erlangen Gedächtnisinstitutionen nur durch individuelle Lizenzvereinbarungen.
 - Die Urheberrechtlichen Schranken lassen sich nicht zweifelsfrei auf den Use Case erhaltungsbedingte Emulation anwenden.
 - Genehmigungsfreiheit geht nicht mit Vergütungsfreiheit einher.
- Die Identifikation der Lizenzinhaber ist möglich, deren Bereitschaft zu Lizenzverhandlungen aber ungewiss.
- Das Dilemma zwischen Urheberrecht und DNB-Gesetz geht im Zweifelsfall zulasten der Erhaltung von Kulturgut.

Ergebnis der Lizenzprüfung und -verhandlung

Lizenzinhaber	System	Lizenzierung
Microsoft Corp.	MS DOS 6.2	Alternativsystem FreeDOS
	Windows 3.11 – Windows 7	Vertraglich definiertes Downgrade-Recht
Apple Inc.	Apple II DOS, System 7	Lizenzbedingungen schließen ein Kopieren des ROM aus
	Mac OS X	Lizenzbedingungen untersagen den Betrieb auf Nicht-Apple-Plattformen
Linux	Suse, Redhat	Freie Software
Cloanto Corp.	C64, Amiga	Individuelle Lizenzvereinbarung
Atari SA	Atari ST TOS	Alternativsystem EmuTOS

Erfahrungen aus den Lizenzverhandlungen der DNB

- Die meisten der von DNB geführten Lizenzanfragen waren erfolgreich.
- Im Negativfall scheitern die Lizenzverhandlungen bereits bei der Kontaktaufnahme.
- Unter Einbeziehung freier (Alternativ-)Betriebssysteme können 75% der von der DNB benötigten Systemumgebungen in EMiL eingesetzt werden.
- Der weitaus überwiegende Teil der multimedialen DNB-Bestände wird langfristig nutzbar bleiben.

Martin Wrigley | The Preservation Action Registry

Martin Wrigley (Open Preservation Foundation, OPF) stellt in seinem Vortrag die OPF und das „Preservation Action Registry“ (PAR) vor. Bei der OPF handelt es sich um eine globale Non-Profit-Vereinigung, die sich für die Förderung von Open-Source-Tools für die Digitale Bestandserhaltungs-Community einsetzt. Sie wurde 2010 gegründet, um die Ergebnisse des PLANETS-Projekts zu erhalten. Von der OPF betreut werden inzwischen unter anderem veraPDF, jpylyzer, fido und JHOVE. Das Ziel der OPF ist es, durch ein Referenz-Toolset und dessen Erweiterungen gemeinsame Lösungen für eine effiziente und effektive digitale Bestandserhaltung zu ermöglichen.

In der Folge schildert er das Problem, das PAR lösen soll: Es gibt gegenwärtig zwar viele „Best Practice“-Lösungen und -Ansätze aus verschiedenen Sparten, diese stehen effektiv jedoch nicht miteinander im Austausch. Das PAR soll einen Mechanismus bieten, der den systemunabhängigen Austausch praktikabler Informationen zwischen Dienstleistern und Institutionen ermöglicht und unterstützt.

Das PAR entstand aus dem Research Data Shared Service Projekt der JISC (einritisches Pendant zur DFG), welches zur Diskussion der Interoperabilität von Format-Grundsätzen zwischen unterschiedlichen Bestandserhaltungssystemen (z.B. Preservica und Archivemata) geführt hat.

Als Kernaussagen seines Vortrags lassen sich folgende Punkte festhalten:

1. Die zuständigen Kolleginnen und Kollegen sollten nicht alle selbst programmieren können müssen. Wichtig ist, dass die vorliegenden Tools, geschaffen von der OPF oder anderen, so verbunden werden, dass auch Nicht-Techniker damit klarkommen.
2. Erhaltung ist nicht nur die Erhaltung der Dateiformate, sondern die Erhaltung des Sinns der Dateien.
3. Formatregister sind untereinander nicht kompatibel.
4. PAR schlägt ein einheitliches Format für Formatprüfung und Erhaltungsaktionen an Dateien vor. PAR soll kein akademisches Projekt werden, sondern ein praktisches!

Abschließend stellt Wrigley die geplanten nächsten Schritte der OPF vor. Sie will Projekte und Zwischenziele genauer definieren, Anwendungsfälle mit gut sichtbarem Nutzen bekannt machen, mehr Organisationen einbinden und das Konzeptmodell PAR dort einbringen.

Diskussion und Fragen:

1. Christian Keitel: Gibt es Spannungen zwischen kommerziellen und nichtkommerziellen Stellen?

A: Kommerzielle Beteiligte benutzen die OPF-Tools und liefern Ideen zur Verbesserung. Zum Beispiel: Fido (Open Source) wird von Artefactual gepflegt. Gegenseitiger Nutzen ist in beiden Richtungen vorhanden. Konkurrenz zwischen Open Source und

Kommerziellen Herstellern ist vorhanden, wobei Open Source (z.B. veraPDF) die Basis ist und kommerzielle Hersteller Mehrwerte über die Basis hinaus liefern (mehr Volumen, spezielle Nischen).

2. Christian Keitel: Im Mittelpunkt stehen Systeme wie Preservica oder Archivemata, während die OPF Tools deren gemeinsam genutzte Peripherie bilden. Plant die OPF, ein Open-Source-Gesamtsystem als Konkurrenz zu den genannten anzubieten?

Martin Wrigley hat dies verneint.

3. Michelle Lindlar: Es wäre schön, wenn alle Hersteller großer Gesamtsysteme in der OPF Mitglied werden.
4. Michael Puchta: Wo ist die Grenze zwischen Gesamtsystem und Zusammenspiel der Einzelmodule?

A: Es hat sich immer wieder als schwierig und misslich dargestellt, anderen Organisationen Standards aufzwingen zu wollen. Daher ist es zielführender, wenn nur die Schnittstellen zwischen den einzelnen Modulen zur Nachnutzung offen stehen.

5. Lambert Kansy: Wie werden die Regeln und deren Umsetzung festgehalten?

A: Das gehört zu den Aufgaben, um die sich in naher Zukunft gekümmert werden muss.



Open Preservation Foundation and The Preservation Action Registry

Martin Wrigley, Executive Director, OPF

Martin Wrigley



30+ years experience delivering software and solutions - mostly in Mobile Telecoms

10+ years experience of managing a membership driven open source association

OPF Executive Director since September 2017

Expanding my knowledge of the finer points of Digital Preservation

Who is OPF?

- A not for profit, global membership association providing stewardship of open-source tools for the digital preservation community.
- Founded in 2010 to sustain the results of the EU PLANETS project
- The OPF reference toolset now includes veraPDF, JHOVE and more

What is OPF's purpose?

OPF Vision

Open sustainable digital preservation

OPF Mission

Enabling shared solutions for effective and efficient digital preservation; the Open Preservation Foundation leads a collaborative effort to create, maintain and develop the reference set of sustainable, open source digital preservation tools.

This set of tools (including software and standards) enables organisations to evaluate, validate, document, mitigate risk, and process digital content to be preserved in line with desired policies and community best practice.

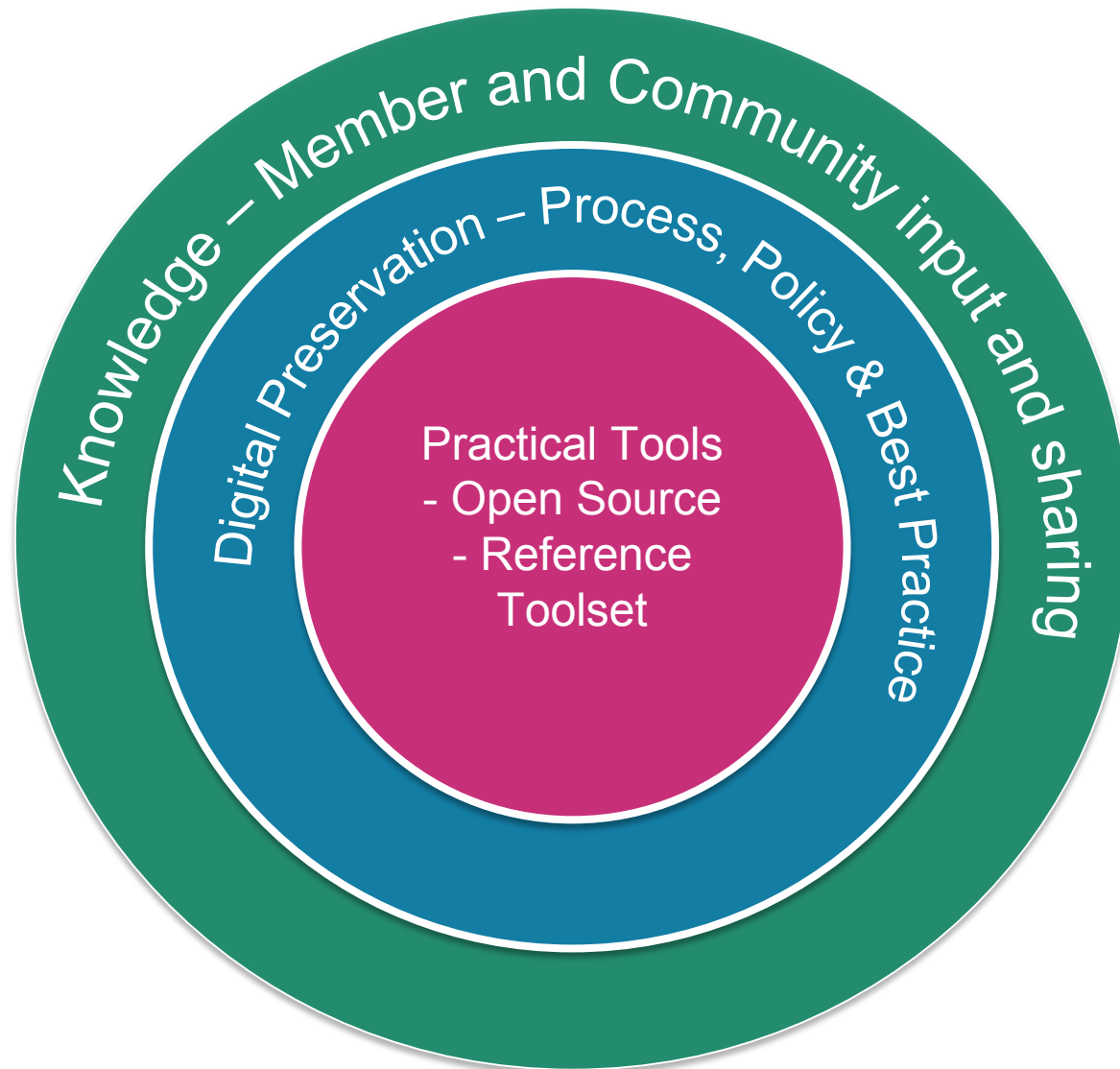
Values

- Open
- Member driven
- Collaborative & Inclusive
- Innovative

What does OPF do?

- Community Knowledge
- Sharing knowledge
- Develop the OPF reference toolset
- Deliver to development roadmaps
- Community engagement
- Webinars and training
- Interest Groups and Tech Clinics
- OPF Software Maturity Model
- Hosting community services e.g. COPTR
- Website, blogs, events

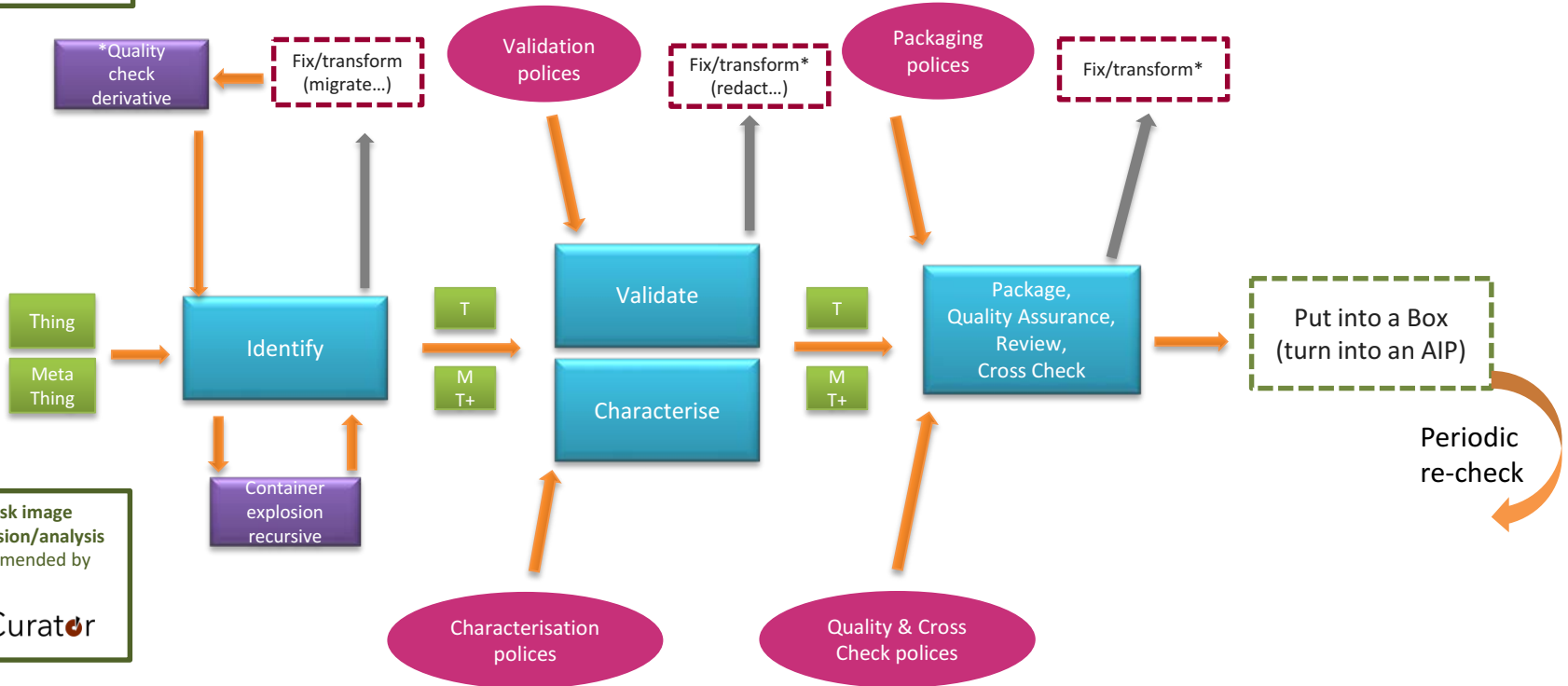
OPF – Digital Preservation Knowledge and Tools



OPF Tool Mapping

Transform
Database archiving /
Extraction tools
Recommended by OPF
**SIARD (SQL database to
XML format)**

Derivative check tools
Maintained through OPF
xcorr sound WAV, MP3



Disk image
explosion/analysis
Recommended by
OPF
BitCurator

Identification tools
Maintained through OPF
fido
Format Sniff
Recommended by OPF
**DROID
PRONOM
FILE**

Validation and Characterisation
tools
Maintained through OPF
veraPDF PDF/A
JHove PDF, JPEG, WAV,
PNG, WARC, AIFF,
UTF8 TEXT, XML,
HTML, GZIP, ASCII
TEXT, MP3, GIF,
JPEG2000
**(DPF Manager)
TIFF module** TIFF
jpglyzer JPEG2000

Information Packaging
tools
TBA

Cross Check tools
TBA

Quality check tools
E-ARK CEF SIP validator

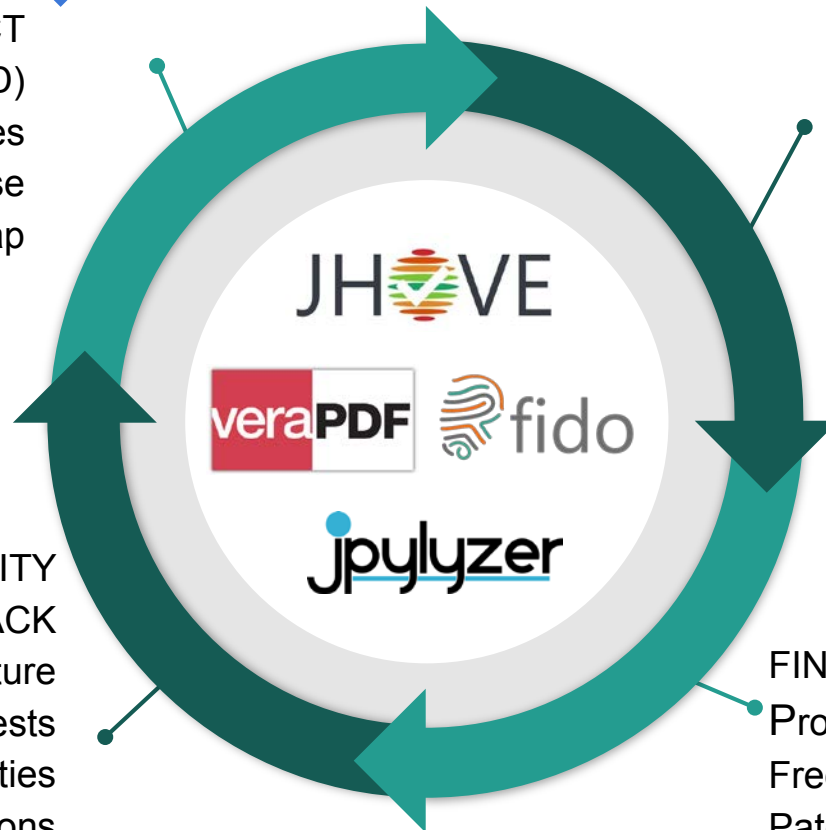
How do OPF projects work?

FUNDING
OPF membership
Donations
Project income



PLANNING (PRODUCT BOARD)
Prioritise fixes and features
Define the release
Manage the roadmap

DEVELOPMENT & TESTING
GitHub for OS development
Build a set of test data
Continuous integration
Quality Assurance



FINAL TEST & RELEASE
Production release
Freely available to community
Patches (essential fixes)

REQUIREMENTS & COMMUNITY FEEDBACK
Bug reports and new feature requests
Hack day activities
Code contributions

Input from OPF interest groups
Contribution of test files



Preservation Action Registry

PAR Background: The problem

- Users want the best advice, wherever it comes from
 - Identification, property extraction, validation, migration, rendering, tools
 - Many sources for current ‘best practice’
 - Products such as Preservica & Archivematica
 - Practitioners
 - Academics
 - Specialists
- but they don't talk to each other effectively

Background: Motivation and Objectives

- To provide a mechanism to exchange good practice information between organisations and preservation system suppliers regardless of which system they use.
- Explicitly: To provide compatibility/ interoperability between JISC RDSS project systems.

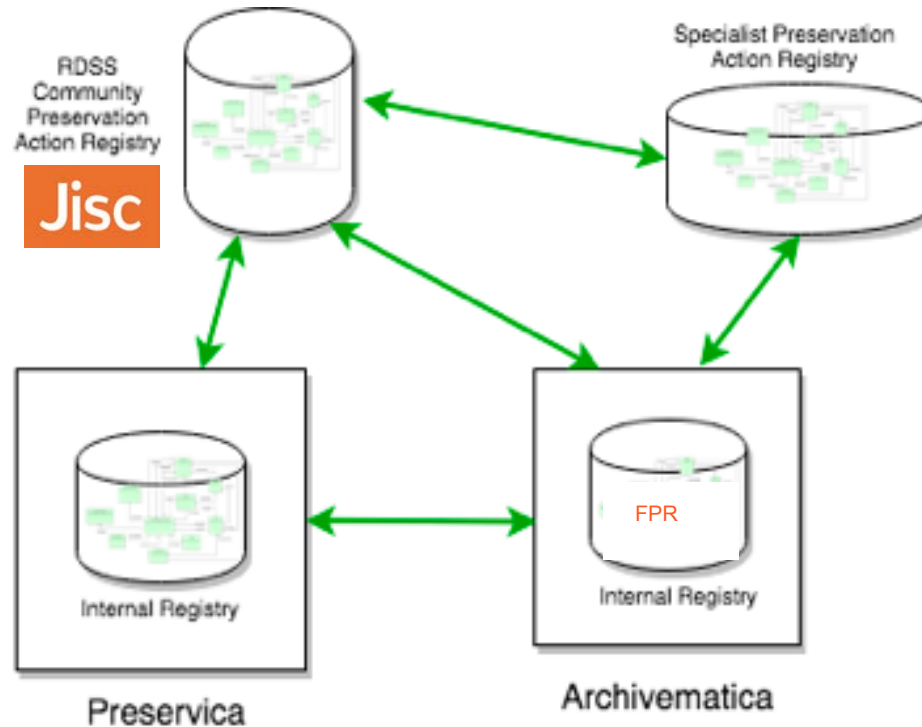
However:

It is not a single 'Best Practice'

It is not 'one registry to rule them all'

Background: Jisc RDSS Project

Development of a multi-vendor shared services platform led to discussions of interoperability of format policies (i.e. “preservation actions”) between preservation systems.



Background: Project Conception

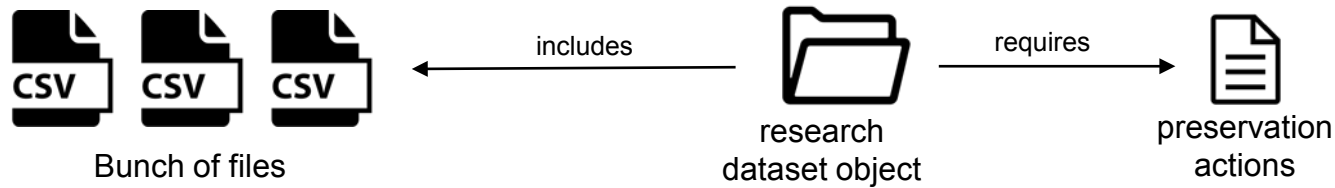
A JISC funded project to initiate the process to deliver benefits to RDSS users

Arkivum, Preservica and Artefactual as RDSS product suppliers

Open Preservation Foundation as respected independent shared DP technology supplier

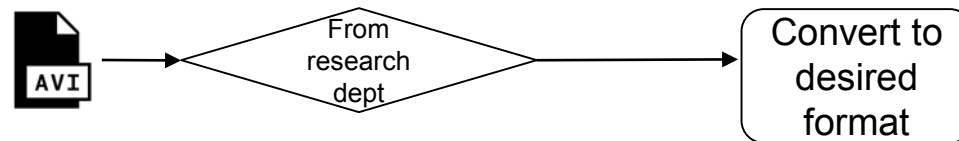
Digital Preservation Actions

Preservation is not just about file formats, it's about making sense of data



The specific action depends on the context, and the policies.

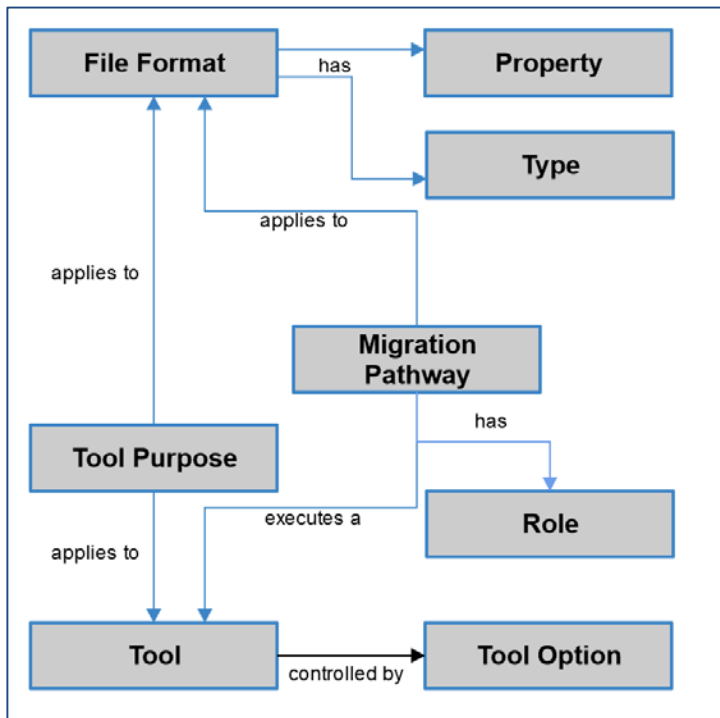
– what action is being taken and why? What is the business rule?



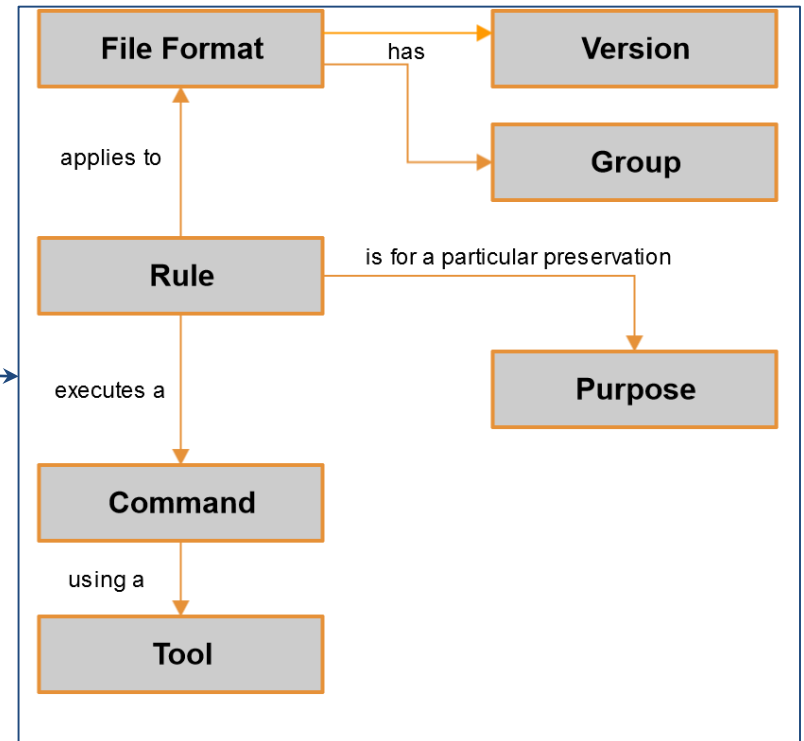
Today - preservation actions are not portable across systems
(e.g. Archivematica, Preservica, others)

Current Registry (In)compatibility

Preservica Registry

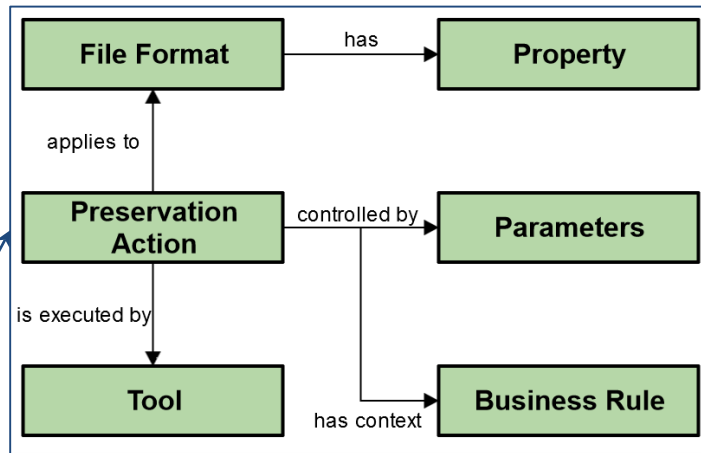


Archivematica FPR



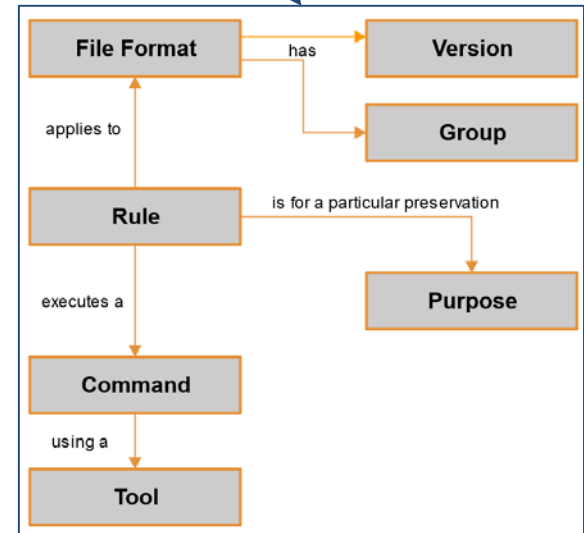
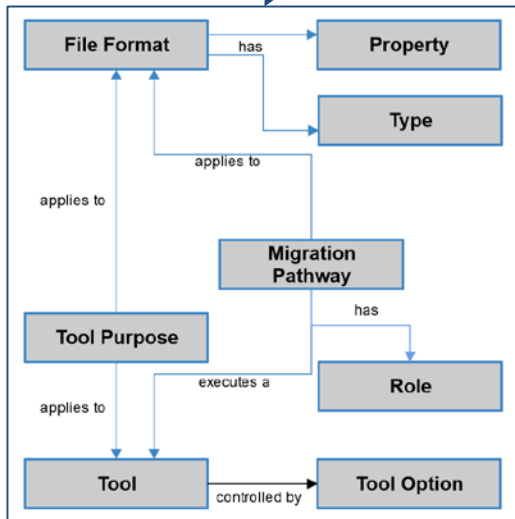
?

Common Language



?

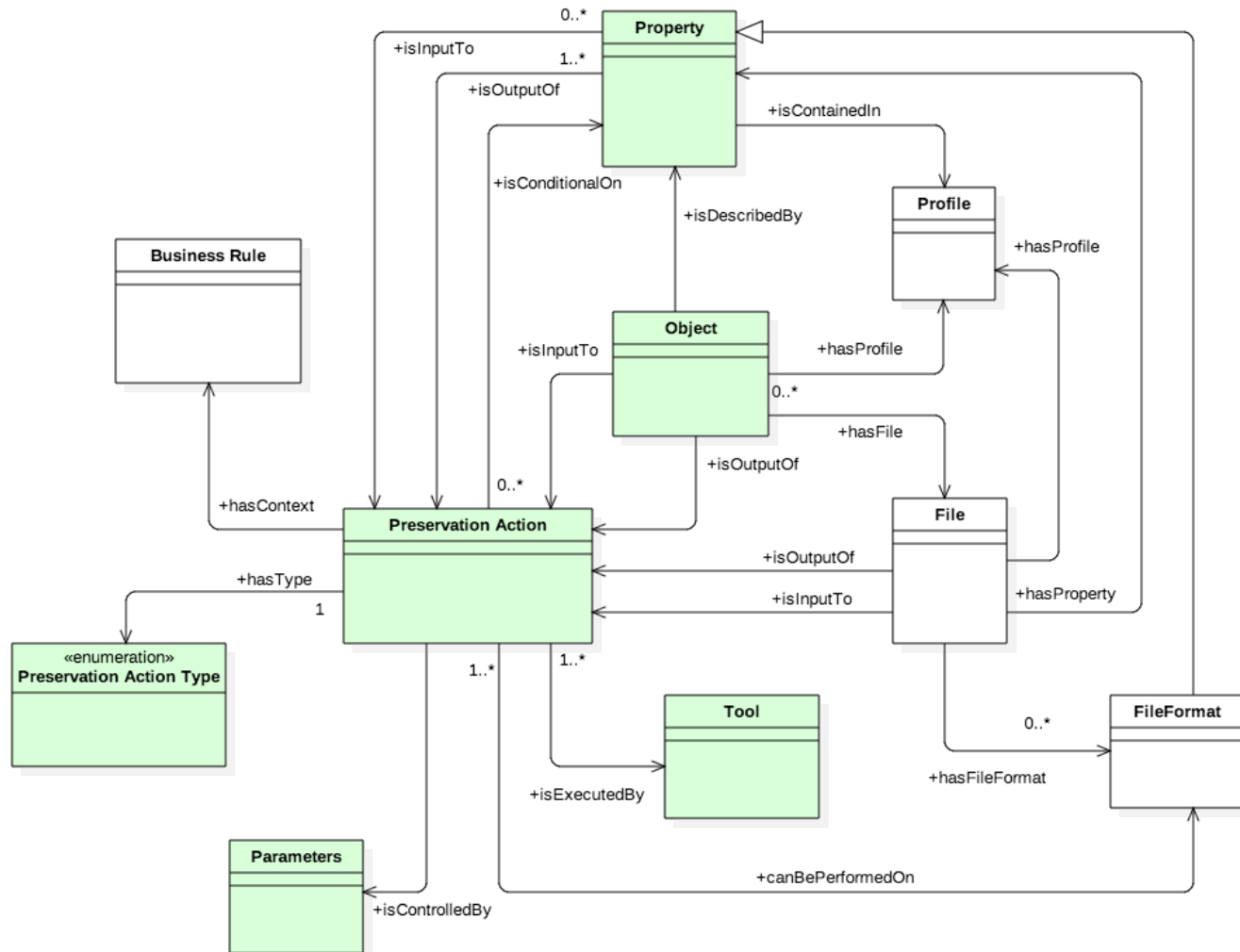
?



What have we produced and why?

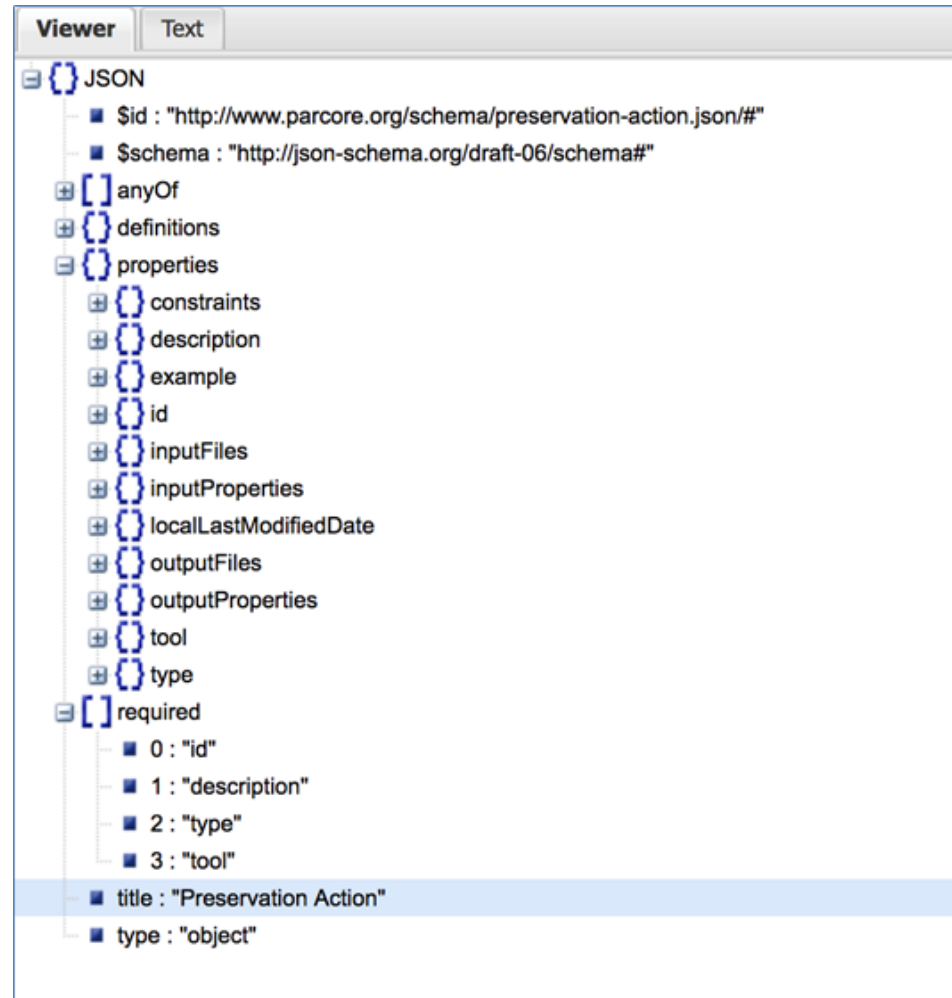
Conceptual Model	<ul style="list-style-type: none">● Common framework for everyone● Language between preservation systems● Still under definition...
Json Schemas	<ul style="list-style-type: none">● Formal definition of the conceptual model● Machine readable, used in API payloads● Used to test and validate interoperability
API	<ul style="list-style-type: none">● Common interface for preservation systems● Well defined way to exchange information
Executable Digital Preservation Actions	<ul style="list-style-type: none">● Cross-platform way to deploy/run tools● Unambiguous and vendor independent
Proof of Concept	<ul style="list-style-type: none">● Reference implementation to share● Make the idea really work between Preservica and Archivematica

PAR Conceptual Model



JSON schemas

- Tool
- Action
- Action Type
- Format
- Property
- Business Rule



The screenshot shows a 'Viewer' window with a 'Text' tab. The main content is a tree view of a JSON Schema. The root node is 'JSON', which is expanded to show its properties. The 'required' array is expanded to show the required properties: 'id', 'description', 'type', and 'tool'. The 'title' property is highlighted in blue.

```
Viewer | Text
├── JSON
│   ├── $id : "http://www.parcore.org/schema/preservation-action.json/#"
│   ├── $schema : "http://json-schema.org/draft-06/schema#"
│   ├── anyOf
│   ├── definitions
│   ├── properties
│   │   ├── constraints
│   │   ├── description
│   │   ├── example
│   │   ├── id
│   │   ├── inputFiles
│   │   ├── inputProperties
│   │   ├── localLastModifiedDate
│   │   ├── outputFiles
│   │   ├── outputProperties
│   │   ├── tool
│   │   └── type
│   ├── required
│   │   ├── 0 : "id"
│   │   ├── 1 : "description"
│   │   ├── 2 : "type"
│   │   └── 3 : "tool"
│   └── title : "Preservation Action"
└── type : "object"
```

APIs

The screenshot shows a web browser displaying the REST API documentation for 'Preservation Actions'. The URL is <https://52.209.71.78/#tag/Preservation-Actions>. The page title is 'Preservation Actions'. The main heading is 'Retrieve all preservation actions'. Below this, there is a description: 'Allow to retrieve the details of all the preservation action'. The 'QUERY PARAMETERS' section lists four parameters: 'limit' (string, integer), 'offset' (string, integer), 'modifiedAfter' (string, datetime), and 'modifiedBefore' (string, datetime). The 'HEADER PARAMETERS' section lists one parameter: 'tool' (string). On the right side, there is a 'Response samples' section showing a sample response for a 200 status code. The response is in 'application/json' format and contains a list of preservation actions with fields like 'constraints', 'description', 'example', 'id', 'guid', 'name', 'namespace', 'inputFiles', 'inputProperties', 'localLastModifiedDate', 'outputFiles', and 'outputProperties'.

Search...

AUTHENTICATION

PRESERVATION ACTION REGISTRY

BUSINESS RULES >

FILE FORMATS >

PAR PROPERTIES >

PRESERVATION ACTION TYPES >

PRESERVATION ACTIONS ▾

- GET Retrieve all preservation actions
- POST Create a new preservation action
- DELETE Delete a preservation action
- GET Retrieve a preservation action
- PUT Update an existing preservation action

TOOLS >

Documentation Powered by ReDoc

Preservation Actions

Retrieve all preservation actions

Allow to retrieve the details of all the preservation action

QUERY PARAMETERS

limit	string	An integer to limit the number of preservation action returned. Default value is zero, which will no filter the result.
offset	string	An integer to specify the offset of the first element of the list of preservation action returned. Default value is zero, which will no filter the result.
modifiedAfter	string	Filter the preservation action to return only the ones having a localLastModifiedDate value GREATER than the one passed in. Allowed datetime formats are YYYY-MM-DD\T\hh:mm:ssTZD and YYYY-MM-DD. In the second scenario, when the time is not provided, it will be assumed to be 00\00\00 UTC time.
modifiedBefore	string	Filter the preservation action to return only the ones having a localLastModifiedDate value LOWER than the one passed in. Allowed datetime formats are YYYY-MM-DD\T\hh:mm:ssTZD and YYYY-MM-DD. In the second scenario, when the time is not provided, it will be assumed to be 00\00\00 UTC time.

HEADER PARAMETERS

tool	string	Filter the preservation actions by the Tool they use, provided its ID. Multiple IDs can be passed in within the same string, linked together using an ampersand symbol as follow: ID_1&ID_2&ID_3. When more than one ID is provided, the endpoint will return all the preservation actions that use one of the IDs in the list. When used together with the preservation_action_type header parameter, the endpoint will return only those preservation actions satisfying both filters' criteria. A list of Tool IDs cannot be combined with another list of Preservation Action Type IDs.
------	--------	---

GET /preservation-actions

Response samples

200

application/json

```
{
  - "constraints": [
    + { _ }
  ],
  "description": "string",
  "example": "string",
  - "id": {
    "guid": "string",
    "name": "string",
    "namespace": "string"
  },
  - "inputFiles": [
    + { _ }
  ],
  - "inputProperties": [
    + { _ }
  ],
  "localLastModifiedDate": "2018-07-31T16:09:25Z",
  - "outputFiles": [
    + { _ }
  ],
  - "outputProperties": [
    + { _ }
  ],
}
```

<https://github.com/JiscRDSS/rdss-par/tree/master/api>

Executable Tool Definitions

- Machine readable spec for running a tool
 - Tool command line
 - Parameters and flags
 - Inputs and outputs
 - Pre and post processing



```
[job mediaInfo2.cwl] completed success
{
  "width": "1280",
  "bitrate": "748253",
  "height": "720"
}
Final process status is success
```

Property extraction

Fixity check

```
[job md5check2.cwl] completed success
{
  "fixity_report": "PASS"
}
Final process status is success
```

Next steps

- OPF coordination
 - Define project deliverables and stages in more detail
- More use cases demonstrating real benefits
- Looking for more organisations to be involved
- Extend the conceptual model to more practical cases that involve more organisations

Make PAR useful to communicate good practice between systems and organisations

Join OPF today!

For more information get in touch...



martin.wrigley@openpreservation.org



<http://openpreservation.org/>



<https://github.com/openpreserve>



[@openpreserve](#)

Newsletter: www.openpreservation.org/subscribe/

For more info on PAR go to

www.openpreservation.org/about/projects/par

Michelle Lindlar | Bestandserhaltung mit Ex Libris Rosetta aus der Perspektive einer großen Forschungsbibliothek

Michelle Lindlar (Leibniz-Informationszentrum – Technik und Naturwissenschaften – Universitätsbibliothek, TIB) berichtet in ihrem Vortrag über die bisherigen Erfahrungen der Bestandserhaltung mit Ex Libris Rosetta. Rosetta wird von der Firma Ex Libris aus Israel vertrieben, die zur ProQuest Group (USA) gehört. Die Software wird aufgrund der Anforderung der Anwendergruppen dynamisch weiterentwickelt; eine davon ist die Digital Preservation Group.

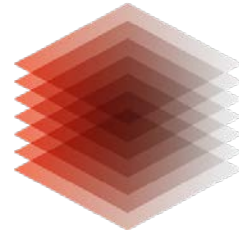
Das Rosetta-System wird an der TIB als Konsortialsystem der drei zentralen Fachbibliotheken betrieben. Dabei werden drei Systemumgebungen (Development, Test, Produktion) redundant auf mehreren Servern - jeweils mit eigener Oracle Datenbank - betrieben. Aktuell verwaltet das System rund 1,8 Mio. Files in circa 500.000 Informationseinheiten. Der Datenbestand beinhaltet 123 verschiedene Formate.

Rosetta greift auf die Inhalte der PRONOM-Datenbank zurück und reichert sie um eigene Metadaten (DNX-Modul) an. Dabei wird ein händisches Nachjustieren von automatischen Identifizierungsergebnissen ermöglicht. Über längere Zeiträume hinweg sind Re-Identifizierungen notwendig, da DROID über den Wechsel der Signaturdaten hinweg unterschiedliche Erkennungsergebnisse geliefert hat (vgl. Lindlar/Tunnat: Time Travel with PRONOM: The 4th Dimension of DROID, iPRES 2018).

Erkenntnisse im Rahmen der Diskussion:

- Identifizierung Multimedia/Webseiten: in der TIB wird ein Image einer CD-ROM in Rosetta eingespielt, aber auch alle enthaltenen Dateien parallel zur Formaterkennung.
- Eine gute Formaterkennung muss die Fragen „Wer?“, „Wer genau?“, „Womit?“, „Womit genau?“ und „Wann?“ beantworten.
- PREMIS kann neuerdings mit 3.0 die Angaben und Regelungen zu „unscharfen“ Formaterkennungen erfassen.
- Validierung ist nach wie vor ein Anliegen, das bei weitem nicht für jedes Dateiformat möglich ist.

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK



TIB

Bestandserhaltung mit Ex Libris Rosetta aus der Perspektive einer großen Forschungsbibliothek

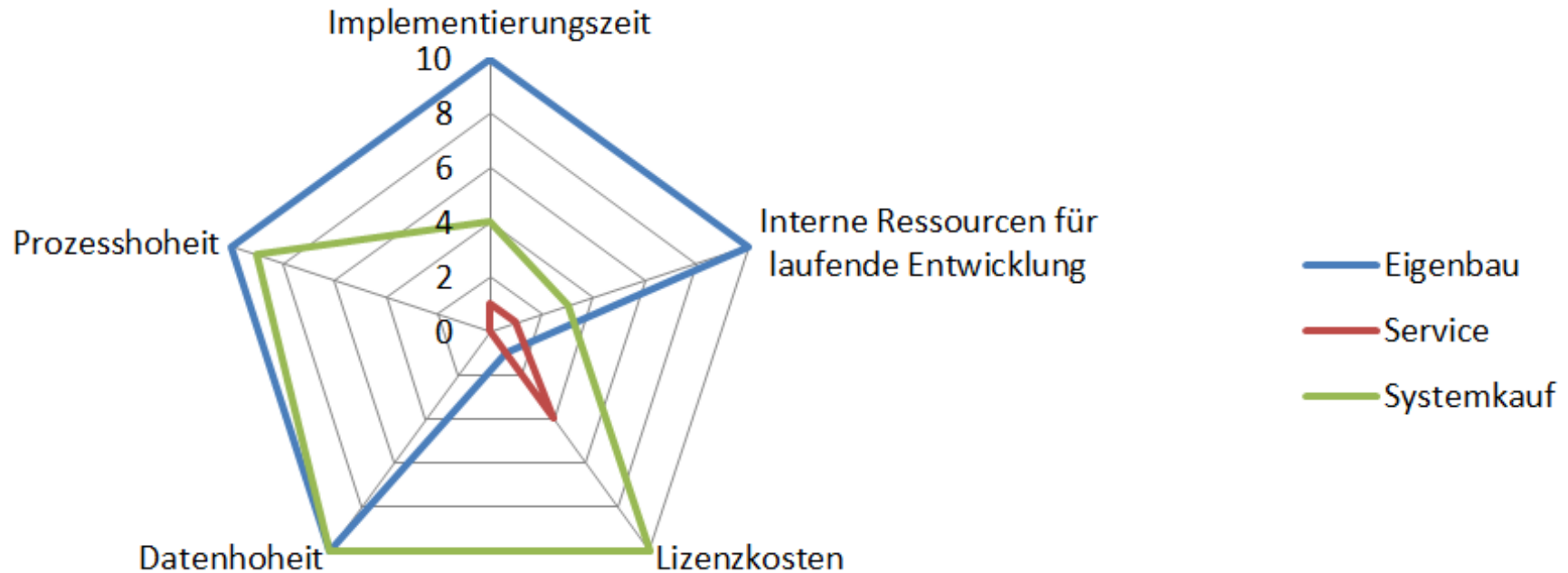
M. Lindlar

Koblenz, 7. November 2018

Workshop: „Von der Theorie zur Praxis:
Bestandserhaltung digitaler Unterlagen“



Warum Rosetta ?

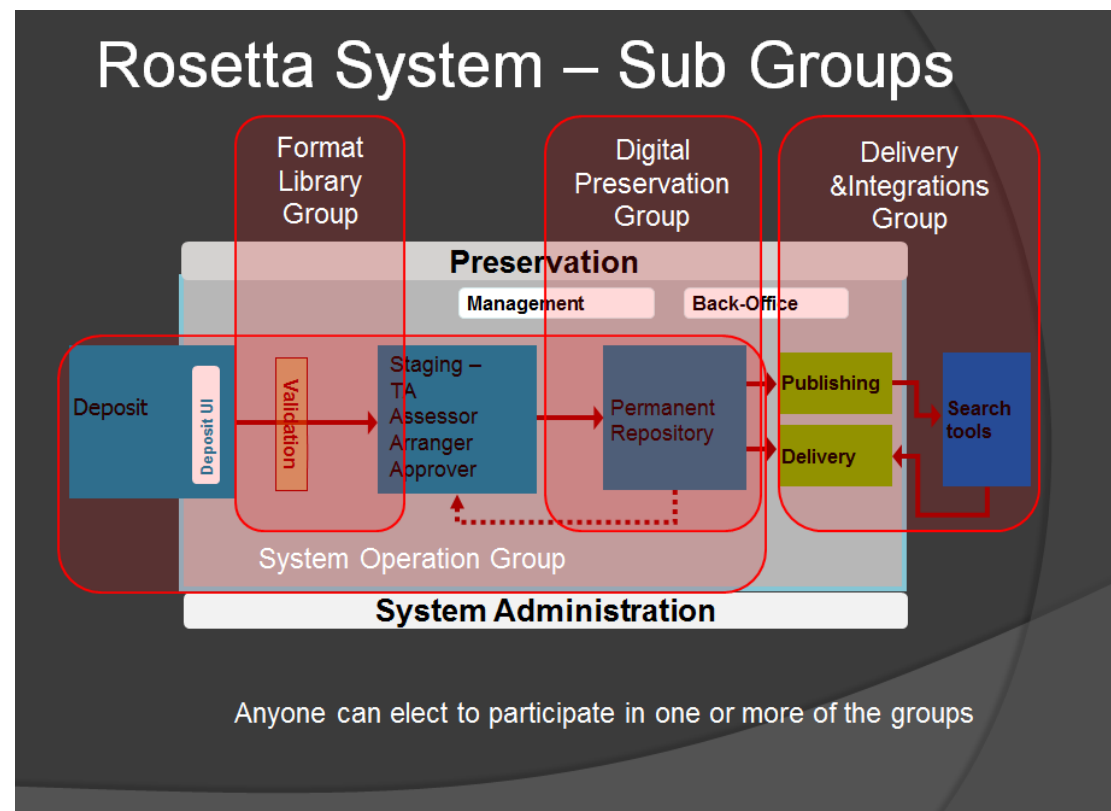


Betrieb an der TIB:

- Als Konsortialsystem der 3 zentralen Fachbibliotheken
- Für TIB Dienstleistung DPaaS (Digital Preservation as a Service)
- 3 Systemumgebungen (DEV, TEST, PROD) redundant ausgelegt mit 2, 4, 4 Application Servern + je mit eigener Oracle DB
- Virtualisierte Solaris Zonen, Storage: spinning disk ZFS via NFS
- Aktuell 1.8 Mio. Files in ca. 500.000 IEs, 123 verschiedene Formate

Rosetta User Group (RUG) Modell

- Kundengesteuerte Weiterentwicklung via Rosetta User Groups
- **Beispiele für Enhancement Requests:**
 - „Enhancing Rosetta’s Load Balancing Behavior“ (SOWG)
 - „SRU Request to get the File PID“ (DIWG)
 - „Semantic Unit Preservation Level Type“ (DPWG)
 - „Display Rule ID in Rule List and Rule Details “ (FLWG)



Dateiformatidentifizierung – so sieht Erfolg aus

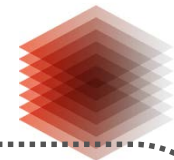
MID	DNX_FL2600341	Metadata Type	dnx:dnx
Created by	lindlarm	Creation Date	05/11/2018 22:07:22
Updated by		Update Date	05/11/2018 22:07:40

object Characteristics	
internal Identifier	
file Fixity	
event	
vs Outcome	
file Virus Check	
file Format	
general File Characteristics	

agent	REG_SA_DKROID
format Registry	PRONOM
format Registry Id	fmt/412
format Registry Role	-
format Name	fmt/412
format Version	2007
format Description	Microsoft Word for Windows
format Note	-
exact Format Identification	true
mime Type	application/vnd.openxmlformats-officedocument.wordprocessingml.document
agent Version	6.4
agent Signature Version	Binary SF v.94/ Container SF v.23
Identification Method	Container
Format Library Version	5.5094

[Back](#)

Dateiformatidentifizierung – was kann passieren?



Mögliche Fehler

Error:
Prozessabbruch

0 Treffer

> 1 Treffer

Erkennung nur
basierend auf
Endung

Endung !=
signaturebasierte
Erkennung

kann

kann

kann

kann

kann

Erneute
Überprüfung

Datei ersetzen

PUID setzen

Ignorieren

Technical Analyst Workbench

Organisatorischer Prozess

Einrichtung
Regel

Neuer
Formateintrag

Update
Formateintrag

Vorbehandlung




Technical Analyst Workbench

SIP ID	301698	Producer	Sanity and Function Tests	Producer Agent	Michelle Lindlar 
Deposit Activity ID	314954	Material Flow Name	Sanity Test Material Flow - Manual	Material Type	TestObjects
Title	FunctionalityTest	Selected Access Rights	Keine Beschränkung	Submitted On	05/11/2018 22:43:17


- Content List
- Metadata
- Notes
- History

Filter Find in


View 1 - 3 of 3 Records

<input type="checkbox"/>	File Name	PID 	Status	Problems 						
1	<input type="checkbox"/> currentlyDROID_unknown_1_2.pdf	FL2600345		1 Tasks failed 	Download	Replace	Recheck	Edit	View	Format

Fixity Check: **Pass**
 Virus Check: **Pass**
 File Format: **Fail** - Multiple formats found for file (vs_Error.5)

2	<input type="checkbox"/> PDF_Pelz.pdf	FL2600346		1 Tasks failed 	Download	Replace	Recheck	Edit	View	Format
---	---------------------------------------	-----------	--	--	--------------------------	-------------------------	-------------------------	----------------------	----------------------	------------------------

Fixity Check: **Pass**
 Virus Check: **Pass**
 File Format: **Fail** - File extension is incompatible with identified format (vs_Error.17)

3	<input type="checkbox"/> Site_issue.pdf	FL2600347		1 Tasks failed 	Download	Replace	Recheck	Edit	View	Format
---	---	-----------	--	--	--------------------------	-------------------------	-------------------------	----------------------	----------------------	------------------------

Fixity Check: **Pass**
 Virus Check: **Pass**
 File Format: **Fail** - Multiple formats found for file (vs_Error.5)

Analyse & Formatentscheidung

Submissions: Technical Issues / Details

The format ID of the file: **currentlyDROID_unknown_1.2.pdf** will be updated. Please choose the new format ID

- Manually assign
fmt/1129
- Manually assign
fmt/14
- Manually assign
fmt/144
- Manually assign
fmt/145
- Manually assign
fmt/146
- Manually assign
fmt/147
- Manually assign
fmt/148
- Manually assign
fmt/15
- Manually assign
fmt/157
- Manually assign
fmt/158
- Manually assign
fmt/16

```
1  [REDACTED]
CT_1a.pdf[REDACTED]
[REDACTED]
[REDACTED]PDF CARON[REDACTED]yyy[REDACTED]
1 P[REDACTED] j f æ ' j f k [REDACTED]
[REDACTED] j f S [REDACTED] %PDF - 1.2
2 %ããïÓ
3 61 0 obj
4 <<
```

Manually assign other format:

Remember my decision concerning the chosen format

Reason

Optional Note

PREMIS Sicht – fileFormat section

```
<section id="fileFormat">
  <record>
    <key id="formatRegistry">PRONOM</key>
    <key id="formatRegistryId">fmt/16</key>
    <key id="formatName">fmt/16</key>
    <key id="formatVersion">1.2</key>
    <key id="formatDescription">Portable Document Format</key>
    <key id="mimeType">application/pdf</key>
    <key id="exactFormatIdentification">>false</key>
    <key id="agentSignatureVersion">Binary SF v.94/ Container SF v.23</key>
    <key id="agentVersion">6.4</key>
    <key id="formatLibraryVersion">5.5094</key>
    <key id="agent">User</key>
    <key id="identificationMethod">Manual</key>
    <key id="formatNote">Other:Verified manually</key>
  </record>
</section>
```

identificationMethod – mögliche Values:

Signature, Container, Extension, Manual, Rule

exactFormatIdentification:

True → 1 Treffer

False → regelbasierte oder manuelle Zuordnung nach 0, >1, ungenauen Treffern

PREMIS Sicht – Event

```

<record>
  <key id="eventDateTime">2018-11-05 22:43:28</key>
  <key id="eventType">VALIDATION</key>
  <key id="eventIdentifierType">DPS</key>
  <key id="eventIdentifierValue">25</key>
  <key id="eventOutcome1">FAILURE</key>
  <key id="eventDescription">Format Identification performed on file</key>
  <key id="linkingAgentIdentifierType1">SOFTWARE</key>
  <key id="linkingAgentIdentifierValue1">REG_SA_DROID , Version 6.4 , Signature version Binary SF v.94/ Container SF v.23</key>
  <key id="eventOutcomeDetail1">FORMAT_ID=fmt/1129 fmt/14 fmt/144 fmt/145 fmt/146 fmt/147 fmt/148 fmt/15 fmt/157 fmt/158 fmt/16
  fmt/17 fmt/18 fmt/19 fmt/20 fmt/276 fmt/354 fmt/476 fmt/477 fmt/478 fmt/479 fmt/480 fmt/481 fmt/488 fmt/489 fmt/490 fmt/491
  fmt/492 fmt/493 fmt/558 fmt/559 fmt/560 fmt/561 fmt/562 fmt/563 fmt/564 fmt/565
  fmt/95;FILE_EXTENSION=pdf;DEPOSIT_ACTIVITY_ID=314954;PID=FL2600345;SIP_ID=301698;PRODUCER_ID=58924046;TASK_ID=48;PROCESS_ID=;MF_ID
  361075;</key>
</record>
<record>
  <key id="eventDateTime">2018-11-05 23:00:12</key>
  <key id="eventType">PROCESSING</key>
  <key id="eventIdentifierType">DPS</key>
  <key id="eventIdentifierValue">38</key>
  <key id="eventOutcome1">SUCCESS</key>
  <key id="eventOutcomeDetail1">key:FORMAT_ID;value:fmt/16;key:FILE_EXTENSION;value:pdf;key:REASON_CODE;value:Other:Verified
  manually;key:PID;value:FL2600345;key:DEPOSIT_ACTIVITY_ID;value:314954;key:TA_NOTES;value:Other:Verified
  manually;key:SIP_ID;value:301698;key:PRODUCER_ID;value:58924046;key:MF_ID;value:6361075;</key>
  <key id="eventDescription">Manually Set Format Library ID on File</key>
  <key id="linkingAgentIdentifierType1">USER</key>
  <key id="linkingAgentIdentifierValue1">lindlarm</key>
  <key id="linkingAgentRole1">TA</key>
</record>

```

- Separate Events für Toolstack (VALIDATION Stack) und manuellen Prozess (PROCESSING)
- Beispiel oben: multiple hits & manuelle Entscheidung für PUID

Beispiel Rule: Adobe Illustrator Extension Mismatch



Rule Editor

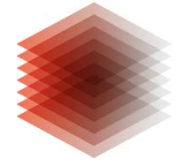
* Name	Extension Mismatch: pdf(fmt/564)		
Description	Files with pdf extension and fmt/564 format will be ignored.		
Created By:	Merle Friedrichsen	Created on:	12/01/2018 14:17:53
Updated By:	Merle Friedrichsen	Last Update on:	12/01/2018 14:19:17

Input General Parameters

* Extension	List Equals	<table><thead><tr><th>1 items selected</th><th>Add</th><th>Remove all</th><th>Add all</th></tr></thead><tbody><tr><td>\$ pdf</td><td></td><td>--</td><td>0</td></tr><tr><td></td><td></td><td></td><td>3dd</td></tr><tr><td></td><td></td><td></td><td>3dm</td></tr><tr><td></td><td></td><td></td><td>3dmf</td></tr><tr><td></td><td></td><td></td><td>3ds</td></tr><tr><td></td><td></td><td></td><td>3ds</td></tr><tr><td></td><td></td><td></td><td>3fr</td></tr><tr><td></td><td></td><td></td><td>3pp</td></tr></tbody></table>	1 items selected	Add	Remove all	Add all	\$ pdf		--	0				3dd				3dm				3dmf				3ds				3ds				3fr				3pp
1 items selected	Add	Remove all	Add all																																			
\$ pdf		--	0																																			
			3dd																																			
			3dm																																			
			3dmf																																			
			3ds																																			
			3ds																																			
			3fr																																			
			3pp																																			
* Mime Type	Any	<table><thead><tr><th>0 items selected</th><th>Remove all</th><th>Add all</th></tr></thead><tbody><tr><td></td><td></td><td>application/applefile</td></tr><tr><td></td><td></td><td>application/dbase</td></tr><tr><td></td><td></td><td>application/dec-dx</td></tr><tr><td></td><td></td><td>application/dicom</td></tr><tr><td></td><td></td><td>application/dwf,application/x-dwf,drawing/x-dwf,image/vnd.dwf,image/x-dwf,model/vnd</td></tr><tr><td></td><td></td><td>application/epub+zip</td></tr><tr><td></td><td></td><td>application/finale</td></tr></tbody></table>	0 items selected	Remove all	Add all			application/applefile			application/dbase			application/dec-dx			application/dicom			application/dwf,application/x-dwf,drawing/x-dwf,image/vnd.dwf,image/x-dwf,model/vnd			application/epub+zip			application/finale												
0 items selected	Remove all	Add all																																				
		application/applefile																																				
		application/dbase																																				
		application/dec-dx																																				
		application/dicom																																				
		application/dwf,application/x-dwf,drawing/x-dwf,image/vnd.dwf,image/x-dwf,model/vnd																																				
		application/epub+zip																																				
		application/finale																																				
* File Size(KB)	Any																																					
* Create Date	Any	dd/MM/yyyy																																				
* Plugin Instance Name	Any	FFDroidIdentifier																																				
* Identification Method	Any	container																																				

Output Parameters:

* Format Name	Ignore (extension mismatch only)
Reason	Problem not relevant for preservation



Play it again, Sam – Re-Identifizierung

Eingang von
Daten in das
Langzeitarchiv

(Re-)Analyse
des Contents

Festhalten der
Ergebnisse /
des Prozesses
in PREMIS
Metadaten und
DB

Preservation /
Technology /
Community /
Organization
Watch

Integration
neuer Tools /
neuer Patterns

```

</section>
- <section id="fileFormat">
  - <record>
    <key id="agent">REG_SA_DROID</key>
    <key id="formatRegistry">PRONOM</key>
    <key id="formatRegistryId">fmt/16</key>
    <key id="formatRegistryRole"/>
    <key id="formatName">fmt/16</key>
    <key id="formatVersion">1.2</key>
    <key id="formatDescription">Portable Document Format</key>
    <key id="formatNote"/>
    <key id="exactFormatIdentification">>true</key>
    <key id="mimeType">application/pdf</key>
    <key id="agentVersion">6.01</key>
    <key id="agentSignatureVersion">Binary SF v.81/ Container SF v.1</key>
    <key id="formatLibraryVersion">4.1081</key>
  </record>
</section>

```



Paul Young @pmyoung84 · 14h

New PRONOM release! V89 now available. 21 new PUIDS, 35 updated entries and 19 new sigs #PRONOM #DROID nationalarchives.gov.uk/PRONOM/Default

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK



**Und jetzt – Diskussion!
Fragen? Anmerkungen!**

Kontakt:

M. Lindlar – TIB Hannover



Michelle.lindlar@tib.eu



0511 762 19826



Lindlarm



mickylindlar



**Welttag Digitale
Erhaltung**

29. November 2018

Yvonne Tunnat | Ob ihr wirklich richtig steht, seht ihr, wenn das Licht angeht – die nestor AG Formaterkennung

Yvonne Tunnat (Leibniz-Informationszentrum Wirtschaft, ZBW Kiel) geht in Ihrem Vortrag zunächst auf die Begrifflichkeiten Formaterkennung und Formatvalidierung ein und stellt dar, wie sich diese Grundfähigkeiten in Risikomanagement, Preservation Planning und Qualitätskontrolle fortsetzen.

Es gibt eine große Zahl Erkennungstools, die Validierung ist hingegen noch längst nicht überall machbar. So ist JHOVE nach wie vor die einzig verfügbare Validierungsfunktion für generisches PDF, denn veraPDF ist nur für PDF/A tauglich.

Tunnat geht in der Folge auf die verschiedenen Tools und ihre Fähigkeiten und Schwächen ein. Auch sie weist deutlich darauf hin, dass eine Re-Identifizierung aller archivierten Dateien als Daueraufgabe ansteht, da sich die Formatsignaturen in allen Tools ständig weiterentwickeln.

Als Anwendungsbeispiel an der ZBW mit Ex Libris Rosetta schildert sie die Formatmigration einer großen Menge von PDF-Dateien, die mit einer Erfolgsquote von 90 Prozent zu Ende ging. Eine Qualitätskontrolle dieser Bestandserhaltungsaktionen konnte in Rosetta noch nicht erreicht werden. Ein Workflow kontrolliert Teile der Qualitätskriterien. Für Rückfragen steht Tunnat gerne zur Verfügung.

Tunnat wirft im Laufe ihres Vortrags, die Frage auf, ob das Validierungstool JHOVE immer noch der einzige Validator für Standard-PDF sei. Die Antwort darauf laute „Ja“, da veraPDF nicht für die Validierung von Standard-PDF tauglich. Zusätzlich zu JHOVE könne pdfaPilot genutzt werden um die Struktur der PDF-Datei zu prüfen. Im Hinblick auf die Validierung von JPEGs durch JHOVE stellt Tunnat heraus, dass defekte JPEGs manchmal nicht erkannt würden. Bei der TIFF-Validierung wurden hingegen keine maßgeblichen Defekte übersehen (DPF Manager sei hier jedoch nutzerfreundlicher). Bei der Validierung von GIF-Dateien wurden vereinzelt falsche Alarme ausgegeben, JHOVE und ImageMagick seien hier jedoch insgesamt auf einem Level zu sehen. Weitere Tipps und Hinweise zur Bedienung von Erkennungstools finden sich im Wiki der Nestor AG Formaterkennung.

Im Anschluss an ihre Ausführungen zum Tool-Benchmarking stellt Tunnat die Arbeit der nestor AG Formaterkennung seit 2014 vor. Die AG setzt sich zurzeit aus 12 aktiven Mitgliedern zusammen, die sich zweimal im Jahr treffen. Wege und Ziele der AG sind unter anderem das Teilen und Erweitern von Wissen, die Beratung der Community, die Zuarbeit zu etablierten internationalen Aktivitäten und die Veröffentlichung von Anleitungen.

Diskussion und Fragen:

1. Paul Wheatley: Die beste Validierung stellt das testweise Öffnen einer Datei dar. (Blogpost: <https://www.dpconline.org/blog/a-valediction-for-validation>)

Können Regeln für den Fall aufgestellt werden, wenn verschiedene Tools unterschiedliche Ergebnisse liefern und ein Ergebnis zum gültigen gekürt werden muss?

A: Entscheidend sind Benchmarks für die Tools, um bereichsweise deren Zuverlässigkeit bestimmen zu können.

2. Tobias Steinke: FITS legt von sich aus die Konkurrenzregeln der Binnentools von FITS fest.
3. Lambert Kansy: Die Kollisionsregeln sollten gemeinsam festgelegt werden.
4. Martin Wrigley lädt zum Beitragen von fehlerhaften oder merkwürdigen Dateien an die OPF bei, um die Validierungsmechanismen der Tools zu verbessern.

Ob ihr wirklich richtig steht, seht ihr, wenn das Licht angeht –

die nestor AG Formaterkennung

07.11.2018

Koblenz

Yvonne Tunnat, ZBW Kiel

Überblick

- Formaterkennung
- Formatvalidierung
- Tool-Benchmarking und andere aktuelle Studien
- nestor AG Formaterkennung



Wer bin ich?

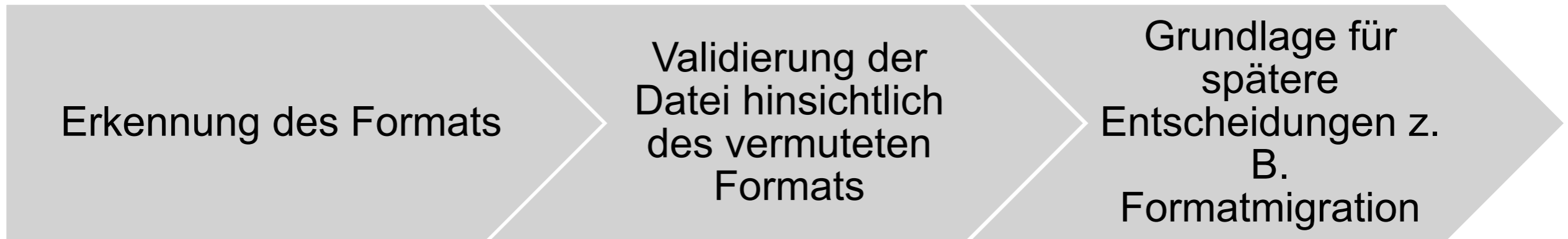
Seit 2011 Projektmanagement Digitale Langzeitarchivierung ZBW

Seit 2014 Leiterin nestor AG Formaterkennung.

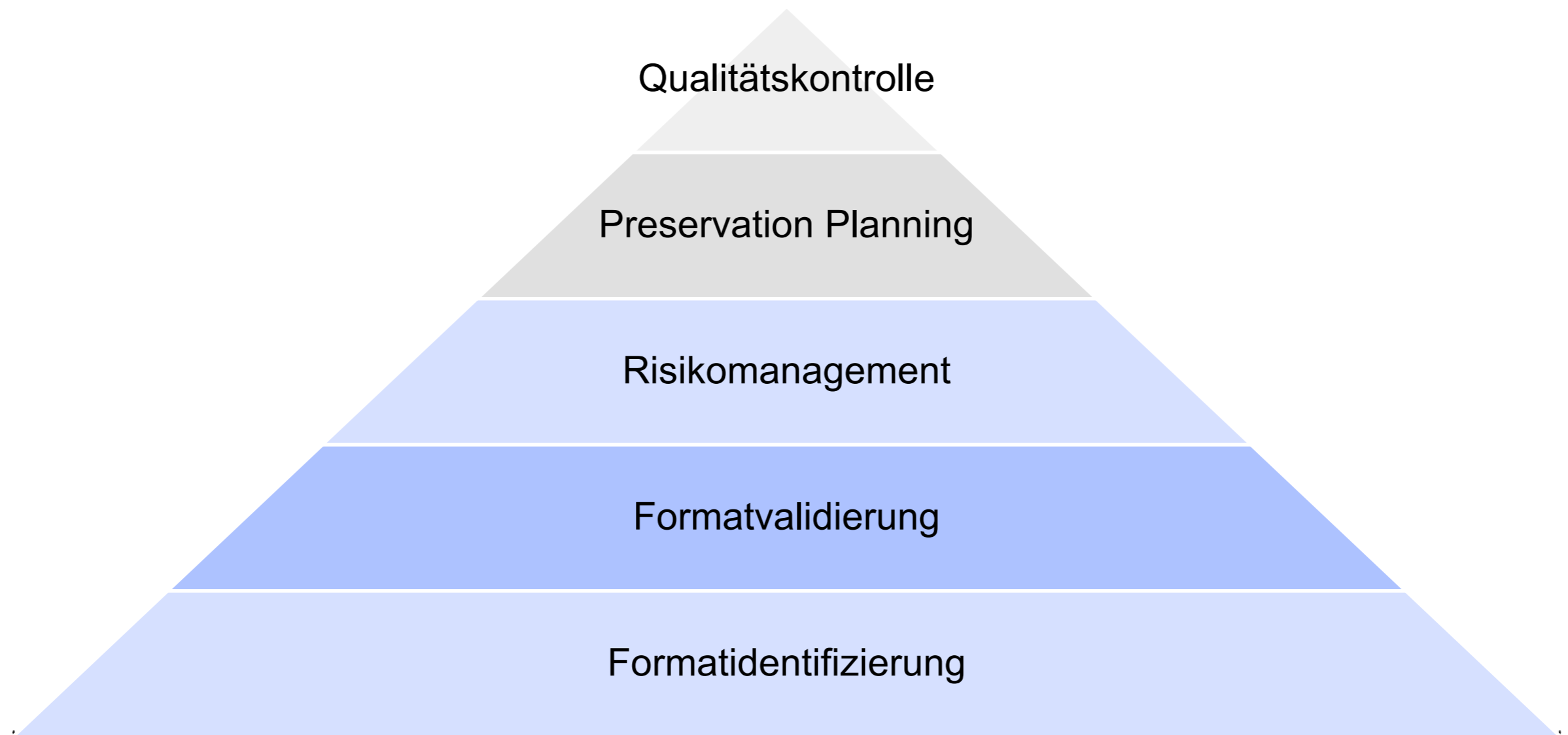
Seit 2014 in der OPF (Open Preservation Foundation) Format Interest Group.

Allgemeines zur Formaterkennung

Was bin ich und bin ich das wirklich?



Basis: Formatidentifizierung

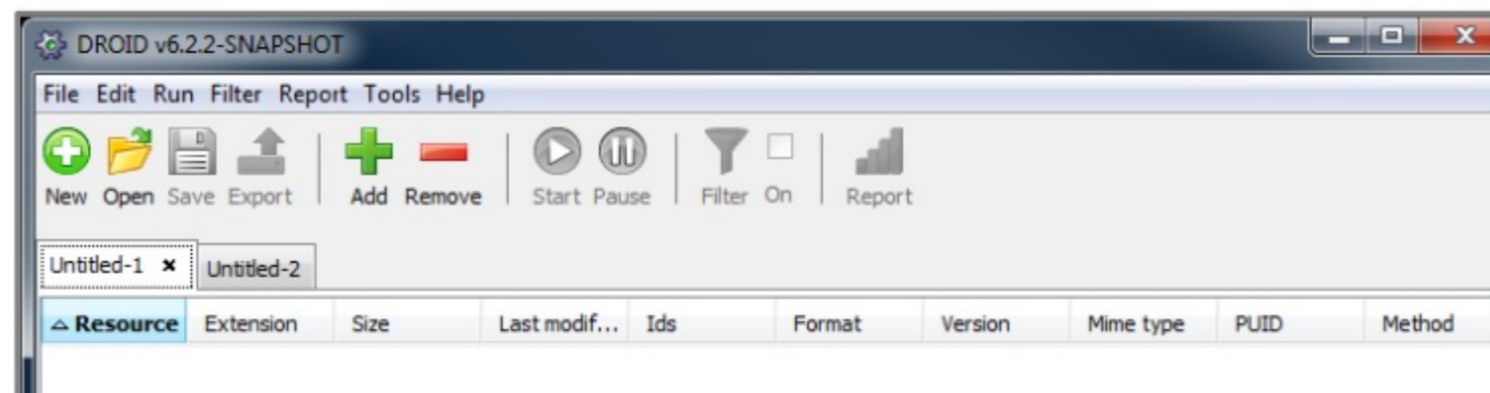


Tools zur Formaterkennung

Name	DROID	Siegfried	Fido	TrID	File	Apache Tika
Bibliothek	PRONOM	PRONOM	PRONOM	eigene	eigene	eigene
Anzahl Formate	1670	1670	1670	10.467	4450	1532
Bemerkungen		Kein GUI	Kein GUI, Benötigt Python	Mittlerweile sogar GUI	aktuelle Version läuft nur auf Linux	Nur MIME-Types

DROID

Erstmals veröffentlicht	PRONOM Signature File	Entwickelt von	Besonderheiten	Nutzungsmöglichkeiten
2006	Aktuell v94, 1670 Einträge	TNA (The National Archives) mit Zuarbeit aus der LZA-Community	Das meistgenutzte Identifizierungstool der Community	<ol style="list-style-type: none">1. GUI2. Kommandozeile (Batch)3. Java Library (wird nicht mehr aktiv unterstützt)



JHOVE (Formatvalidierung)

Erstmals veröffentlicht	Entwickelt und betreut von	Mit Unterstützung der Community via	Validierungsmöglichkeiten	Nutzungsmöglichkeiten
2005	<ul style="list-style-type: none">• Harvard University Library• Gary McGath• OPF (seit 2014)	<ul style="list-style-type: none">• OPF Product Board• Format Interest Group• JHOVE Hack Days	15 Module (u. a. PDF, TIFF, JPEG, ...)	<ol style="list-style-type: none">1. GUI2. Kommandozeile (Batch)3. Java Library



Tool-Benchmarking und andere Studien





&



Frage: Ist JHOVE immer noch der einzige Validator für Standard-PDF?

Antwort: Ja

veraPDF taugt nicht für Standard-PDF

pdfaPilot kann zusätzlich genutzt werden, um PDF-Struktur zu prüfen.
Aussage darüber, ob im Viewer (Adobe) zu öffnen.



&



Erkannte Probleme & Fazit

JHOVE entdeckt defekte JPEGs manchmal nicht.



Ergänzung nutzen.

Bad Peggy





&



Erkannte Probleme & Fazit

JHOVE ist eher streng, keine argen Defekte übersehen

DPF ist nutzerfreundlicher



&

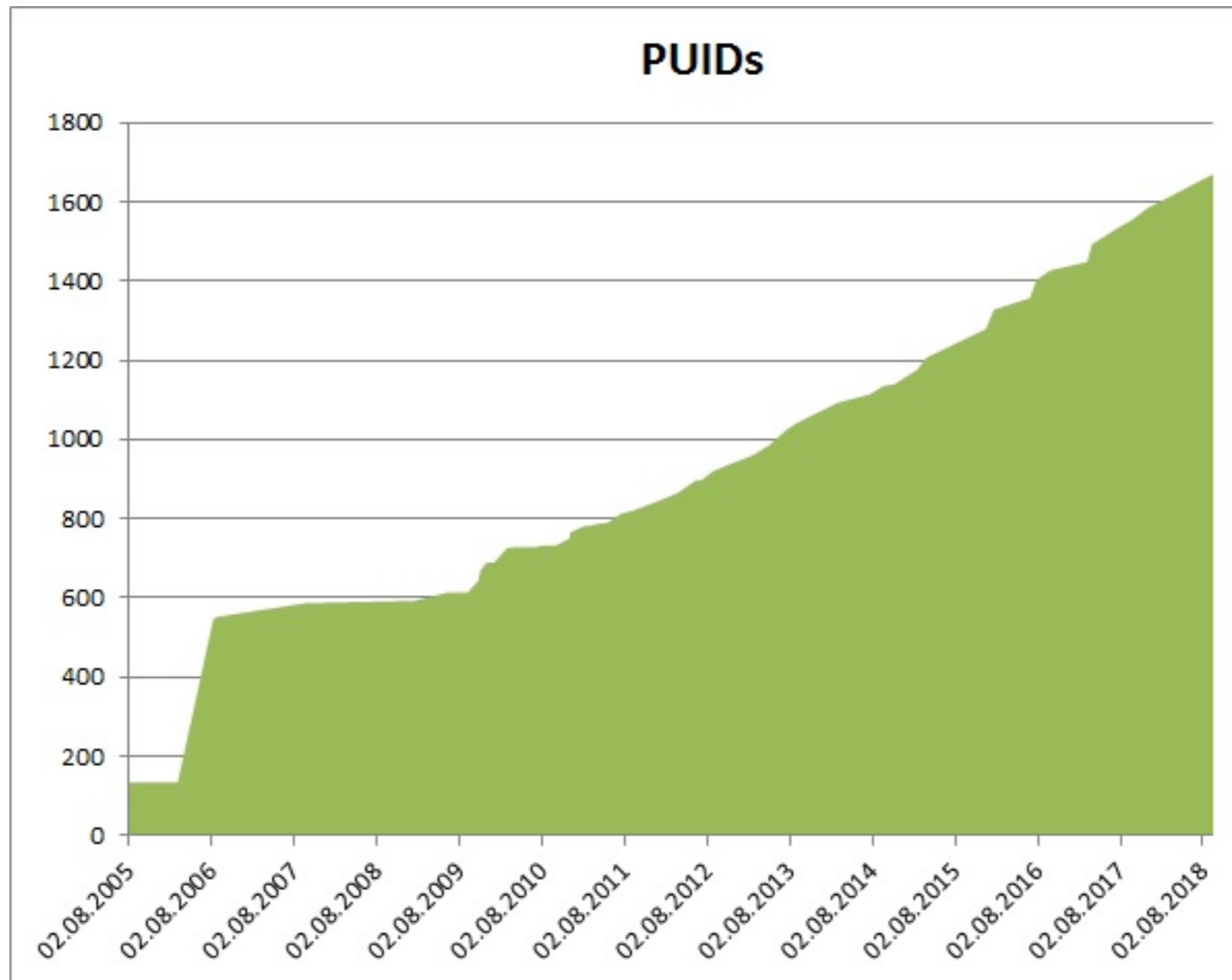


Erkannte Probleme & Fazit

Ein paar false positives (falscher Alarm)

JHOVE und ImageMagick gleichauf

Warum Re-Identifizierung?



nestor AG Formaterkennung (seit 2014)

nestor

- Seit 2003 Deutsches Kompetenznetzwerk zur digitalen Langzeitarchivierung
- 20 Partner mit 11 aktiven AGs

nestor AG Formaterkennung

- Ca. 12 aktive Mitglieder
- Bisher 10 Treffen
- Rückfragen aus der Community
- Umfangreiches wiki



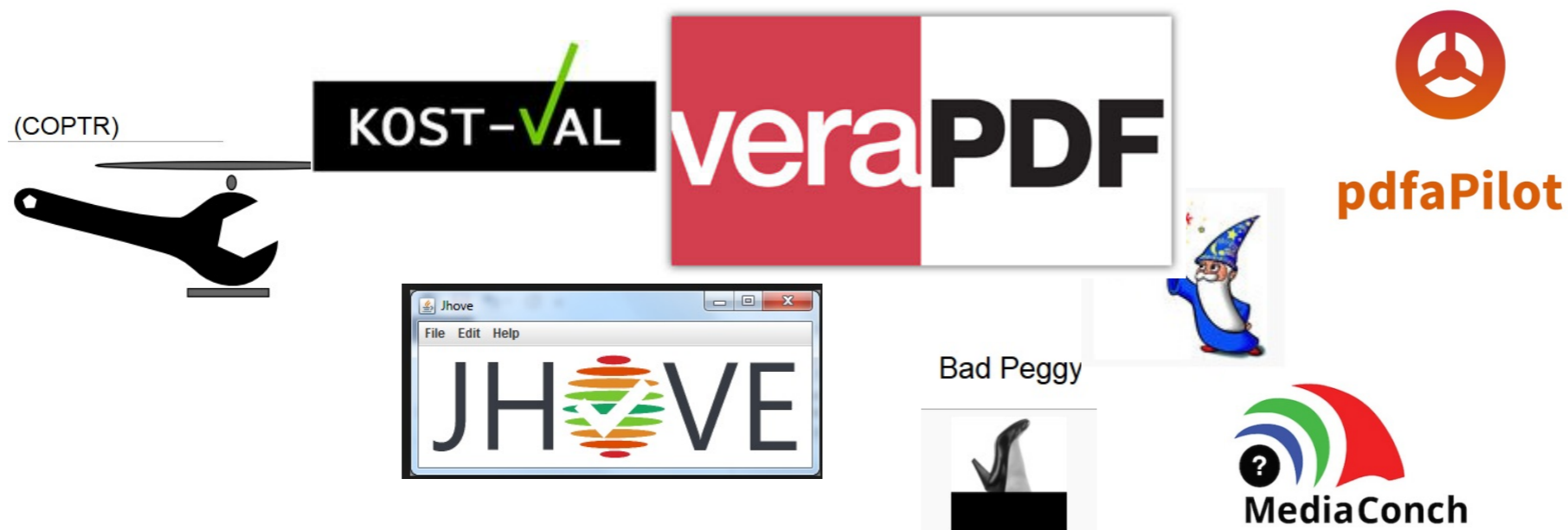
Wege und Ziele der nestor AG Formaterkennung

- Wissen teilen & erweitern
- Beratung der Community
- Zuarbeit zu bereits etablierten internationalen Aktivitäten
- Anleitungen im wiki, Blogposts, Vorträge, Paper, Rundmails, ...
- Treffen 2x im Jahr mit internen Vorträgen zu Oberthema (Pre-Ingest, Format Policy)



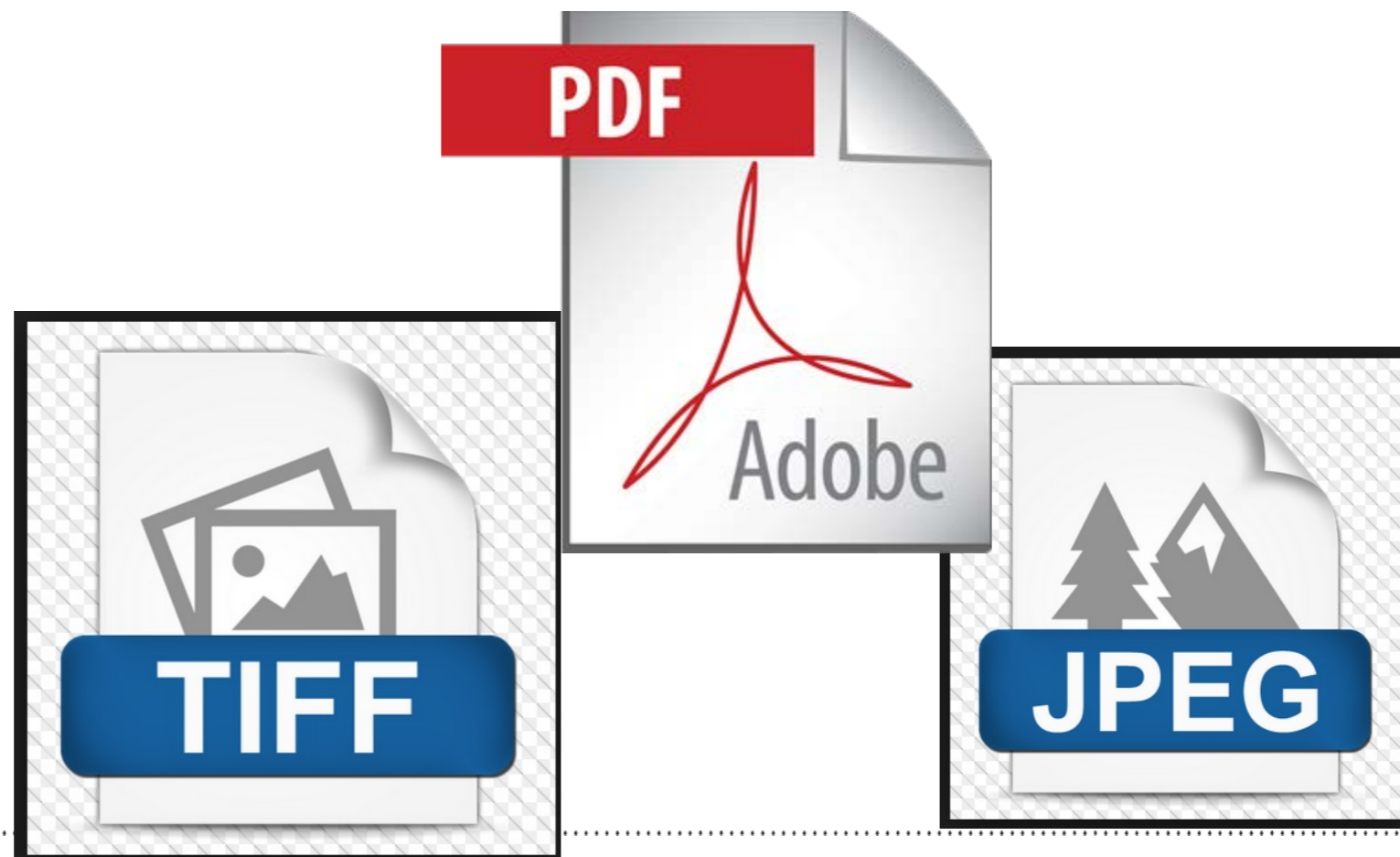
Beispiele für Aufgaben: Tool Registries

Anzahl Mitarbeitende	Laufende Aktivität
8	Zuarbeit zu COPTR



Testsuites (zum Testen und Evaluieren von Tools)

Anzahl Mitarbeitende	Laufende Aktivität	Abgeschlossene Aktivität
3 (+ alle)	Es können weiterhin Dateien hinzugefügt werden	<u>Dateien bei GitHub hinterlegt</u>



R.O.S.T

Anzahl Mitarbeitende	Laufende Aktivität	Abgeschlossene Aktivität	Bemerkungen
4 (+2)	Auswertung, geplante 2. Runde	Erste Runde der Umfrage	iPRES-Poster in 2019 / Amsterdam

Auswertung und Kontaktbörse aus der ersten Runde



Format Registries

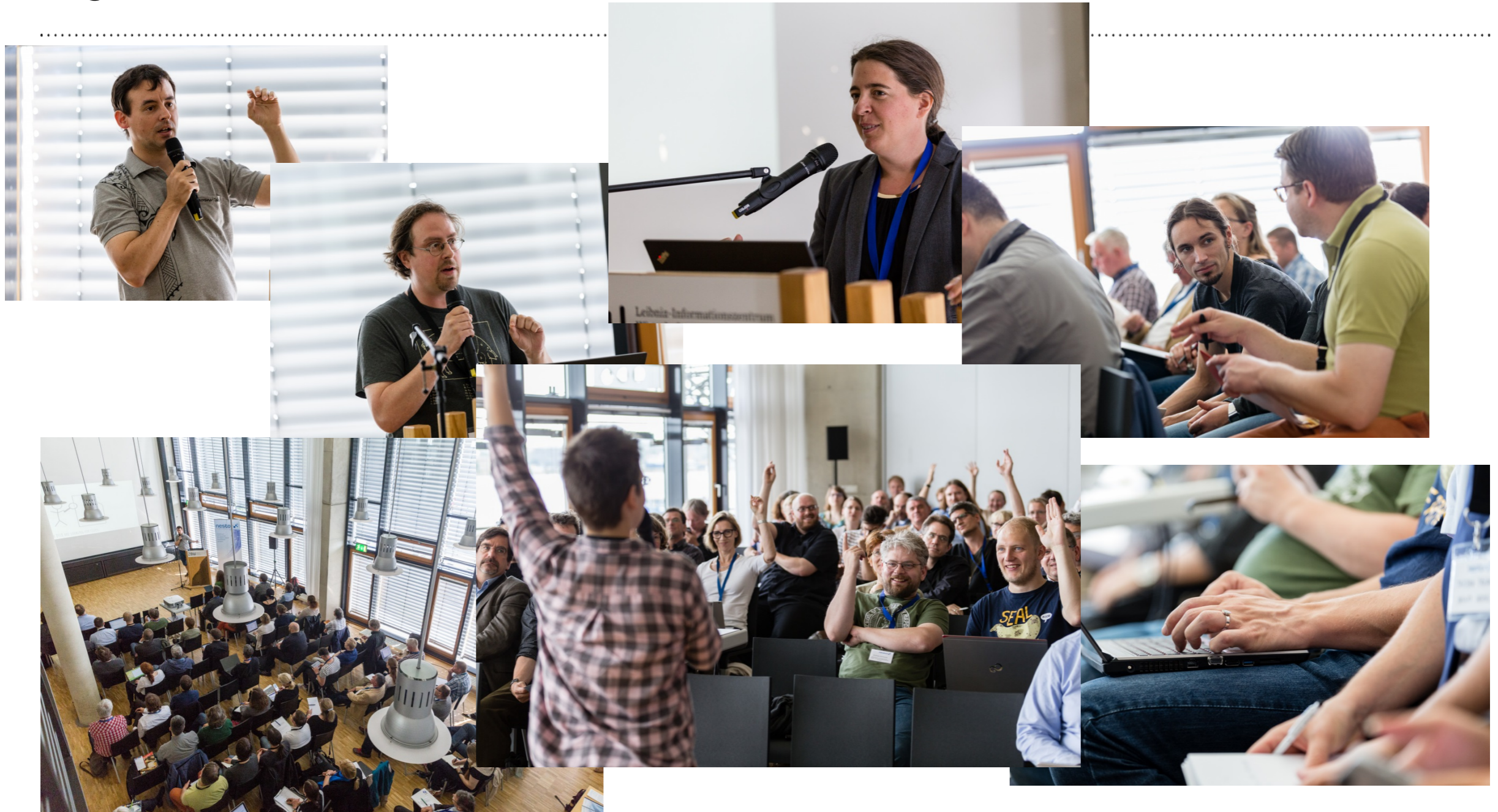
Anzahl Mitarbeitende	Laufende Aktivität	Abgeschlossene Aktivität
4	Zuarbeit zu DROID	<u>Eigene Formatsignaturen für DROID und Zuarbeit zu PRONOM</u> im öffentlichen Wiki.

The **technical registry**
PRONOM

 HARVARD LIBRARY

Global Digital Format Registry (GDFR)

Highlight: nestor Praktikertag + Workshop im Juni 2017



Ausblick

- JHOVE Workshop im Herbst 2020 beim Landesarchiv Dresden (mit der OPF, englischsprachig)
- Ausbau des wikis und der bisherigen Arbeitspakete
- Fokus auf Best Practice der praktischen Formatidentifizierung



Referenzen

DROID User Guide

Blogposts zur Identifizierung und Validierung:

- Sherlock Carriage – PRONOM's blind spot on (some) PDFs from 2010 to 2014 (DROID)
- JHOVE – the one and only PDF validator (PDF)
- Good GIF Hunting: JHOVE's GIF validation skills (GIF)
- TIFF format validation: easy-peasy? (TIFF)
- Error detection of JPEG files with JHOVE and Bad Peggy – so who's the real Sherlock Holmes here? (JPEG)

Fragen?



YVONNE TUNNAT
Projektmanagement
Langzeitarchivierung

ZBW –Leibniz-Informationszentrum
Wirtschaft

Düsternbrooker Weg 120
D-24105 Kiel

Germany

T: +49 431. 88 14-610

E: y.tunnat@zbw.eu

Jenny Mitcham (via Skype) | Making WordStar files accessible: An imperfect file migration case study

Jenny Mitcham (Borthwick Institute for Archives) eröffnet Ihren Vortrag mit einer Vorstellung des 1953 gegründeten Arbeitgebers. Sie berichtet, dass die digitale Archivierung dort zurzeit anläuft, bislang jedoch kein geeignetes digitales Bestandserhaltungskonzept etabliert werden konnte. In den letzten Jahren war Archivemata im Testbetrieb, wurde jedoch nie in den Produktivbetrieb übernommen.

Sie gibt einen exemplarischen Überblick über die bisher übernommenen genuin digitalen Datenbestände. In Hinblick auf die zu betrachtenden Datenbestände stellt sie vor allem die Arbeit mit Dateien heraus, die mit dem Textverarbeitungsprogramm Wordstar generiert worden sind. Die Dateien konnten mit heute gängiger Software aufgrund von Kompatibilitätsproblemen nur unter großem Aufwand gelesen werden.

Eine Unzulänglichkeit in Versionen vor WordStar 5.0 war, dass in jedem Byte das erste Bit anzeigte, ob dieses Byte den letzten Buchstaben eines Worts repräsentiert oder nicht. Das passte nicht gut zur Internationalisierung der IT, weil es die Verwendung von 8-Bit Zeichensätzen jenseits von ASCII unmöglich machte. Auch sorgte dieses Design dafür, dass WordStar-Texte in anderen Programmen etwas seltsam aussahen, mit merkwürdigen Zeichen an jedem Wortende.

Um solche Dateien in gängige Formate umzuwandeln, wurde zunächst vor allem die Software Quick View Plus 12 (QVP 12) benutzt. Nach einem Update auf QVP 2017 wurde WordStar jedoch nicht mehr unterstützt. Ein Versuch das Problem im Benehmen mit dem Hersteller zu lösen, scheiterte zunächst, sodass die grundsätzliche Nutzung von QVP zunehmend risikoreich wurde. Das Scheitern der Arbeit mit QVP führte dazu, dass Mitcham eigene Migrationsversuche unternahm. Diese Versuche führten zwar zu keinem vollständig perfekten Ergebnis, aber sie ermöglichten doch eine Auswertung der Objekte. Der Migration schloss Mitcham eine eingehende Analyse der signifikanten Eigenschaften von Original und migrierter Repräsentation sowie eine Qualitätssicherung an. Die Analyse ergab, dass die Migration nicht alle signifikanten Eigenschaften übertragen hatte.

Aus den gemachten Erfahrungen heraus wurden im Anschluss funktionelle Anforderungen an das Archivierungssystem Archivemata definiert. Jenny Mitcham hat über die gewonnenen Erfahrungen gebloggt (<http://digital-archiving.blogspot.com/2018/11/testing-manual-normalisation-workflows.html>). Aus diesem Blog können weitere Informationen gezogen werden.

Diskussion und Fragen:


Zitat von Tim Gollins: Alle Migrationen sind unperfekt, nur manche sind unperfekter als andere! (10. Juli 2018 via Twitter)

Making WordStar files accessible:

An imperfect file migration case study



Jenny Mitcham (Digital Archivist) - Borthwick Institute for
Archives, University of York, UK

 @Jenny_Mitcha
m

The Borthwick Institute for Archives

- 1953 - established
- 2012 - 1st digital archivist appointed
- We are now 'doing' digital archiving but we do not have a 'digital preservation system'
- We have been testing and exploring Archivematica for several years - but not in production



Background - Marks and Gran digital archive

Marks & Gran

HOME ABOUT US CREDITS BLOG CONTACT

Credits

Television

Theatre
Radio

Click on the shows below to reveal more



- SNAKES AND LADDERS** - YORKSHIRE TELEVISION for CHANNEL 4. SITUATION COMEDY, 1 SEASON - 1989.
- BIRDS OF A FEATHER** - ALOMO PRODUCTIONS for BBC 1. SITUATION COMEDY, 9 SEASONS, 7 SPECIALS - 1989 - 1999. 2 BAFTA nominations.
- SO YOU THINK YOU'VE GOT TROUBLES** - ALOMO PRODUCTIONS for BBC 1. COMEDY-DRAMA, 1 SEASON, 1991.
- LOVE HURTS** - ALOMO PRODUCTIONS for BBC 1. COMEDY-DRAMA, 3 SEASONS, 1992-1994. BAFTA nominated.
- GET BACK** - ALOMO PRODUCTIONS for BBC 1. SITUATION COMEDY, 2 SEASONS, 1992-1993.
- WALL OF SILENCE** - ALOMO PRODUCTIONS for BBC 1. TV MOVIE, CRIME DRAMA, 1993.
- GOODNIGHT SWEETHEART** - ALOMO PRODUCTIONS for BBC 1. SITUATION COMEDY, 6 SEASONS, 1 SPECIAL, 1993-1999.
- MOSLEY** - ALOMO PRODUCTIONS for CHANNEL 4. DRAMATIC MINI SERIES, 1998.
- UNFINISHED BUSINESS** - ALOMO PRODUCTIONS for BBC 1. SITUATION COMEDY, 2 SEASONS, 1998-1999.
- STARTING OUT** - ALOMO PRODUCTIONS for BBC 1. 1 SEASON, 1999.
- DIRTY WORK** - ALOMO PRODUCTIONS for BBC 1. COMEDY-DRAMA, 1 SEASON, 2000.
- BELIEVE NOTHING** - ALOMO PRODUCTIONS for BBC 1. 1 SEASON, 2002.
- MUMBAI CALLING (Pilot)** - ALLAN McKEOWN PRESENTS LTD for ITV - 2007.
- BIRDS OF A FEATHER** - RETORT for ITV - 2014.



Background - what was on the disks?

D1_THE BRETTS (filecopy) Ep2 Ep10 Ep11	05/08/2013 16:44	File folder
D2_BRETTS Ep 11	05/08/2013 16:44	File folder
D3_THE BRETTS	05/08/2013 16:44	File folder
D4_Bretts Ep 10, Aut 1	05/08/2013 16:44	File folder
D5_My ACCOUNTANT MASTER SYSTEM...	05/08/2013 16:44	File folder
D6_WORKING CLASS HEROES	05/08/2013 16:44	File folder
D7_IDEAS	05/08/2013 16:44	File folder
D8_Copy of MASTER WORDSTAR DISK (...)	05/08/2013 16:44	File folder
D9_BOOT DISC - W.S. + SMARTKEY	05/08/2013 16:44	File folder
D10_Relative Strangers Series 2 (copy A)	05/08/2013 16:44	File folder
D10_Relative Strangers Series 2 (copy B)	05/08/2013 16:44	File folder
D11_Wordstar 2000 Demo WS2MONO	16/05/2017 16:49	File folder
D12_Part 1 HIGH PLACES Oct 1986 Edit	05/08/2013 16:44	File folder
D13_Part 2 (B) HIGH PLACES Oct 1986 Edit	05/08/2013 16:44	File folder
D14_HIGH PLACES 1987	05/08/2013 16:44	File folder
D15_2HP A	05/08/2013 16:44	File folder
D16_HP A WARD MOVIE	05/08/2013 16:44	File folder
D17_PUNCH I	05/08/2013 16:44	File folder
D18_SNAKES & LADDERS 1	05/08/2013 16:44	File folder
D19_WILLIAM	05/08/2013 16:44	File folder

176 files in total (mostly WordStar)
Dates appear to range from 1984 to 1987

Content:

- Scripts and title pages
- Character lists and location lists
- Plot outlines
- A few ideas and notes
- 1 article

Why are the WordStar files hard to access?

```
Á COUPLÁ OÆ DAYÓ LATER@ IÔ IÓ THÁ  
COCKTAIÏ HOUR@ CHARLEÓ ANÄ LYDIÁ ARÄ  
HAVINÇ Á LITTLÁ DRINKIE¬ ANÄ READING  
THROUGHÈ WILLIÄ SUTHERLAND'Ó NE× PLAY¬  
ALOUND@ AÓ THERÄ IÓ ONLÙ ONÄ COPY¬ THEÙ  
HAVÄ TÏ SIÔ SIDÄ BÙ SIDÄ OÏ Ä SOFA.
```

LYDIA

```
"Ignorancá œ thá worlä ió thá inevitablá  
producó œ á publiã school education..."
```

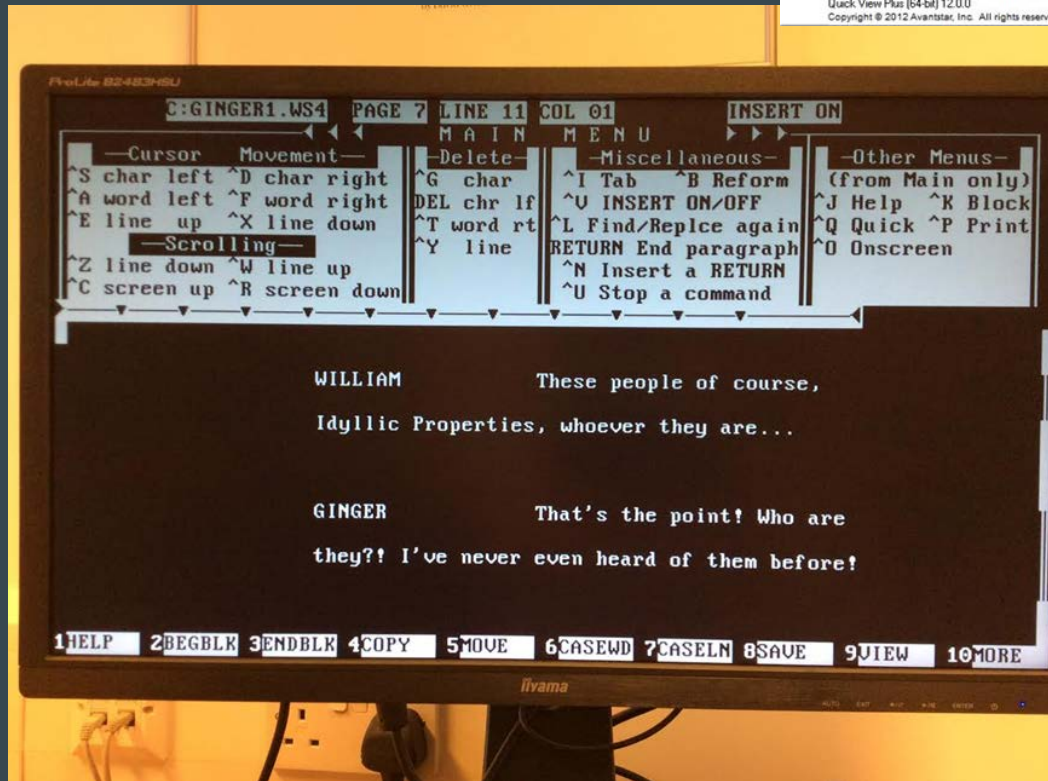
CHARLES

```
"Ignorantç Iç É caí telì á '9· clareó  
witè mù eyeó shut!"
```

“One quirk present in versions prior to 5.0 was its use of the high bit of each byte of its files to denote the last letter of a word.This interfered with internationalization, since it prevented the use of extended character sets beyond ASCII, and also resulted in WordStar files having characters at the end of words that looked like gibberish in other programs which interpreted the characters via some 8-bit encoding. Eventually this "feature" was dropped...”

<http://justsolve.archiveteam.org/wiki/WordStar>

...but I can view the files in WordStar a



The problem

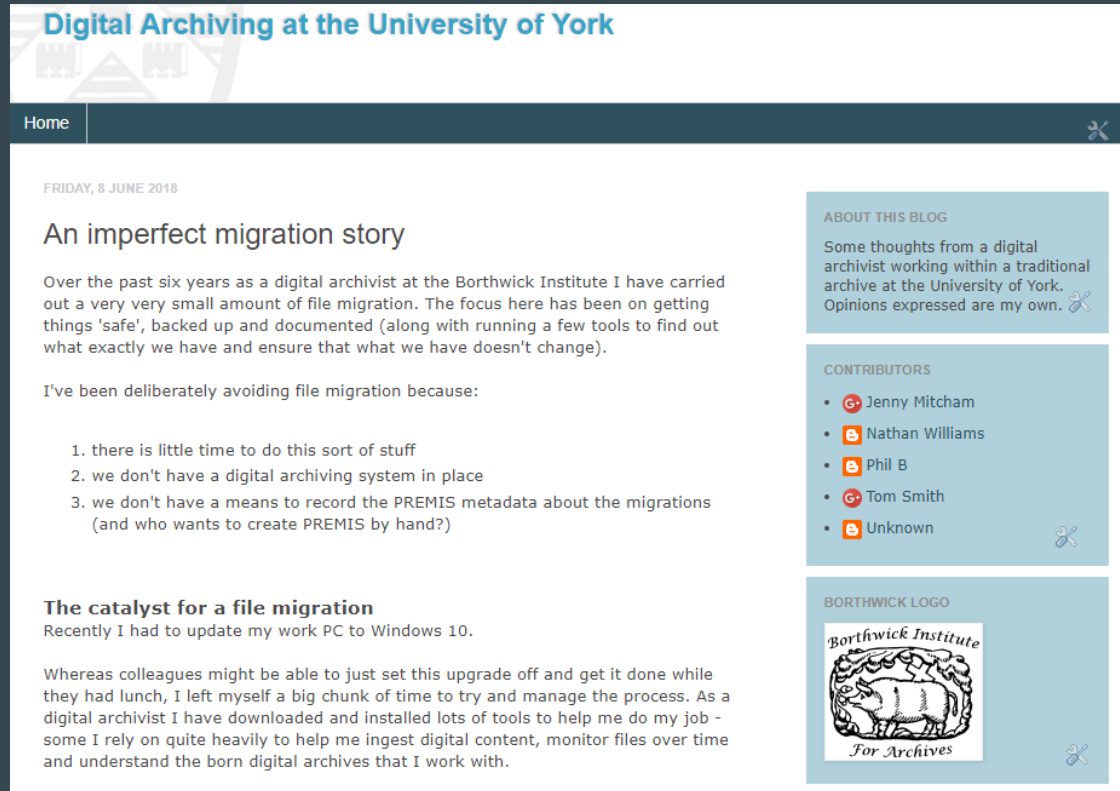
- When I upgraded my office PC to Windows 10, I couldn't reinstall QVP 12 - it wasn't compatible
- I installed a trial copy of QVP 2017 - it didn't support WordStar (despite the documentation saying it did)
- I asked the suppliers if this bug would be fixed and I was told *“Unfortunately, information is not available on future versions at this time.”*
- I was no longer comfortable with the risk of not being able to access these files

Marks and Gran migration

So, I carried out a file migration this year

More of a rescue mission really - with limited resource and technical skill available

My migration was imperfect but better than nothing.



The screenshot shows a blog post on the website 'Digital Archiving at the University of York'. The page has a dark blue header with the site name and a 'Home' link. The main content area is white with a dark blue sidebar on the right. The post title is 'An imperfect migration story' dated 'FRIDAY, 8 JUNE 2018'. The text discusses the author's experience as a digital archivist at the Borthwick Institute, mentioning file migration challenges and a Windows 10 update. The sidebar contains sections for 'ABOUT THIS BLOG', 'CONTRIBUTORS' (listing Jenny Mitcham, Nathan Williams, Phil B, Tom Smith, and Unknown), and 'BORTHWICK LOGO' (featuring a boar illustration and the text 'Borthwick Institute For Archives').

Digital Archiving at the University of York

Home

FRIDAY, 8 JUNE 2018

An imperfect migration story

Over the past six years as a digital archivist at the Borthwick Institute I have carried out a very very small amount of file migration. The focus here has been on getting things 'safe', backed up and documented (along with running a few tools to find out what exactly we have and ensure that what we have doesn't change).

I've been deliberately avoiding file migration because:

1. there is little time to do this sort of stuff
2. we don't have a digital archiving system in place
3. we don't have a means to record the PREMIS metadata about the migrations (and who wants to create PREMIS by hand?)

The catalyst for a file migration

Recently I had to update my work PC to Windows 10.

Whereas colleagues might be able to just set this upgrade off and get it done while they had lunch, I left myself a big chunk of time to try and manage the process. As a digital archivist I have downloaded and installed lots of tools to help me do my job - some I rely on quite heavily to help me ingest digital content, monitor files over time and understand the born digital archives that I work with.


ABOUT THIS BLOG

Some thoughts from a digital archivist working within a traditional archive at the University of York. Opinions expressed are my own.

CONTRIBUTORS

- Jenny Mitcham
- Nathan Williams
- Phil B
- Tom Smith
- Unknown

BORTHWICK LOGO



Borthwick Institute
For Archives

<https://digital-archiving.blogspot.com/2018/06/an-imperfect-migration-story.html>

Significant properties and Quality Assurance

After the migration I did a lot of QA looking at the original documents in WordStar and comparing with the migrated files.

...the results were pretty patchy

Key:

Green: property is captured

Yellow: property is partially captured

Pink: property is not captured

SIGNIFICANT PROPERTY			
Property	PDF files	DOCX files	TXT files
Carriage returns	Mostly captured though there are occasion examples of this not being the case - for example on 2nd page of MOBIL1.A in which CHARLES is not on a new line as it should be before first line of dialogue	Mostly captured though there are occasion examples of this not being the case - for example on 2nd page of MOBIL1.A in which CHARLES is not on a new line as it should be before first line of dialogue	Captured
Alternate pitch	Not captured	Not captured	Not captured
Underlined text	Captured	Captured	Not captured
Hard page breaks	For the most part these have been captured. In the sample of documents that were checked, there was one occasion (on page 17 of SAFE1) where this page break was not present.	Not captured	Not captured
Text	Horizontal text	Once the font was	Horizontal text

An imperfect migration?

With the skill set and resource I had available I failed to find a migration strategy that adequately preserved **all** the significant properties of the files.

However I have created versions

- PDF/A
- DOCX
- TXT

...but which are for
access and which
are for preservation?
...it depends



How would a digital preservation system handle this?

In order for Archivemata to work for us, it needs to be able to:

- Incorporate manually migrated files in the SIP, AIP and the DIP
- Record the manually migrated files with an appropriate level of metadata



The screenshot shows the Archivemata website interface. At the top is a blue header with the Archivemata logo and a navigation menu with links for Home, Downloads, Documentation, Community, Development, News, Wiki, and Demo. Below the header is a yellow banner announcing 'Archivemata 1.7.2 is our latest release.' The main content area is divided into two columns. The left column is titled 'Manual normalization' and contains text explaining that some users may use tools other than those provided in Archivemata. It describes two manual normalization workflow options: one where normalization occurs before objects are brought into Archivemata, and another where objects are processed through Ingest and then normalized. It also mentions that PREMIS event detail metadata can be added. The right column is titled 'Archivemata 1.7.2' and contains a 'CONTENTS' section with a list of links: 'Manual normalization', 'Option 1: Manual normalization prior to transfer to Archivemata', 'Option 2: Manual normalization during Ingest', 'Adding PREMIS eventDetail for manual normalization', and 'Normalizing files with the same name'. There is also a 'SEARCH' section with a search box and a link to the 'general index'.

archivemata

Home Downloads Documentation Community Development News Wiki Demo

Archivemata 1.7.2 is our latest release.

Manual normalization

Some users may have access to tools other than those provided in Archivemata that they wish to use for normalization actions.

There are two manual normalization workflow options. First, the user can do some normalization before bringing any objects into Archivemata. Second, the user can process objects through to Ingest and then choose to do manual normalization on some of the objects. Below are the workflow steps to test each of these options.

In either option, it is also possible to add PREMIS event detail metadata about the manual normalization event. See below for details.

Option 1: Manual normalization prior to transfer to Archivemata

The manual normalization feature allows the user to prepare transfers that already contain manually normalized

Archivemata 1.7.2

[Table of contents]

CONTENTS

- Manual normalization
 - Option 1: Manual normalization prior to transfer to Archivemata
 - Option 2: Manual normalization during Ingest
 - Adding PREMIS eventDetail for manual normalization
 - Normalizing files with the same name

SEARCH

Open the [general index](#) or type your search in the search box.

Testing manual normalisation workflow in Archivemática

- Files must be placed in the directory structure required by Archivemática
- It must be told whether migrated files are for access or preservation
- It will accept 1 access file and 1 preservation file per original
- CSV file must be supplied to associate files if migrated filenames don't exactly match original
- Additional info about the migration can be added to PREMIS eventDetail field

```
/topDirectory  
  
    /manualNormalization  
  
        /access  
  
            file1.jpg  
  
        /preservation  
  
            file1.tif  
  
file1.tga  
file2.doc  
file3.docx
```

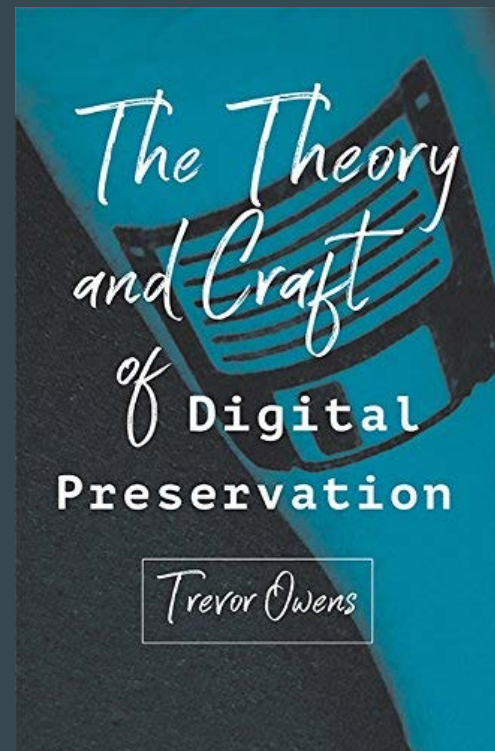
Does this meet my needs?

- Not if I have 3 imperfect versions of the files and want to be able to provide access to any one of them depending on what the user requires
- Not if I can't decide which are 'preservation' copies
- Not if I want to add full PREMIS metadata (including date and time of each event, who carried it out, what the outcome was etc)
- Not if I have pages of documentation about the significant properties, migration methodology and recording QA findings ...how do I associate these with the migrated files?

Trevor Owens - Theory and Craft of Digital Preservation

“Tools can get in the way just as much as they can help.

Specialized digital preservation tools and software are just as likely to get in the way of solving your digital preservation problems as they are to help.”



From theory to practice....

Digital preservation systems are rigid and complex. They need clear rules. They can not be endlessly flexible.

Digital archives are unique and different. There is always an exception to every rule. They don't always fit neatly into a rigid way of thinking.

Digital archivists are imperfect. None of us can see into the future. We do things to 'hedge our bets'. In the words of Trevor Owens, what we practice is more of a 'craft' than a 'science'.

Thanks for listening



<https://digital-archiving.blogspot.com>



@Jenny_Mitcha
m

Siegfried Fößel | Nachhaltige Videoformate

Herr Siegfried Fößel (Fraunhofer IIS – Erlangen) geht in seinem Vortrag zu nachhaltigen Videoformaten zunächst auf die technischen Änderungen und Entwicklungen der letzten 15 Jahre ein. Hierbei hebt er unter anderem den Wandel von analoger zu digitaler Aufnahmetechnik sowie die Fortschritte bei Bild- und Speichertechnik hervor.

In der Vergangenheit wurden die fertigen Lieferformate archiviert, heute wird Archivierung schon am Beginn der Produktion mitbedacht. Es wird angestrebt, das „Kameraoriginal“ oder wenigstens ein Master-Format zu sichern.

Bei professionellen Produktionen sind derzeit Digital Cinema Package (DCP) und ... Interoperable Master Format (IMF) die beherrschenden Formate. Das IMF bündelt die verschieden zu einem Filmprodukt gehörenden Komponenten wie mehrsprachige Tonspuren, Untertitel, Bonusaufnahmen und Kapitel mittels XML. Das IMF hat verschiedene Ausprägungen, die derzeit populärste benutzt JPEG2000 für die Bilddaten und PCM-Tonspuren. Das Material Exchange Format MXF ist als Containerformat Teil dieses Pakets.

Das TC 457, eine Arbeitsgruppe des CEN (European Committee for Standardization) erarbeitet derzeit eine Norm zur langfristigen Erhaltung von digital produzierten Filmwerken

Rein proprietäre Formate sind vom Markt verschwunden, Standardformate dominieren. Die Schöpfer der Standardformate vereinbaren eine prozentuale Beteiligung an den Patentgebühren, die während der Laufzeit des Patentschutzes fällig sind. Nach Ablauf des zwanzigjährigen Patentschutzes sind die Formate für alle Benutzung frei.

Bei der Auslieferung von AV-Material gibt es nach wie vor eine große Vielfalt an Formaten, die automatisiert erstellt werden können.

Diskussion und Fragen:

1. Tobias Steinke: Sollen alte Filme in das neue Format übertragen werden?

A: IMF#4 gilt in den Kinematheken als sinnvolle Option. Sonst sind noch keine Tendenzen erkennbar.

2. Michelle Lindlar: Die ETF hat im Preforma-Projekt als Alternative zu MXF Matroska mit FFV1 empfohlen.

A: Das MXF Format ist zugegeben etwas ausführlich und damit schwerfällig. Die Filmindustrie und die Archive scheinen hier etwas unterschiedliche Wege zu gehen.

Das AV-Objekt muss stets auf seine Eigenschaften hin geprüft werden. Aus unterschiedlichen Ausgangseigenschaften (Auflösung, Farbtiefe) ergeben sich unterschiedliche Formatlösungen.

Matroska ist in den Normierungsgremien der ISO eher unbekannt. Viele Gremien konkurrieren um die Durchsetzung “ihrer” Erfindung.

3. Fr. Dr. Ulrike Gutzmann: Das Grundproblem ist, dass die Produktion sich stets ein aktuelles Format wünscht, sodass die alten Formate sehr schnell wieder obsolet werden.

A: Es ist stets gut, sich an ISO-Normen zu halten, denn diese sind spätestens nach 20 Jahren lizenzfrei.

4. Wie lange dauert es noch bis zur Fertigstellung eines IMF Archiv-Standards?

A: Mit der Fertigstellung ist in ein bis zwei Jahren zu rechnen.

NACHHALTIGE VIDEOFORMATE - ARCHIVFORMATE FÜR BEWEGTBILD

Siegfried Foessel, Fraunhofer IIS - Germany



Outline

- Technology changes
- Workflow examples during movie production and distribution
- Formats
- Summary

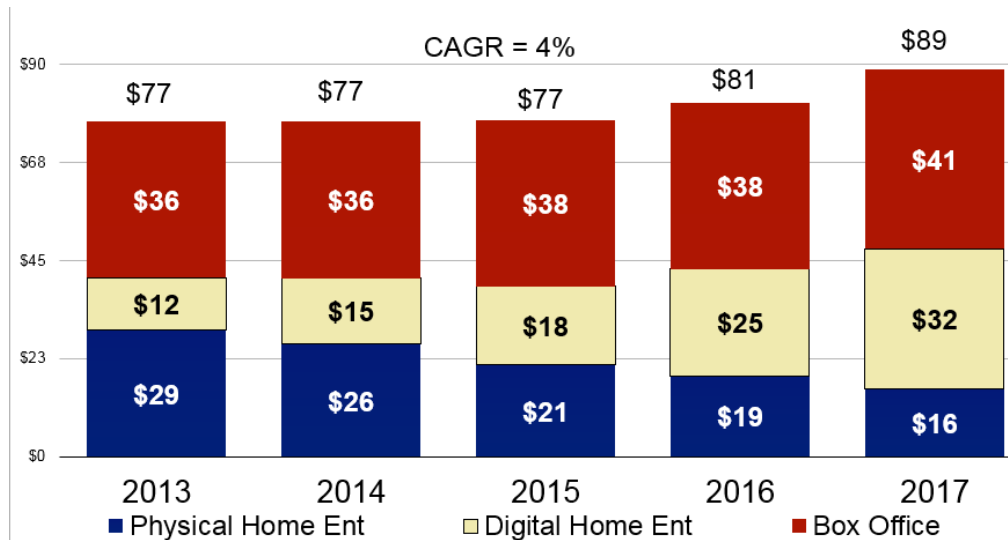
Technology changes in the last 15 years

- From analog to digital
 - no tape-based video formats, no film any more
 - file based workflows, storage media agnostic
- Improved image sensor technology (from 10 f-stops to 16 f-stops, from SD resolution to 8k resolution, from BT.601 to BT.2020)
 - More data during capturing
- Improved display technology (from CRT to LCD displays to OLED/Laser)
 - More data during distribution
- Increased storage capacity

Technology changes in the last 15 years

■ Global film revenue changes

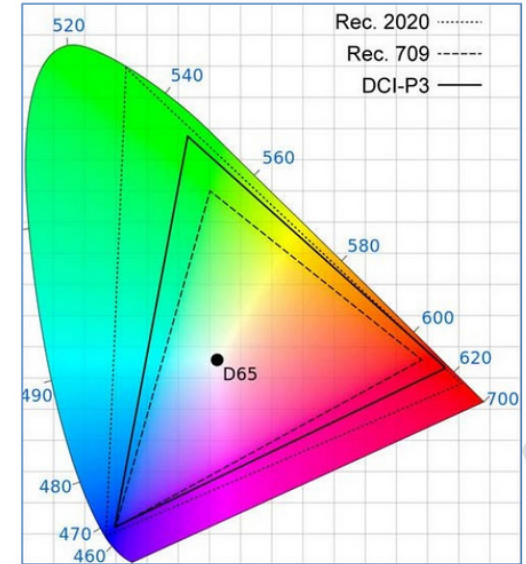
- Growth on box office caused by China (Germany constant)
- Digital Home growths nearly factor 3 in 4 years
- Decline of Physical Home (Germany down to 600 videotheken from 3000)



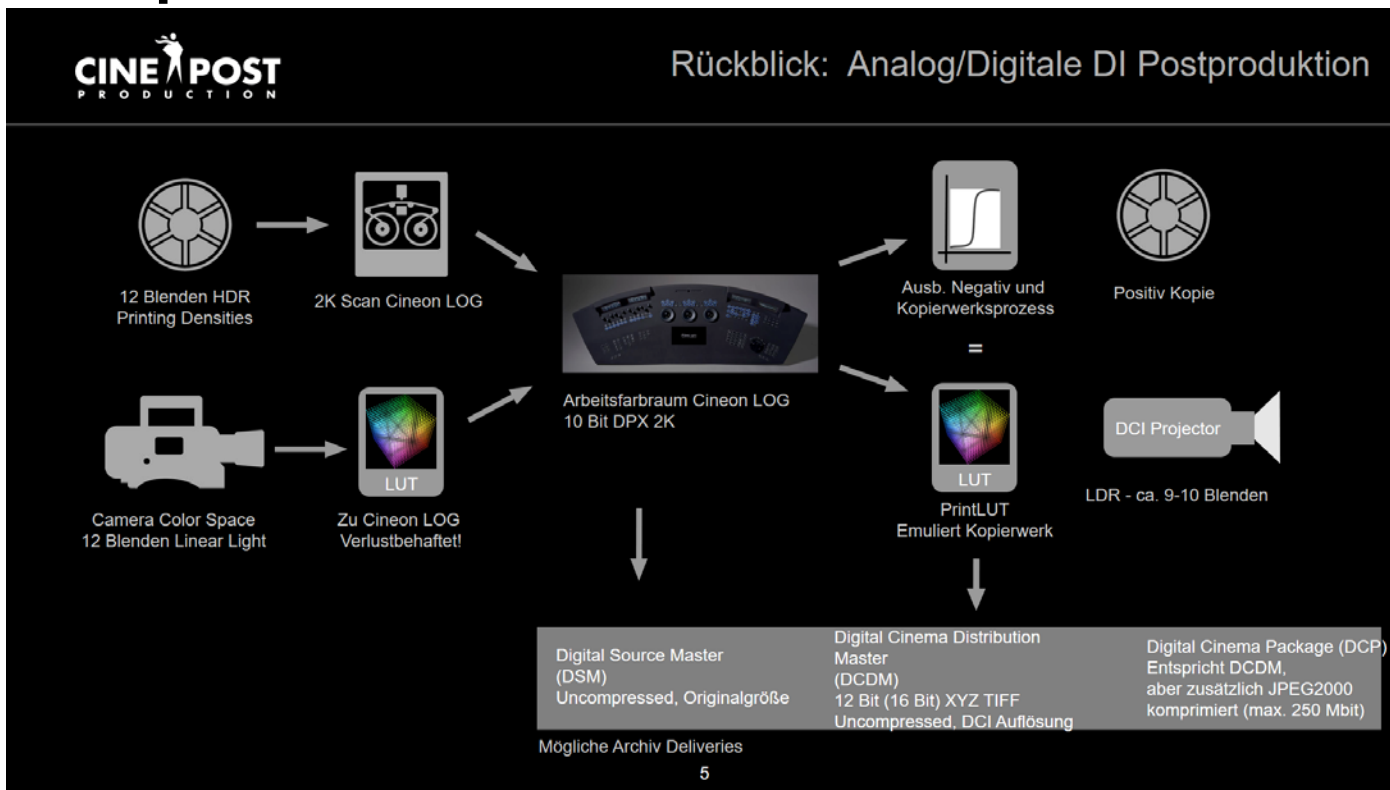
Technology changes in the last 15 years

Format	Res. / FR	Bitrate (uncompr.)	Bitrate (compr.)
SD	720x576/50i	270 Mbps	1-2Mbps
HD	1920x1080/50p	3 Gbps	6-15Mbps
UHD-1	3840x2160/60p	12 Gbps	15-50Mbps
UHD-2	7680x4360/120p	96 Gps	
Dcinema 2k	2048x1080/24p	1.5 Gbps	250Mbps
Dcinema 4k	4096x2160/60p	12 Gbps	250-500Mbps

- Today compression 10:1 to 30:1 accepted for HQ Master
- HDR: Today 16/32Bit processing or half float
- Every 10 year: Storage capacity increase by factor 8-10, compression efficiency increase by factor 2

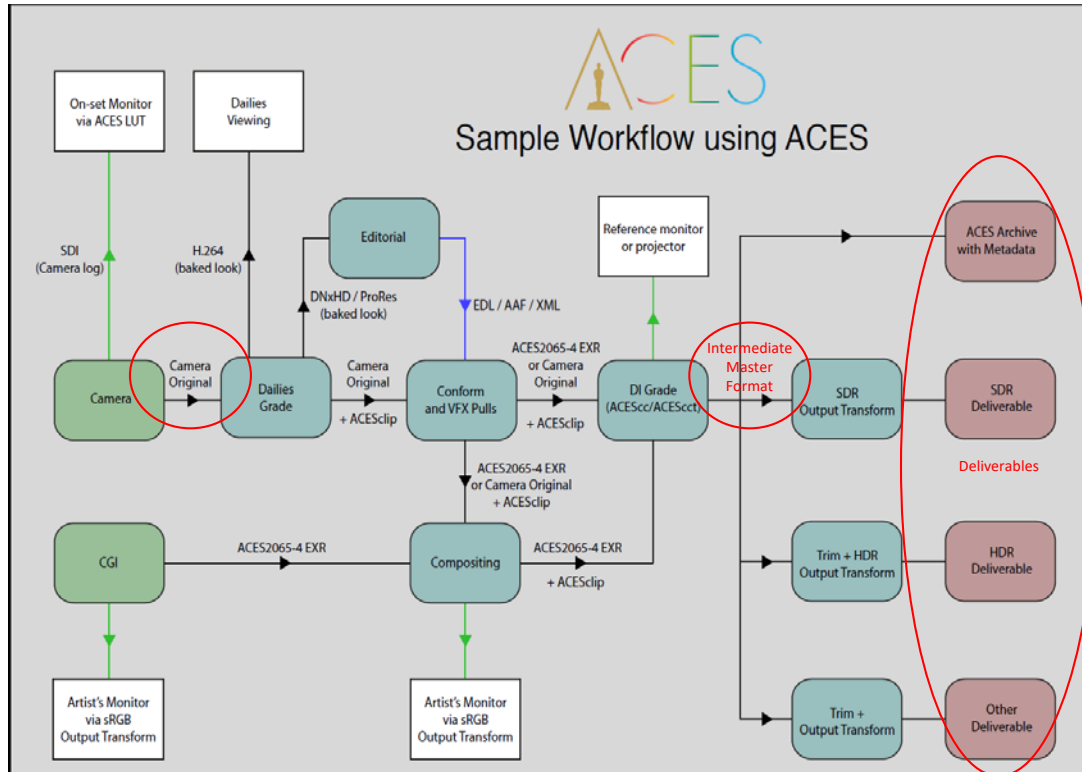


Example: CinePost Production Workflow 2012



Reference: CinePostproduction Presentation at Bundesarchiv 2012

Example: ACES Workflow 2018

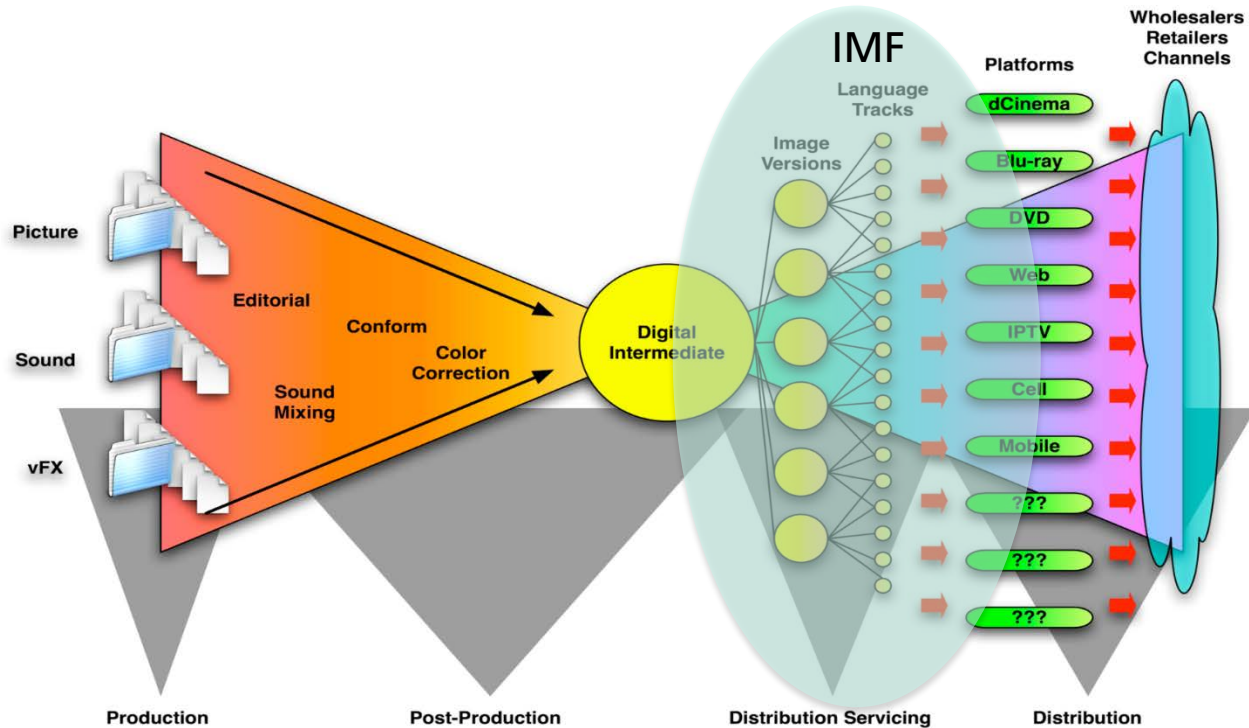


Reference: <https://acescentral.com/aces-primer-glossary-and-quick-start-guides-get-up-and-running-with-aces/1433>

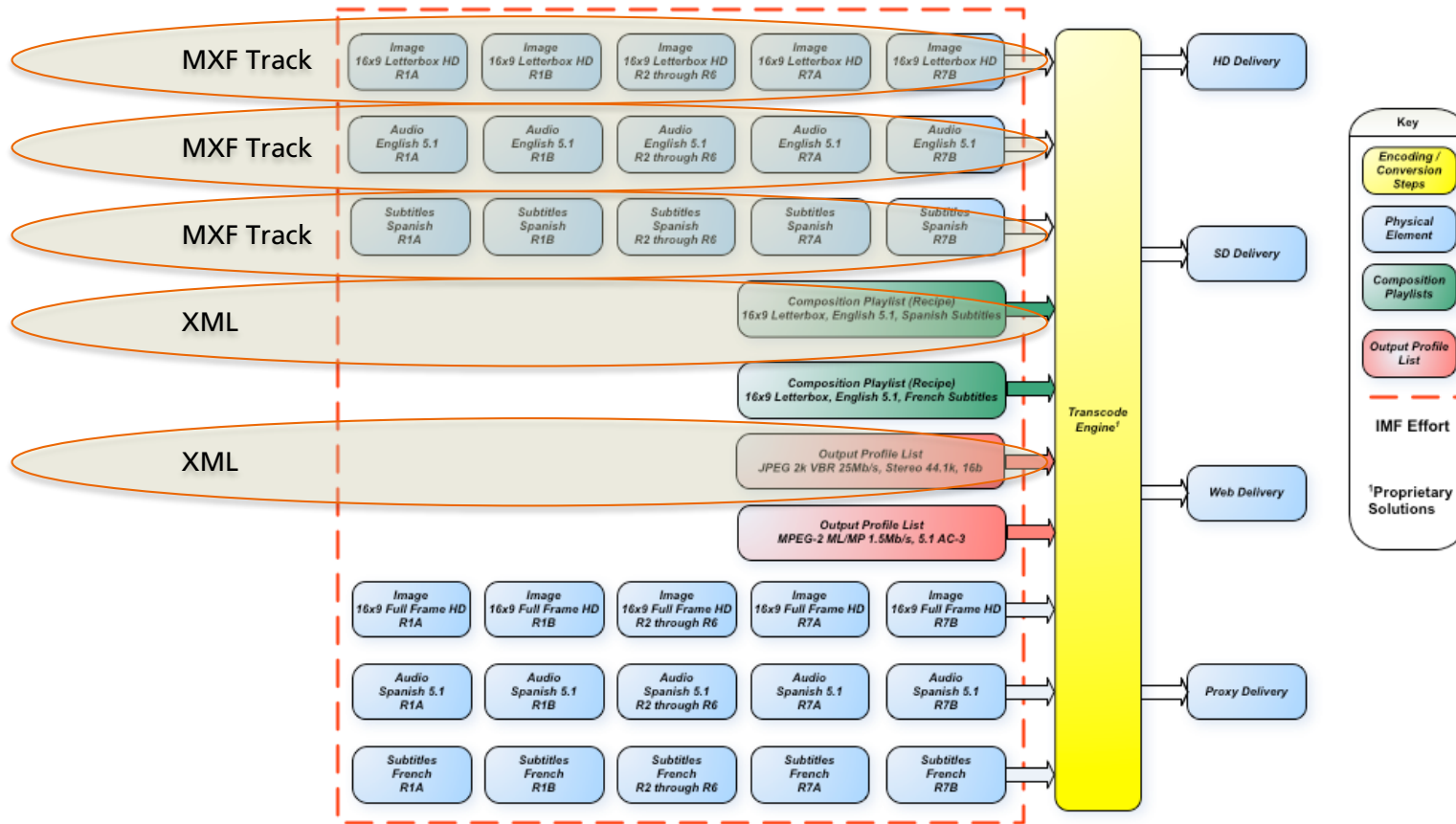
The Digital World - Pandoras Box is open

- New file formats every day
 - MPEG2, H.264, AVC-I, HEVC, H.266, MXF AS-02, DCP, IMF, HEIF
- New consumer devices every day
 - Smartphone, Tablets, Phablets, Ultrabook, HDTV, UHD TV,...
- New movie formats
 - 2D, 3D, 4D, High Frame Rate (HFR), High Dynamic Range (HDR), Extended Dynamic Range (EDR)
 - 2.0, 5.1, 22.1, Atmos, IOSONO, Auro, object based audio,
- New technologies
 - OLED, laser projection, wavefield synthesis, DASH

Example: IMF Workflow since 2012



Reference: <https://www.etcenter.org>



Reference: <https://www.etcenter.org>

IMF Flavours

- Application #1: Uncompressed Images, Audio PCM
 - Not defined
- Application #2: JPEG2000 Broadcast Profile, Audio PCM
 - Most used today for postproduction exchange
- Application #3: MPEG4 Visual Simple Studio Profile, Audio ITU-R BR.1352-2
 - Not used
- Application #4: JPEG2000 Archive Profile, Audio PCM
 - Driven by CST
- Application #5: ACES, Half Float EXR, Audio PCM
 - Driven by AMPAS

IMF

Essence/dynamic Metadata

Image

- MXF Track File

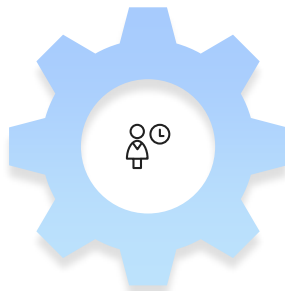
Audio

- MXF Track File

Subtitles

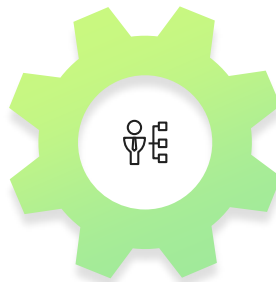
- MXF Track File

Metadata



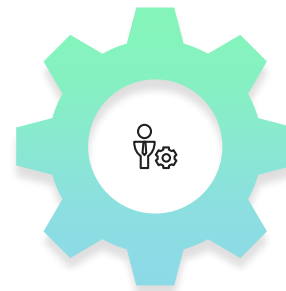
CPL 1

Like an Edit Decision List (EDL) or “recipe” Points to various track files to synchronize and create the proper playback of content



PKL 2

PKL – Packing List
An XML document which lists and describes the files contained therein. It provides a means for authentication as delivered in the package.



OPL 3

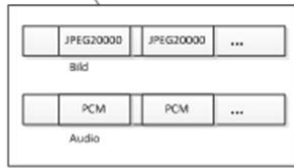
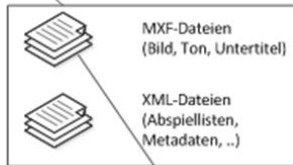
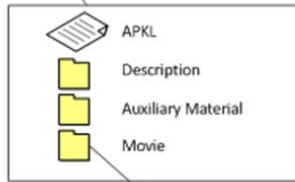
Set of global information or instruction set to specify certain output preferences to a downstream device
Human readable as XML

New working group CEN TC457

- Goal: Defining a long-term archiving standard based on DCP, IMF and package principles for OAIS
- Creation of open source reference code
- Participation of European archives
- Secretary lead by DIN

Format Example AIP

Top Level – Packages: e.g. AIP



- AIP is one complete package:
 - Inventory list APKL
 - Movie (Image and sound) e.g. DCP Play lists, Sub-Title, ..
 - Variants
 - Descriptive meta data
 - Additional elements (Poster, advertising,...)
 - Check sums, ECC, ...

Conclusion

- In the past distribution formats were archived, now archiving starts already at the begin of the workflow, (camera original or at least Master Format)
- In professional production DCP and IMF are the dominant formats (MXF plus XML packages), enriched by additional metadata
- Proprietary formats are gone, standardized or open formats are used today
- In distribution there is a manifold of formats, which will be generated more and more automatically by scripts

Thank You

Siegfried Foessel

Moving Picture Technologies

Fraunhofer IIS

Erlangen, Germany



Martin Hoppenheit | Qualitätssicherung bei der Datenbankarchivierung

Herr Martin Hoppenheit (Landesarchiv NRW, Duisburg) geht in seinem Vortrag zunächst auf relationale Datenbanken an sich und dann das SIARD-Format ein.

Eine Qualitätssicherung ist aufgrund eigener Erfahrungen notwendig, weil das Abgabeformat (Datenbankzugang oder SQL-Dump) in der Regel nicht dem Archivierungsformat (SIARD) entspricht, das häufig erst im Archiv entsteht. Die Wichtigkeit der Qualitätssicherung resultiert auch aus der Verantwortung des Archivs für die Integrität der Daten.

In der Theorie sollte SIARD auf das produktive Datenbankmanagementsystem (DBMS) zugreifen. In der Praxis der Datenbankarchivierung wurde dies oft verweigert, weshalb zunächst ein von der Behörde abgegebener Datenbankdump in das DBMS des Landesarchivs importiert wurde, woraus dann ein SIARD-Container entstand.

Martin Hoppenheit suchte nun nach einem Verfahren, um die Inhalte des SIARD-Containers mit dem gelieferten Dump auf inhaltliche Identität zu prüfen. Aus dem SIARD-Container wurde zunächst die Datenbank erneut in das DBMS des Landesarchivs importiert, so dass eine Repräsentation A vor dem Ingest und eine Repräsentation B danach vergleichbar vorlagen.

Als Vergleichsmittel fiel Hoppenheit zunächst auf das Tool „mysqldbcompare“, das zwei Datenbankinstanzen miteinander abgleicht. Das Tool schied aber aus, weil es beim Vergleich keine Möglichkeit gab, inhaltlich äquivalente, aber nicht identische Werte wie „ä“ und „ä“ als gleichwertig zu definieren.

Ein flexiblerer Weg ergab sich aus dem Vergleich des von der Behörde gelieferten Dumps mit einem neuen Dump, der aus der Repräsentation B erzeugt wurde. Die beiden Textdateien wurden mit dem Linux-Befehl diff gegeneinander abgeglichen.

Herr Hoppenheit betont, dass bei einfachen Datenbanken Tabellendefinitionen und Tabelleninhalte die Bereiche darstellen, die von eigentlichem Interesse seien.

In den verschiedenen Repräsentationen einer Datenbank (DBMS, Dump, SIARD) werden dieselben Werte teils unterschiedlich dargestellt bzw. formatiert. Akzeptable Abweichungen sollen bei der Qualitätssicherung toleriert beziehungsweise normalisiert werden (führende Nullen, Sekundenbruchteile, etc., sofern sie keine Relevanz besitzen). Inakzeptable Abweichungen sowie mögliche Datenverluste hingegen müssen bei der Qualitätssicherung auffallen.

Fragen und Diskussion:

1. Tobias Steinke: Warum wird kein SQL-Dump to SIARD-Converter entwickelt?

A: Das wäre möglich, aber eine derartige Entwicklung war vom Ansatz her nie vorgesehen. Ein Dump ist eine Textdatei, die im proprietären SQL-Dialekt des Herstellers geschrieben ist. Deshalb wäre für jeden Hersteller ein eigenes SIARD-Modul nötig. SIARD setzt für die Vereinheitlichung an der für alle Hersteller identischen ODBC-Schnittstelle an, was den Aufwand verringert.

2. Jörg Filthaut: Warum erfolgt keine Übernahme von CSV-Dateien?

A: Neben dem SQL-Dump wurden von der Behörde zusätzlich PDF-Exporte abgegeben. Eine Übernahme von CSV-Exporten wäre auch machbar gewesen, aber SIARD hatte den Vorteil, dass nicht nur die Tabellen, sondern auch Metadaten über die einzelnen Inhalte mitgeliefert werden.

3. Lambert Kansy: Migrations- und Konvertierungsschritte sollten gering gehalten werden. Man sollte daher versuchen, an die laufenden Systeme heranzukommen.

A: Das wäre wünschenswert, jedoch ist es in der Praxis meist so, dass kein Zugriff auf die Systeme erfolgen kann.

Jörg Filthaut: Das neue Archivgesetz in Thüringen verpflichtet die Behörden, Einsicht in die zur Aussonderung anstehenden Unterlagen zu gewähren.

4. Kai Naumann: Welche Übernahmefrequenz ist bei der Übernahme von Datenbanken beim angesprochenen Verfahren zu erwarten?

A: Man ist optimistisch, dass sich das Verfahren als praktikabel herausstellen wird. Grundsätzlich muss bei Datenbanken immer geschaut werden, ob die extrahierte Datenbank abseits des beinhaltenden Systems verständlich ist. Dies muss gemeinsam mit der abgebenden Stelle beurteilt werden.

5. Kai Naumann: Welche Vorteile hat die Präsentationskomponente des Database Preservation Toolkit?

Martin Hoppenheit: Das Database Visualization Toolkit (dbvtk) besitzt eine webbasierte Oberfläche, die einen Einblick in die SIARD-Container gewährt. Es ist eine Alternative zur SIARD-GUI, mehr aber nicht.

Zusatzinformation: Nachdem die Software SIARD-Suite lange Zeit unter einer proprietären Lizenz stand, ist SIARD inzwischen vollständig quelloffen.

Von der Theorie zur Praxis: Bestandserhaltung digitaler Unterlagen
Koblenz, 07./08.11.2018

Qualitätssicherung bei der Datenbankarchivierung

Martin Hoppenheit, Landesarchiv Nordrhein-Westfalen



Grundlagen

Relationale Datenbank: Menge von Tabellen

SIARD: Abbildung von Struktur, Inhalt und Metadaten einer Datenbank in softwareunabhängiger* Form



* d.h. unabhängig von einem konkreten Datenbanksystem wie MySQL oder Oracle

Warum überhaupt Qualitätssicherung?

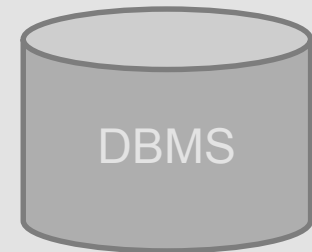
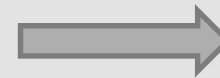
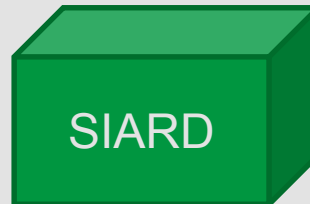
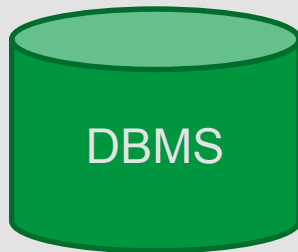
Abgabeformat \neq Archivformat



Verantwortung für die Integrität der Daten

Ein Beispiel aus dem echten Leben

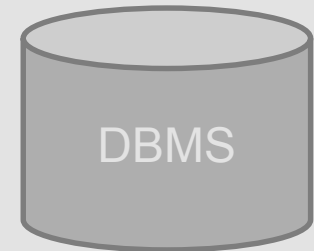
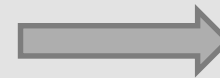
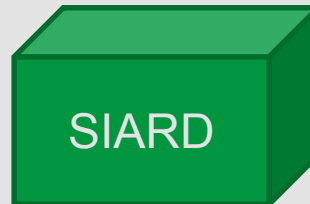
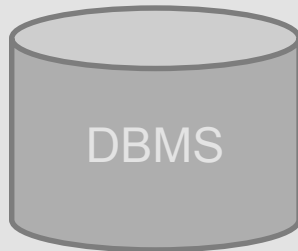
SQL-Dump



- Abgabe: SQL-Dump (MySQL)
- Import des Dumps in DBMS
- Export im SIARD-Format

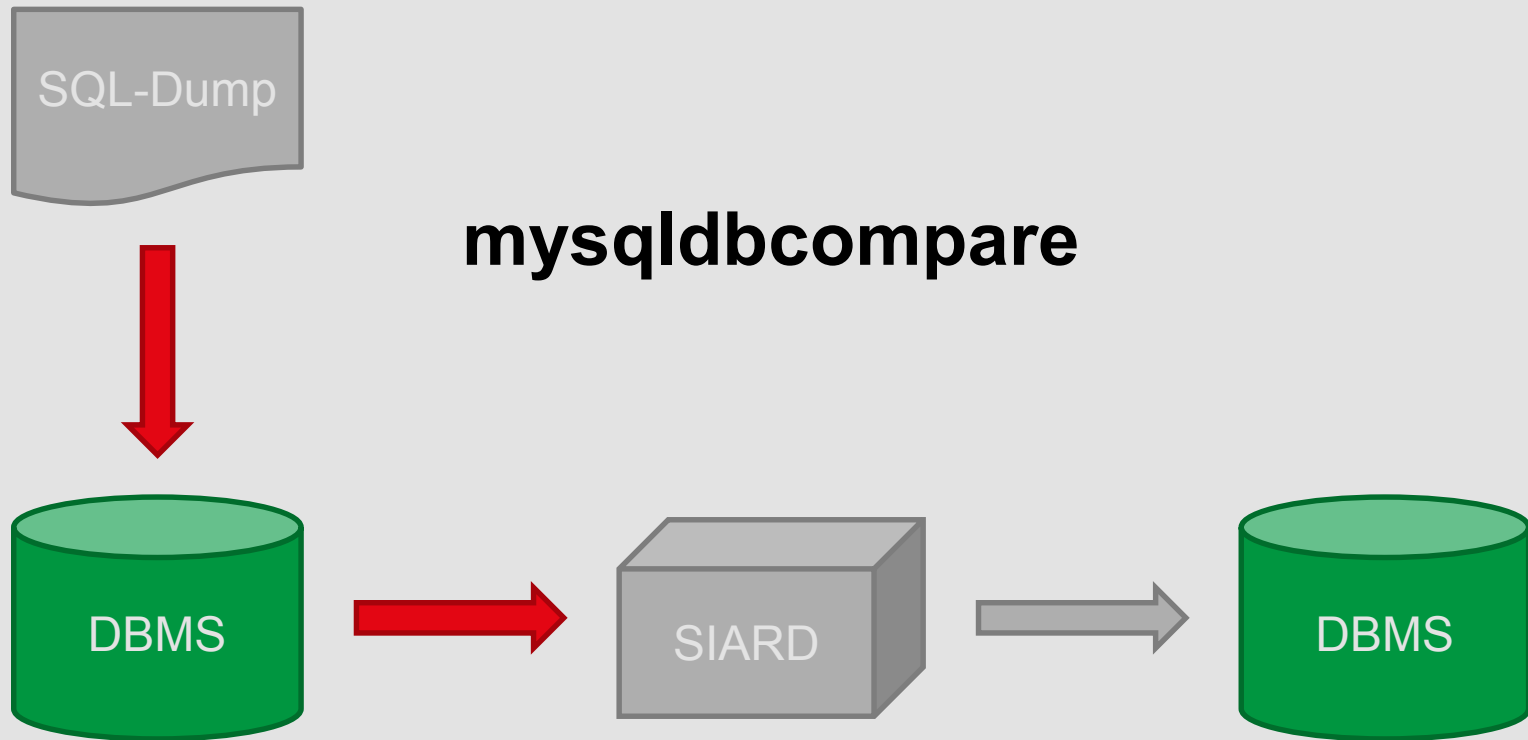
Vergleich von SQL-Dump und SIARD-Datei

SQL-Dump

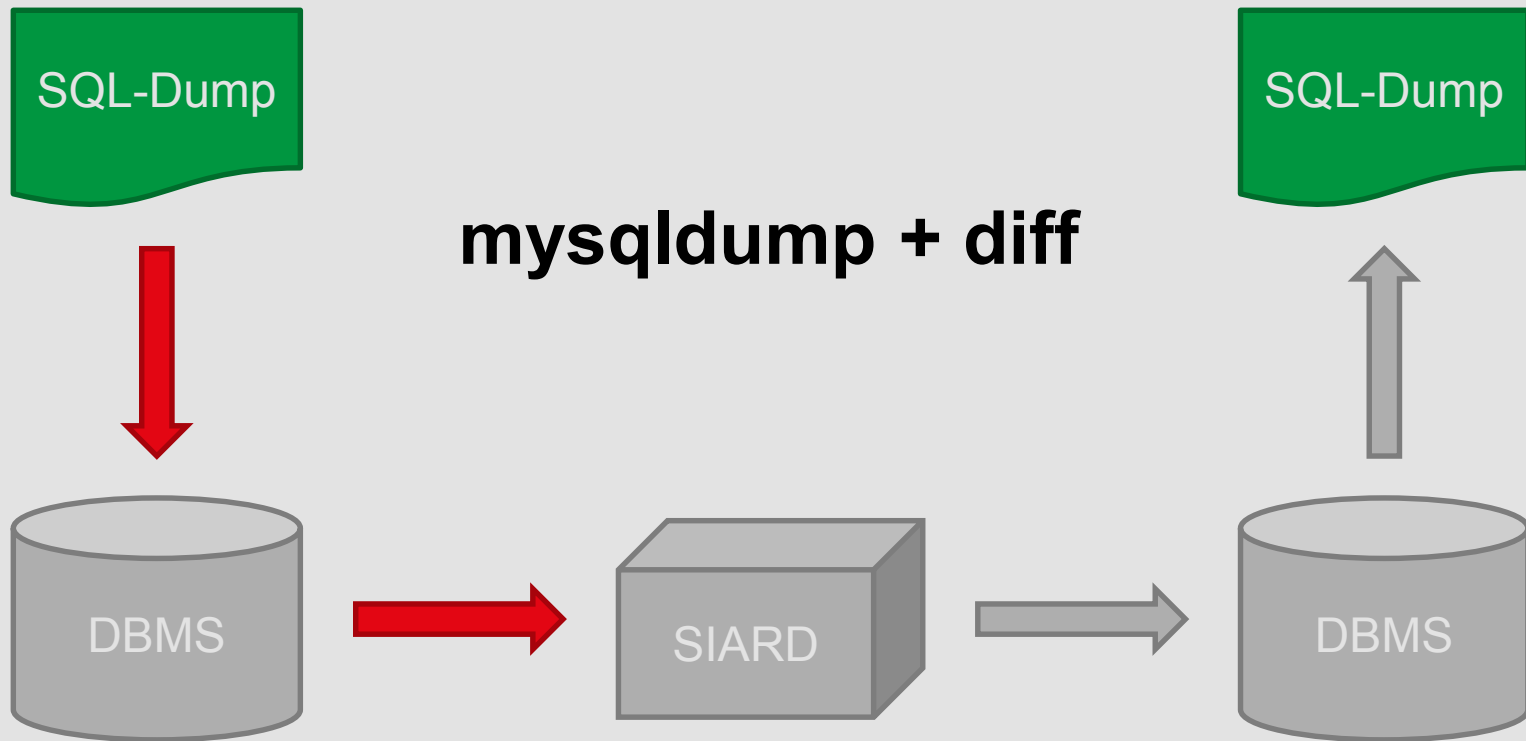


keine fertigen Tools ...

Vergleich der Datenbanken



Vergleich der Datenbankinhalte



Beispiel: SQL-Dump

```
CREATE TABLE `foerderbereich` (  
  `foerderbereich` int(8) NOT NULL,  
  `bezeichnung` varchar(255),  
  PRIMARY KEY (`foerderbereich`)  
);
```

Tabellendefinition



```
INSERT INTO `foerderbereich` VALUES(0, 'Allgemein');  
INSERT INTO `foerderbereich` VALUES(2, 'Schulinfrastruktur');  
INSERT INTO `foerderbereich` VALUES(4, 'Weiterbildung');  
INSERT INTO `foerderbereich` VALUES(8, 'L&auml;ndliche Infrastruktur');  
INSERT INTO `foerderbereich` VALUES(10, 'Informationstechnologie');  
INSERT INTO `foerderbereich` VALUES(42, 'Pommesbuden');  
INSERT INTO `foerderbereich` VALUES(99, 'Sonstiges');
```

Tabelleninhalte



✓ Akzeptable Abweichungen

00012 ≈ 12

09:17:08.000000 ≈ 09:17:08

ä ≈ \u00e4 ≈ ä

0000-00-00 ≈ NULL

→ **Sollen bei der QS ignoriert werden. (Normalisierung)**

X Inakzeptable Abweichungen (Datenverlust)

0 = 0

1 = 1

2 ≠ 1

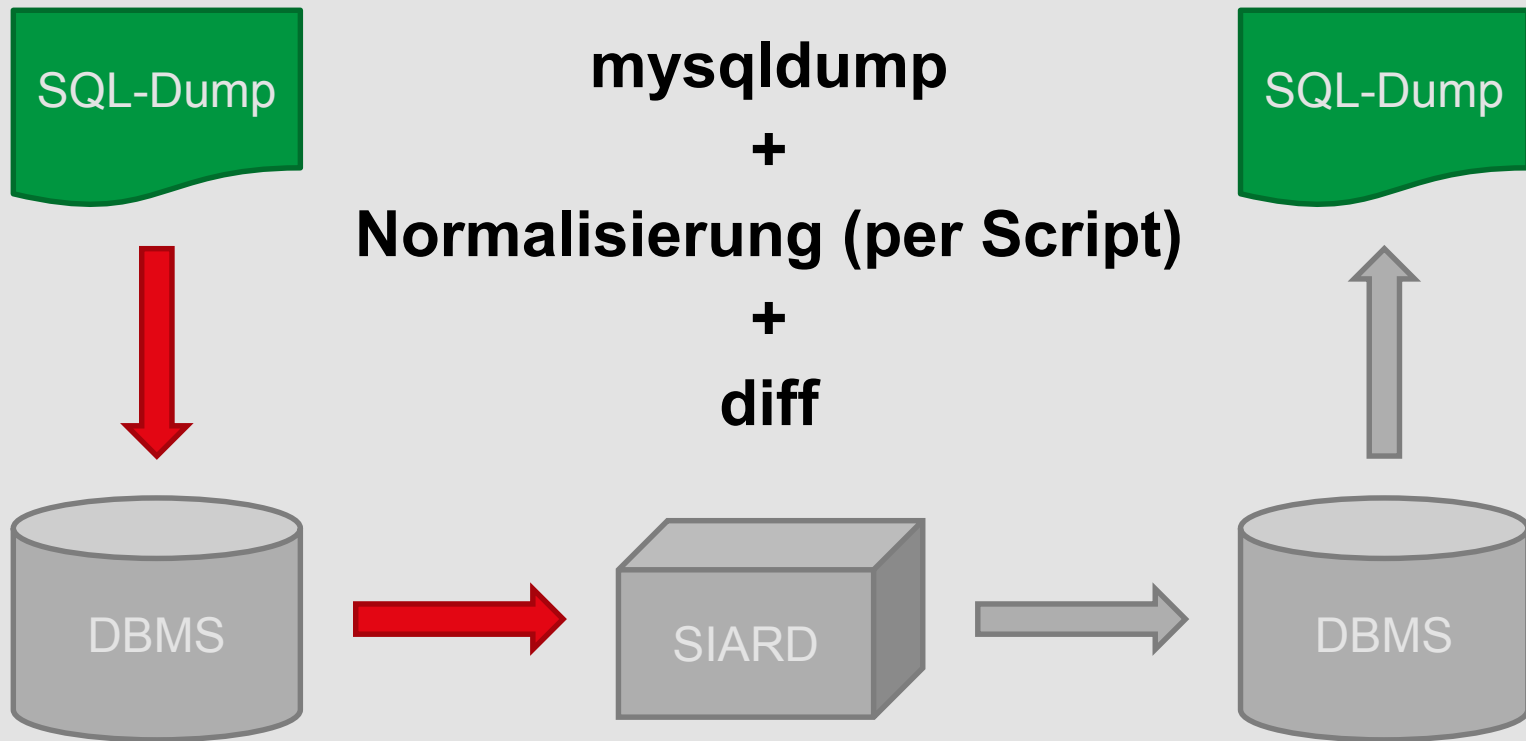
3 ≠ 1

4 ≠ 1

...

→ Müssen bei der QS auffallen!

Vergleich der Datenbankinhalte



Zusammenfassung

- 1. Welche Instanzen vergleichen?
Dump, Datenbank, SIARD?**
- 2. Welche Eigenschaften vergleichen?
Struktur, Datentypen, Inhalte?**
- 3. Welche Werte sind „gleich“?**

Fazit

Nutzen?

- **Datenverlust entdeckt und vermieden.**
- **SIARD Suite und dbptk verbessert.**

Aufwand?

- **Umsetzung relativ einfach.**
- **Eigenheiten von Datenbank und DBMS erfordern Vorkenntnisse oder Einarbeitung.**

→ **Es lohnt sich!**

Danke für die Aufmerksamkeit! Fragen? Anregungen?

Martin Hoppenheit, LAV NRW F4
martin.hoppenheit@lav.nrw.de



Andreas Trautheim-Hofmann | Erhaltung von 3D Ingenieurdokumenten und Produktdaten

Andreas Trautheim-Hofmann (PROSTEP AG) stellt in seinem Vortrag die Standardisierung von Abläufen als Grundlage für die Produktion umfangreicher und komplexer Produkte dar. Die PROSTEP AG ist Dienstleister für das Product Lifecycle Management mit Schwerpunkt auf der Fertigungsindustrie.

Das Anwachsen des Datenumfangs und der Migrationsschritte erschwert die langfristige Sicherung. Daten-Content muss über die Jahre so aufbewahrt werden, dass die Inhalte von damals auch heute und in Zukunft noch lesbar und integer sind. Durch den Federal Code of Civil Procedures wurden für die USA Regeln geschaffen, nach denen elektronische Informationen zusätzlich zu papiergebundenen Informationen zu Beweis Zwecken herangezogen werden dürfen (sogenannte eDiscovery Verfahren). Daraus resultiert vor allem bei kostspieligen Produkthaftungsprozessen die gesteigerte Wichtigkeit im Hinblick auf die digitale Langzeiterhaltung.

Mögliche Lösungsansätze bieten sich Trautheim-Hofmann zufolge in der Adaption etablierter Archivierungsmethoden und Prozessen/Lösungen für 3D- und PDM-Daten. Maßgeblich seien dafür das OAIS-Modell (ISO14721), das Prinzip der digitalen Signierung sowie die Anwendung diverser international anerkannter Standards.

Problematisch sieht er, dass LZA-PDF-Container, wenn sie neben dem reinen Viewing einen funktionalen Nutzerkomfort bieten sollen, wie z.B. Hyperlinks zwischen den Dokumentbestandteilen im Container oder dem Suchen in den 3D-Inhalten, in der Regel nicht PDF/A-fähig sind.

Diskussion und Fragen:

1. Ulrike Gutzmann: Die Volkswagen AG führt solche 3D-Daten nicht in ihrem Archiv. Das Vorgelegte bedeutet für sie nicht Langzeitarchivierung sondern eher Aufbewahrung, da die Daten nicht wirklich archivfähig sind. Hier besteht Diskussionsbedarf, weil nicht archivfähiges dauerhaft gesichert werden soll.
2. Christian Keitel: Im Ansatz von Herrn Trautheim-Hofmann besteht ein anderes Verständnis von Archivierung, wobei das grundsätzliche Denken jedoch sehr ähnlich ist. Im Kern geht es um die dauerhafte, sichernde Aufbewahrung zu Nachweiszwecken während und nach der effektiven operativen Nutzung.

A: Den Produzenten ist das grundsätzliche Problem bekannt. Das Handeln geschieht quasi "aus der Not" heraus. Da eine PDF/A-Archivierung jedoch nicht möglich ist, erfolgt quasi eine dauerhafte Aufbewahrung auf Basis PDF/E.

3. Hr. Dr. Michael Puchta: Existiert bei einzelnen Kunden eine Hochrechnung, wie umfangreich die kumulierte Datenmenge sein würde?

A: Bisher besteht eine solche Hochrechnung nicht. Alte Datenbestände werden momentan vielerorts sukzessive durch den geschilderten Prozess bearbeitet. Zudem besteht ein juristisches lizenzrechtliches Problem, da die Daten in vielen Jahren womöglich durch externe Dienstleister migriert und konvertiert werden müssen.

4. Fr. Eszter Kiss: Bisher haben sich keine Begrifflichkeiten auf diesem Gebiet der Archivierung etabliert, weshalb es nach wie vor unterschiedliche Interpretationen der Begriffe „Archivierung“ und „Langzeitspeicherung“ gibt.

A: Die Begrifflichkeiten-Diskussion hält bislang an. Irgendwann erfolgte aus pragmatischen Gründen die Festlegung von Arbeitsbegriffen. Es wurde versucht, den Begriff „3D-Dokument“ als juristischen Begriff zu etablieren. Gewünscht wird ein ausgedehnteres Engagement für die Anpassung der juristischen Grundsätze!

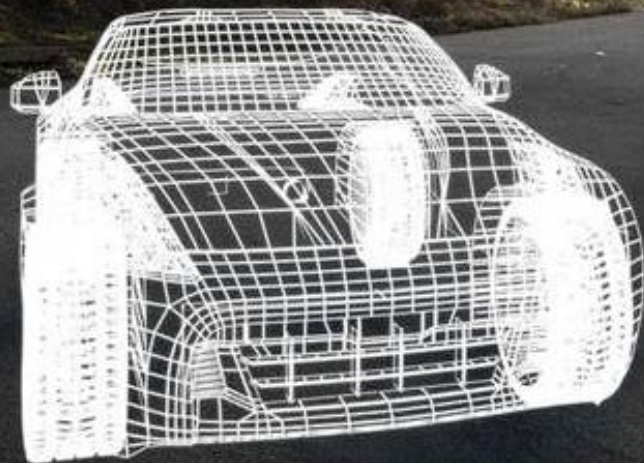


LONG TERM ARCHIVING

of 3D engineering documents and product data

Andreas Trautheim-Hofmann

KLA-Ausschuss „Digitale Archive“, Koblenz, 08.11.2018



Andreas Trautheim-Hofmann

PLM Principal Consultant & Project Manager



PLM Strategy & Processes

Dipl.-Ing. (Technical University)

- Mechanical engineering, construction engineering
- Computer science / information processing

28 years of experience in the area of PLM & Process Consulting

Competences

- Project Management and coordination (classic & agil)
- Strategic and methodical PLM consulting especially within the engineering of industrial customers and boards
- Business and IT process consulting
Product development processes, Engineering Change Management
- Data modelling
- Technical concept papers and specifications
- Partner / system selection and implementation
- Development of engineering methods
- Change processing & management
- Moderation, coaching, user training
- Support, coordination implementation, test, operational roll-out

Selected Projects

- Engineering Change Management Processes, multi-disciplinary and company-wide; Automotive Industry; Coordinator, Consultant VDA4965 (ECM)
- Functional Design & Integration at BMW; Consolidation methods & system landscape; Optimizing Support services (maintenance, projects) for Business & IT (organizational & strategical consulting)
- PEP-PDM Audit at BMW
- Roadmap for the selection & introduction of PLM functions for „Digital Master“ (aerospace industry, plants)
- PDM / PLM Strategy development & benchmarking, architecture planning & integration projects (all sectors)
- Drawingfree (Release-) Processes, Documentation & Longterm-Archiving (LZA) of digital 3D and PLM data (Airbus, BMW, Daimler, VW)
- Consulting & Coordination PDM User Group, LOTAR, LZA (VDA 4958), „Zeichnungslose Prozesse“ (VDA4953-2)

Languages

German (native)
Russian (basics)

English (business fluent)

PROSTEP - 100% PLM

Who we are – What we offer

PROSTEP is

- an expert for PLM processes and solutions
- the leading vendor-independent PLM consultant for the discrete manufacturing industry
- 25 years of industry background – more than 200 PLM experts - 200 customers - 9 locations
- Shareholders:



- Partner of leading PLM vendors:



Selected reference customers



PROSTEP offers

- PLM strategy consulting
- PLM concepts & solution architecture
- PLM implementation
- own products to build PLM integration & collaboration solutions



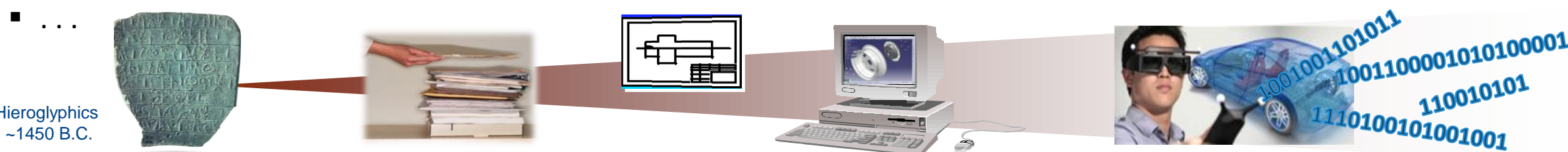
PROSTEP - 100% PLM

Beratungs- und Lösungsangebot

Strategie	Konzept & Lösungsarchitektur	Implementierung IT & Prozess	
PLM Einführungsstrategie	PLM-Architektur & -Prozesse Benchmark & ROI-Analyse Bebauungsplanung PLM-Prozess- & -Systemintegration	PLM Migration & Integration OPENPDM® OPENDESC.com® OPENDXM® GLOBALX	– PLM Systemauswahl – PLM Einführung – PLM Prozessoptimierung
	PLM für Digitale Transformation	PLM für IoT/I 4.0 Lösungen Digital Master / Digital Twin	
PLM für Kollaboration		Model Based Enterprise 3D Master / Systems-Engineering	3D PDF basierte Lösungen ZLP, Maintenance & Repair, LZA, TDP, 3D Master 3D PDF
	Unternehmensübergreifendes PLM PLM für Merger & Acquisitions PLM für Joint Ventures Partnerintegration	PLM Kollaboration Automatisierte PLM-Regelversorgung PDM & CAx Datenaustausch OPENPDM® OPENDESC.com® OPENDXM® GLOBALX	– Merger & Acquisitions – Joint Venture – Konsortialprojekte

Our activities on standardization (excerpt) for about 20 years related to LTA / DFP

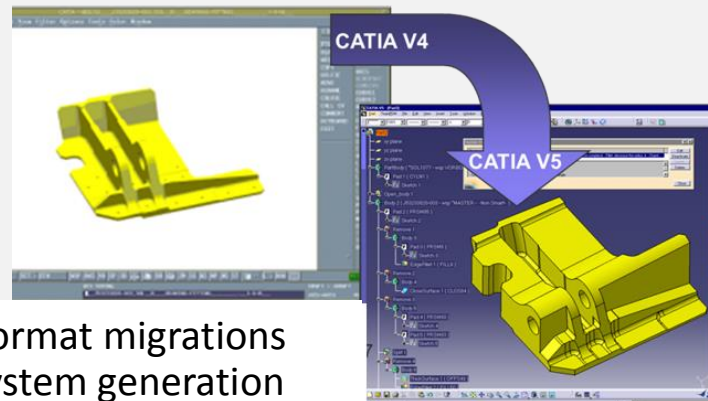
- EN/NAS 9300 series – Long Term Archiving and Retrieval (LOTAR)
- ISO10303 series – STEP (e.g., AP242, 212, 214, 239, 209 ...)
- ISO 14306 – Industrial automation systems and integration; JT file format specification for 3D visualization
- ISO 16792 – Technical product documentation, Digital product definition data practices
- ISO 32000 – PDF (ISO19005 PDF/A; ISO15930 PDF/X; ISO24517 PDF/E; ISO14739 PRC; ECMA363 U3D)
- VDA 4953 / 4953-2 – Simplified Drawings / Drawing Free Processes
- VDA 4955 / SASIG PDQ – Product Data Quality
- VDA 4956 / SASIG – Product Data Exchange; Assembly Data Exchange (incl. incremental representation)
- VDA 4958 – Long Term Archiving of digital Product Data (which are not based on technical drawings)
- VDA / SASIG 3D Annotated Model Standard („CPV“ Collaborative Product Visualization; „DEV“ Digital Engineering Viewing; incl. Guidelines for combining 3D models and 2D CAD documentation)



Hieroglyphics
~1450 B.C.

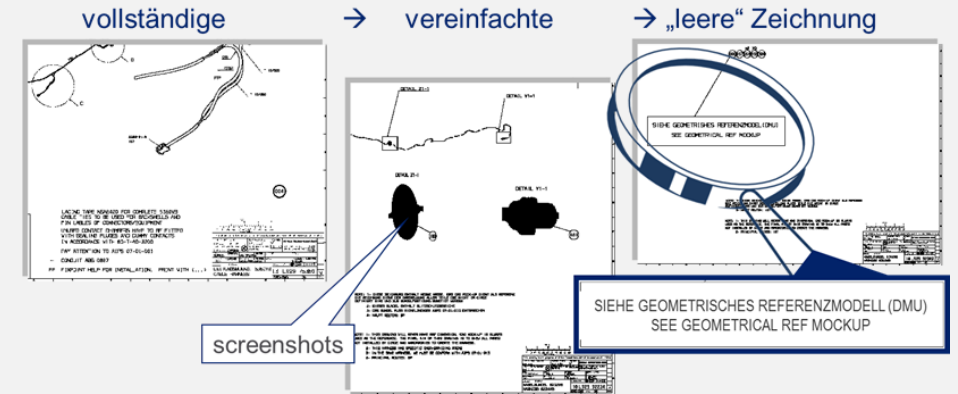
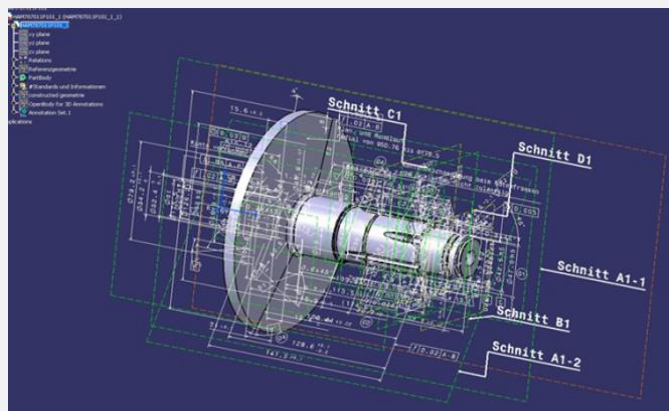
State of the art Application strategies in a dead end?

long term not applicable



native (CAx) formats or format migrations may fail in a new (CAx) system generation (3...10 years), wrong or no result vs. product life cycle 10 ... >60 years

huge effort & expensive
product definition/
(re)presentation
of PMI's, meta data,
parameters etc.
only within 3D ?



- 3D technology (CAx) is state of the art
- product data will be used further more in an electronical way (re-use)
- technical drawings contain no longer "all" information

careless ... nonsensical

... up to
3-dimensional modeled
(2D) title block, tables &
free texts

nonsensical

Demands for action: Organization and Processes

Legal regulations require changes in technical processes



Legal Regulations

Liability for Documentation and Retention of relevant product data, e.g.,

for product liability (10 years) /
manufacturer's liability (+30 years)

permission, supply of services,
operating companies etc.

within the entire partner network



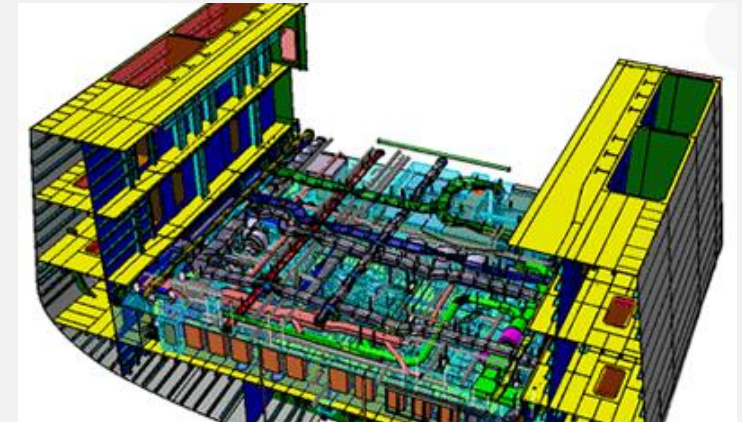
FRCP

Extensions of the „Federal Rules of Civil Procedures“ (FRCP)

since 2006 also applicable to secure
evidence digitally

covering „electronic discovery“
processes (US market)

requirements for the availability of
electronic information in timely
manner and appropriate forms



Product Development

... it includes:

product documentation of design and
added value partners

the electronic representations of the
entire technical information like
2D/3D geometry, CAE/CAT data,
metadata, schemata, product
structures, configuration parameters,
electric/ electronic/ (control)software

...

Key Approaches & Success Factors

to preserve non-drawing based, digital product data

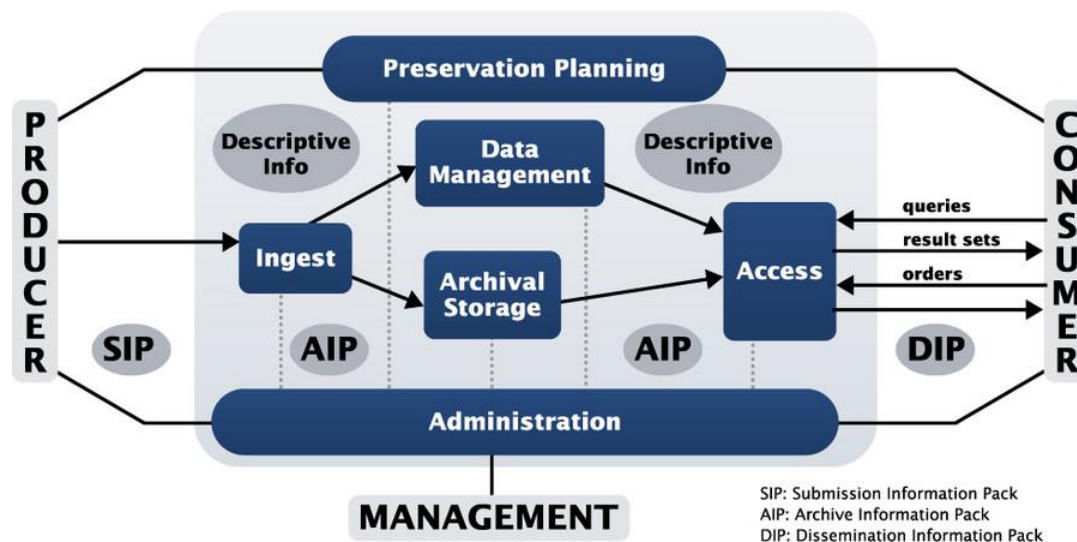
Adaption of established / robust document based (2D) archiving methods & processes and proven solutions of the 3D & PDM data exchange

Digital Signature o.e. acc. German Signature laws to preserve the unchangeability of the (data) content / intent

ISO14721 (OAIS)

adaption as functional framework of the architecture plus specific validation-verification of CAx & PDM information

International acknowledged standards to map reference data and descriptions into appropriate (re)presentation formats enabling a long term preservation without loss of semantics and avoid conflicts with (SW) licenses or obsolescence of supporting vendor companies



e.g.
 PDF - ISO 32000
 STEP - ISO 10303
 JT - ISO 14306



LONg Term Archiving and Retrieval an international consortium of Aerospace manufacturers (since 2002)

PROSTEP
is the project coordinator and
consultor of LOTAR

- LOTAR is an international consortium of Aerospace manufacturers
- Prime objective is creation and deployment of the **EN/NAS 9300 series of standards** for long-term archiving and retrieval of digital data, based on standardized approaches and solutions.
- Integration of LOTAR requirements in software tools ensured by close cooperation with:

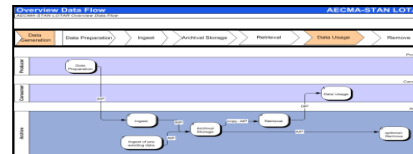
CAX Implementor Forum (CAX-IF):

- Facilitated by AFNeT, PDES, Inc. and prostep ivip
- Consists of CAD, CAE, STEP Translator & Validation Tool vendors
- Supports STEP (ISO10303) AP203, AP209, AP214, AP242

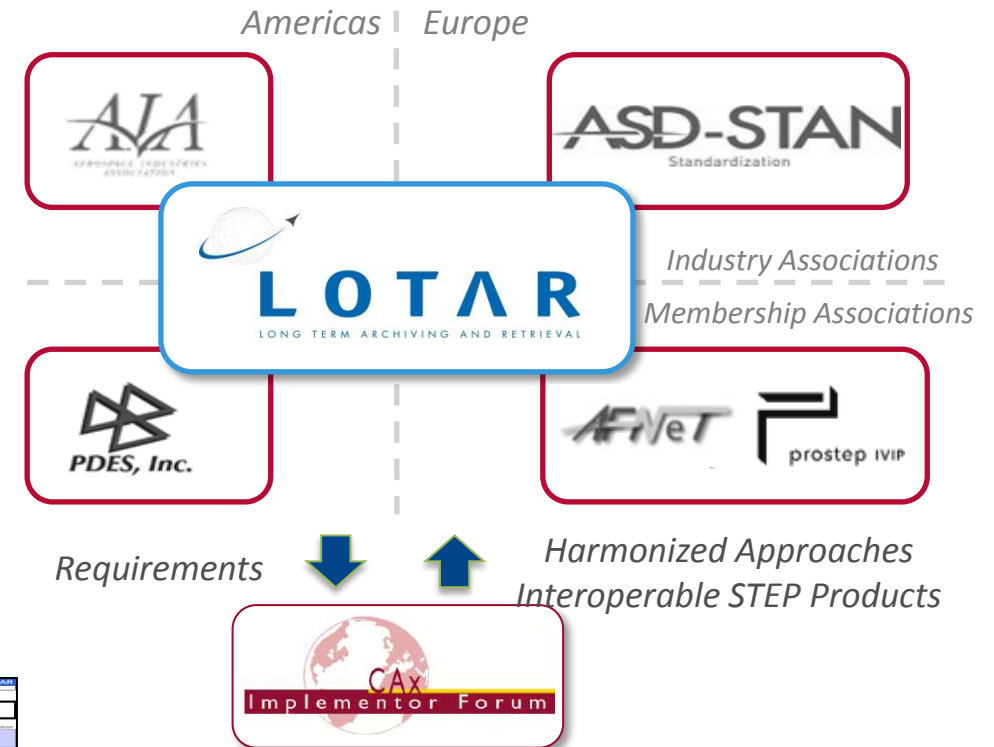
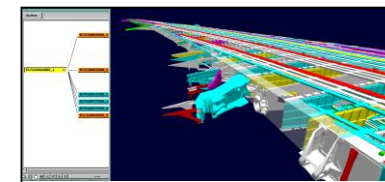
PDM Implementor Forum (PDM-IF):

- Facilitated by AFNeT and prostep ivip
- Consists of PDM and STEP Translator vendors
- Supports STEP AP242 BO Model XML

based on
LTA Reference Processes



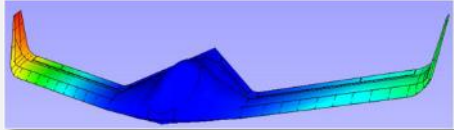
relevant engineering
elements for retention
and re-use along the
entire product life cycle



www.lotar-international.org

EN = European Norm (Standard)
NAS = National Aerospace Standard
CAX = Computer Aided "x" (Design, Engineering...)
PDM = Product Data Management

Seven LOTAR Working Groups at present



Engineering Analysis and Simulation

EN/NAS 9300-6xx series
ISO STEP AP209 ed2

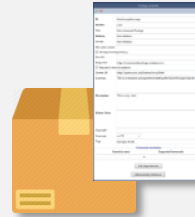
2014 launch



Wiring Harness

EN/NAS 9300-4xx series
STEP AP242 ed2

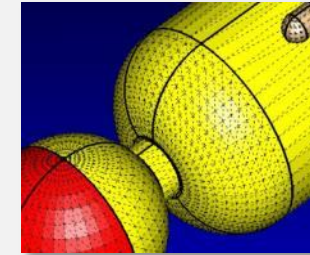
2012 launch



Meta Data for Archive Packages

EN/NAS 9300-21
STEP AP239 ed3
STEP AP 242 ed2

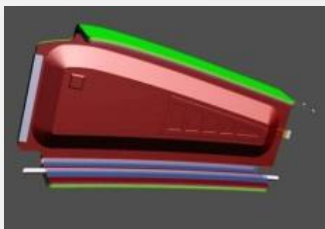
2012 launch



3D Visualization

Requirements and Compliance Documents

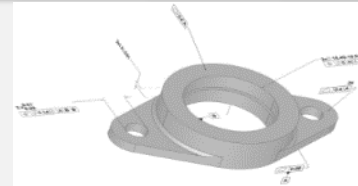
2012 launch



Composites and Advanced Manufacturing

EN/NAS 9300-3xx series
STEP AP203 ed2
STEP AP242 ed1

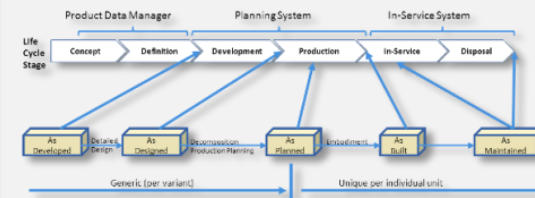
2009 launch



Mechanical 3D CAD with Product and Manufacturing Information (PMI)

EN/NAS 9300-1xx series
STEP AP203 ed2
STEP AP214 ed3
STEP AP242 ed1

2004 launch



Product Data Management (PDM)

EN/NAS 9300-2xx series
STEP AP239
STEP AP242 ed1

2004 launch

Long term archiving (LTA) of digital product data which are not based on technical drawings

VDA4958-1: Requirements and general recommendations

for digital product data retention

VDA4958-2: LTA Reference Process

to design the processes

(data preparation, ingest, archive, access and reprocess the archived data)

VDA4958-3: Relevant LTA Data and Models

defines the minimum of data/documents to be archived

(CAD, product structure etc.) and mapping onto standard representations/formats

VDA4958-4: Certification

recommendations for the verification of reliable LTA workflows

Core team: **DAIMLERCHRYSLER**
VOLKSWAGEN AG

protics
technical information
consulting & support
A DaimlerChrysler Company

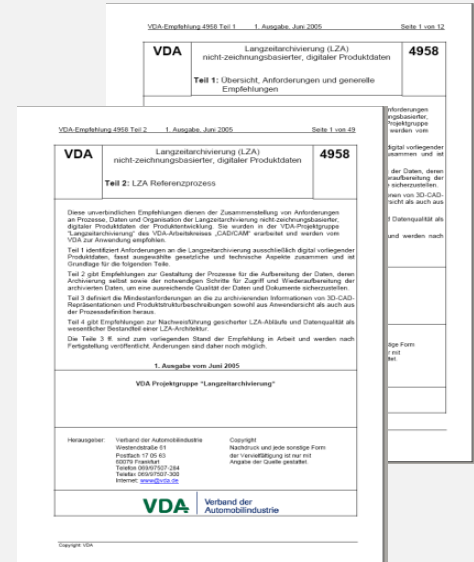
DELPHI
KARMANN

BOSCH

Project coordination / consulting by

PROSTEP
integrate the future

Project data: 2003-2007, 37 Workshops (most 2 days), 3 Events for presentation of results, 1 special for harness



Cooperation with:



Project LOTAR

Long Term Archiving and Retrieval of Digital Technical Product Data

EN9300-xxx



Holistic Consideration - Requirements Cluster

Technical requirements of systems

of authoring systems, conversion tools, management systems, interfaces, retention, process control

(VDA4853-2, ISO14721 OAIS ...)

Safeguarding Quality

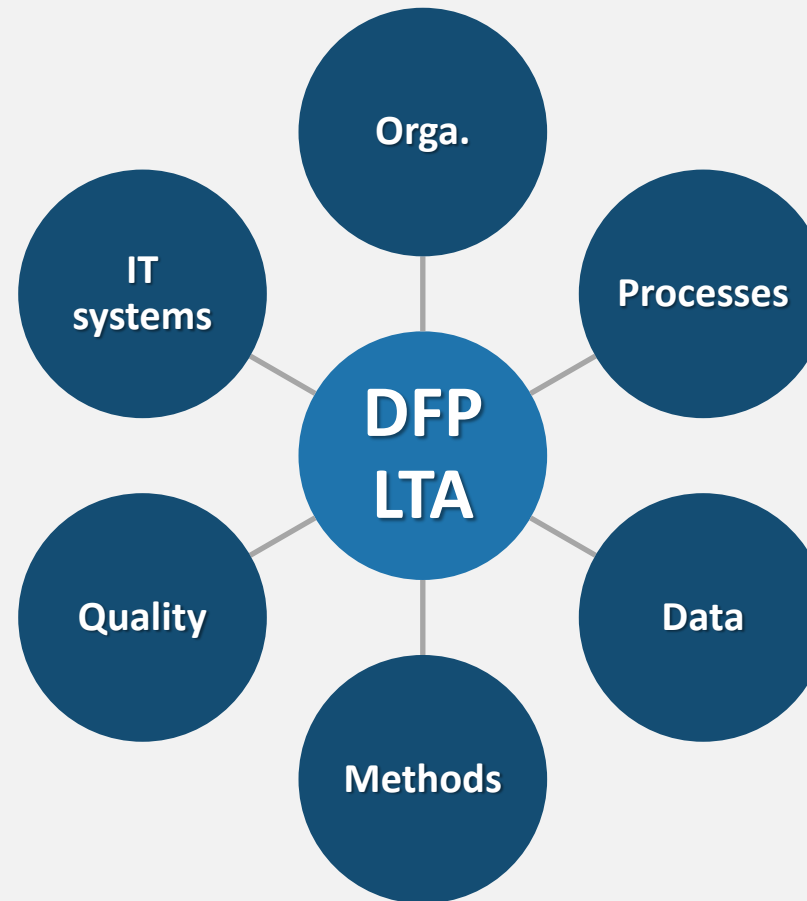
Data validation, verification, safeguarding of processes, certification

(VDA4958-4, VDA4955, DIN EN ISO9001, PK-DML, EN/NAS9300, ...)

Methodical requirements

Functional mapping of content, representational and presentational requirements, (schemata, formats)

(ISO16792, VDA4953-2, ISO10303-242, EN/NAS 9300 ...)



General requirements

strategic requirements, legal aspects, verification, respites, organization, rules & (access) rights, responsibilities

(ProdHaftG, VDA4958-1, ISO14721 OAIS ...)

Processual requirements

Use cases, core / domain processes, support processes ...

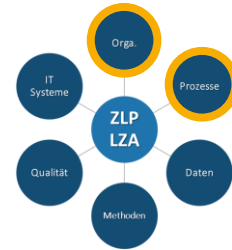
(EN/NAS 9300, VDA4958-2, VDA4953-2 ...)

Data requirements

(user) data content, management information, process and control information ...

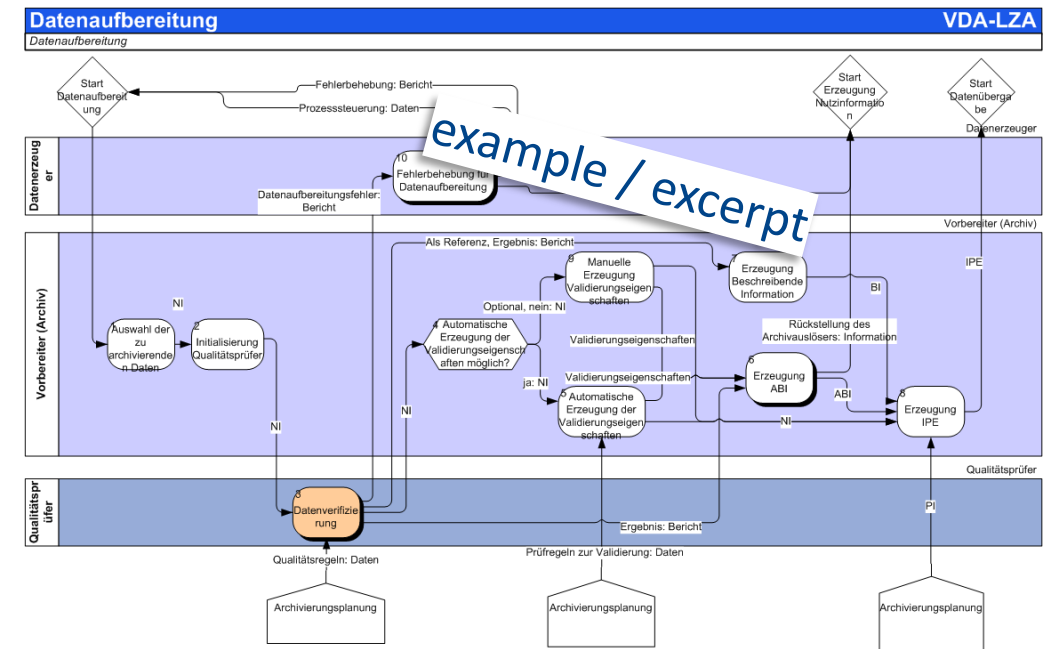
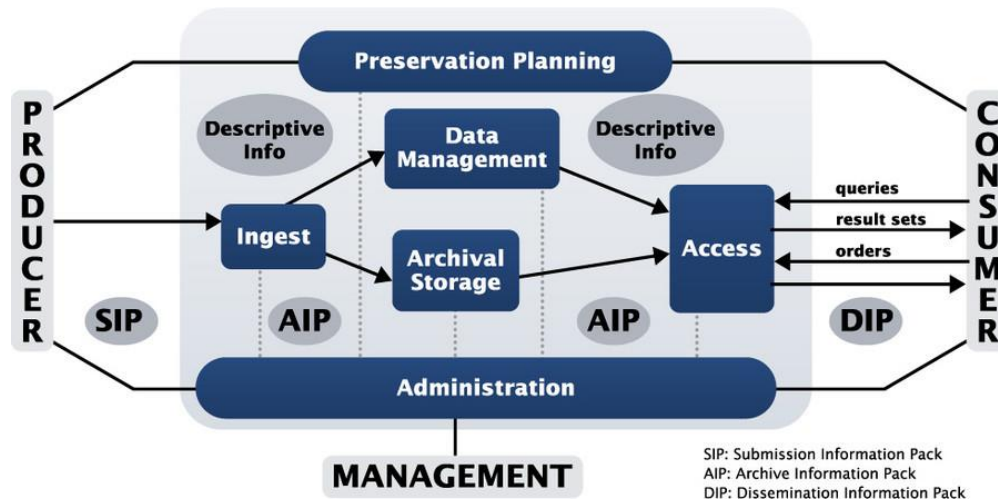
(EN/NAS 9300, VDA4948-3, VDA4953-2, ISO10303-242 ...)

ISO 14721 OAIS (Open Archive Information System)



Reference Model as one of the fundamental approaches of LOTAR / VDA ZLP

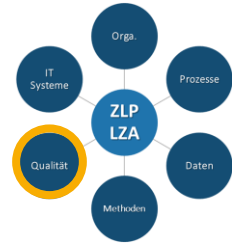
- developed by Aerospace & Defense Industry's "Consultative Committee for Space Data Systems"
- Recommended Practices for OAIS available at www.ccsds.org



- functional framework is extended for Aerospace (LOTAR) & Automotive (VDA LZA) archiving standards enabling the application specific context of 3D and PLM data

Quality Management - Validation & Verification

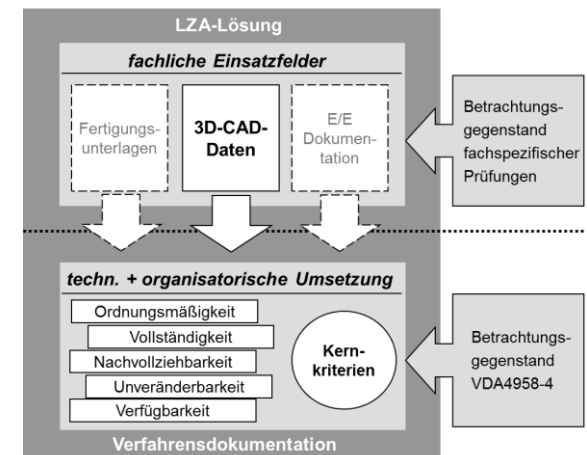
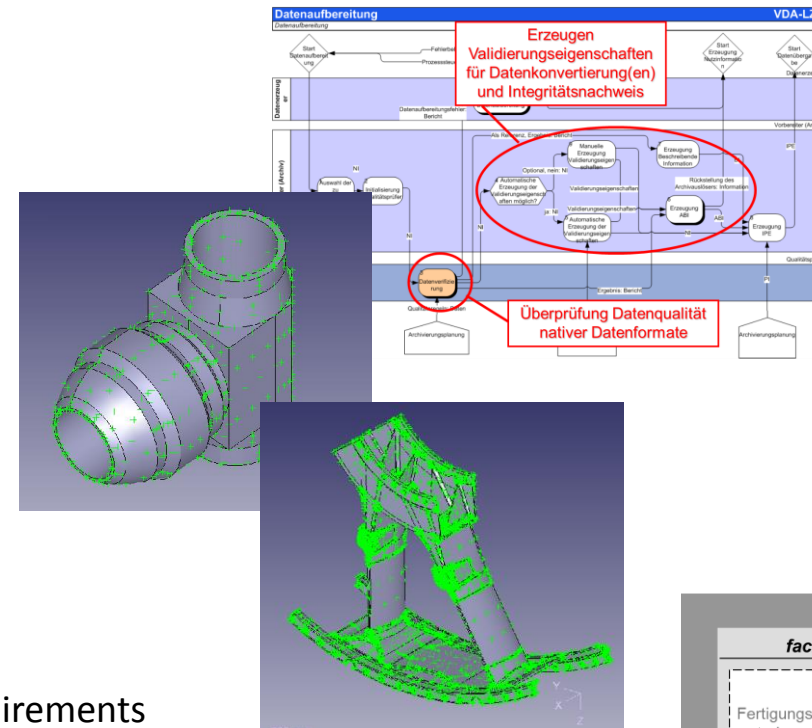
acc. VDA 4958 and VDA 4953-2 (excerpt) → Derivation of QM Levels



- Safeguarding the (application) data quality and consistence
 - simple syntax up to (company) specific rules
e.g., via data check tools like in use @ data exchange
 - Validation Properties of 3D artifacts
 - Check Sums of Meta data
 - QM2 more powerful methods
(e.g. Cloud of Points of 3D, Meta Data acc. LOTAR)

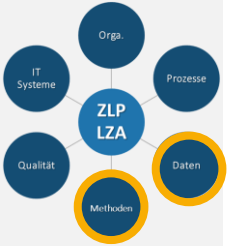
- Safeguarding the (3D / ZLP-Container) Document via Digital Signatures, Blockchain or similar

- Safeguarding the process by Certification
 - e.g., acc. DIN EN ISO9001 Quality Management Systems; Requirements
 - PK-DML Prüfkriterien für Dokumentenmanagementlösungen, 2. überarbeitete Auflage, VOI Verband Organisations- und Informationssysteme e.V.



VDA Project Group „ZLP“ (2013-2015)

Participating companies and projects



DFP

“ZfE - Zeichnungsfreier Entwicklungsprozess” (productive)



BOSCH



Drawing free In-House development



Concept and pilot “Simplified drawing with metadata”



Coordination and Consulting of the project group



User support at BMW Drawings & 3D



Feasibility study



“3D Zeichnungsloser Prozess” (productive)



DAIMLER

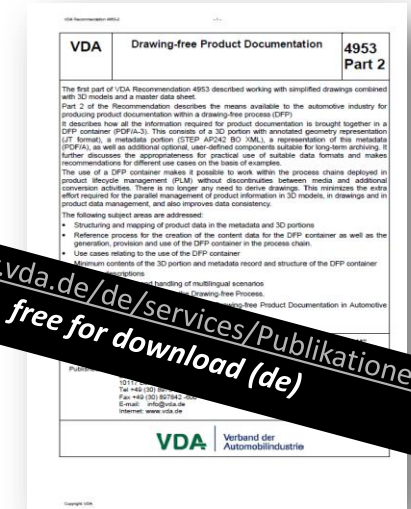
“3D Master” (productive)



VDA 4953-2 Drawing-Free Product documentation

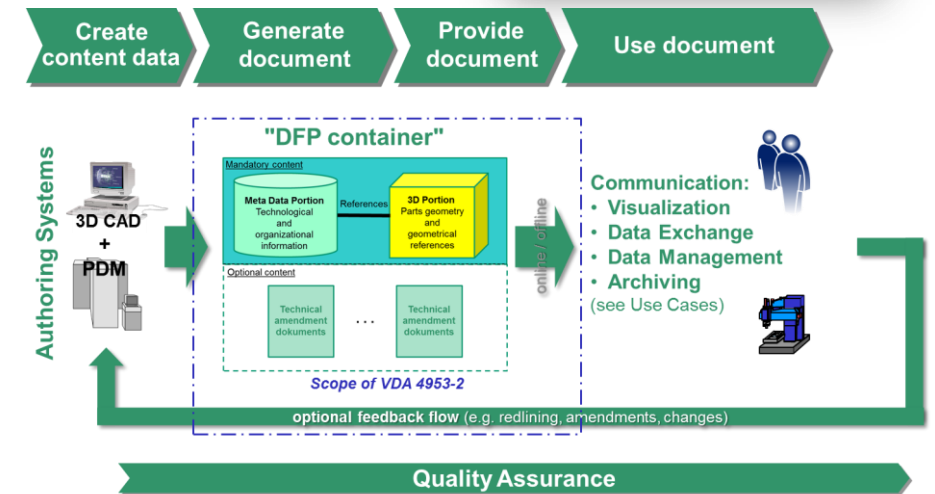
a harmonized solution (2014), coordinated and consulted by PROSTEP

- Product representation by **3D model and associated metadata set**
- Definition of the **minimum content and execution** of the 3D portion and the metadata
- Recommendations for the usage** of the 3D model and metadata collection within typical user scenarios
- Description of **specific use cases** within a drawing-free process (e.g., mirror part, variant part, joining elements)



Architecture of a DFP/ZLP Container *:

- **3D portion** with annotated representation of geometry (JT acc. ISO14306)
- **Metadata** portion as
 - record of structured metadata, separated from the 3D portion (STEP AP242 BO XML format acc. ISO10303-242)
 - presentation of these metadata (PDF/A acc. ISO19005)
- **optional elements**
 - further technical amendment documents incl. references (standard formats suitable for long-term archiving)

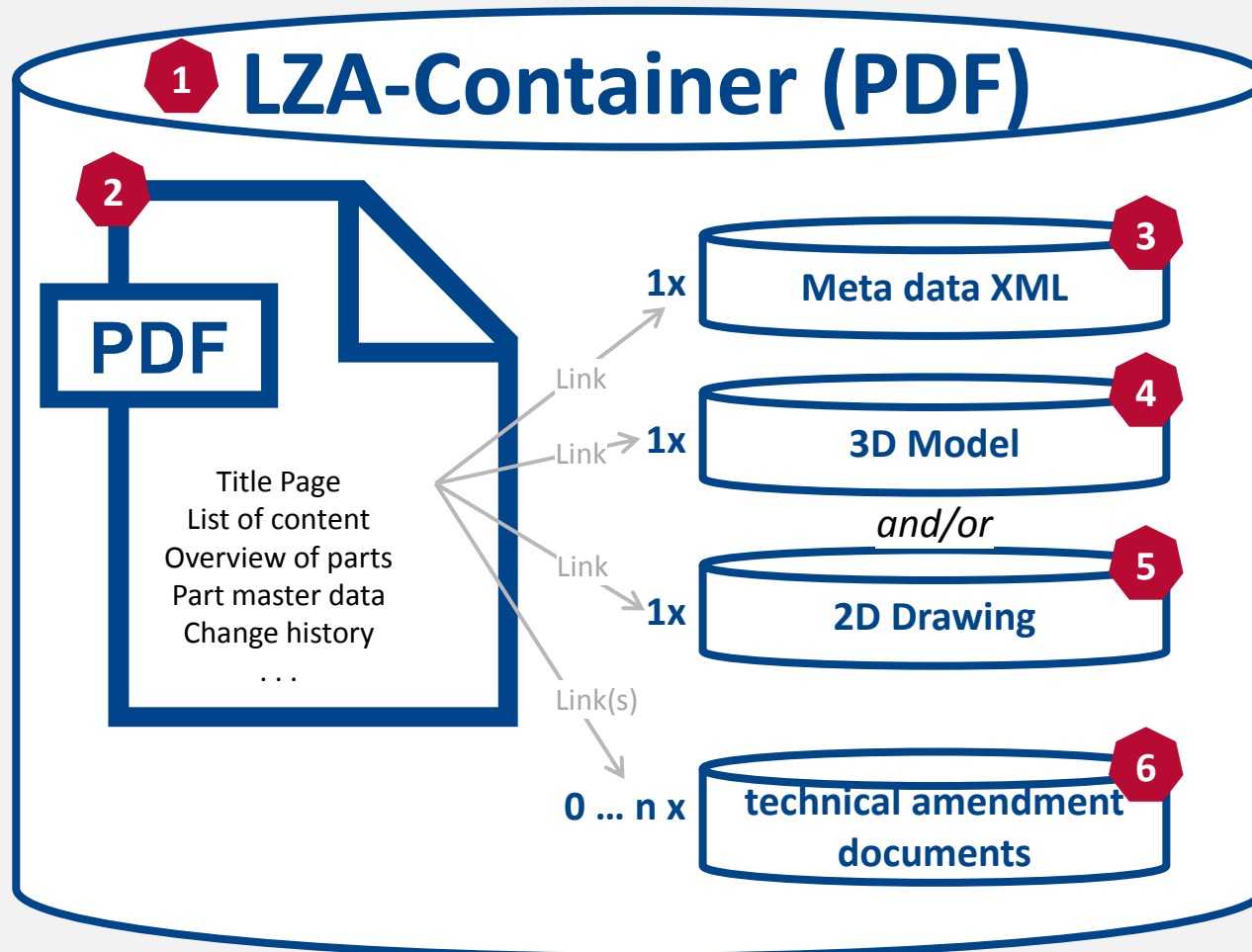


Source: VDA project group „ZLP“

* DFP - Drawing Free Process / ZLP - Zeichnungsloser Prozess Container = „3D document“

Reference Project: “LZA” @ an Automobil OEM

LZA-Container content



- 1 PDF-Container of a part version containing:
- 2 Meta data visualization (“rendering” based on 3) PDF pages of the LZA-Container
- 3 Meta data computer interpretable acc. ISO 10303-242 (STEP AP 242 BO XML)
- 4 3D geometric model (JT without BREP and/or
- 5 2D drawing (TIFF format)
- 6 Technical amendment documents of standardized formats (PDF), e.g., MS Office documents in neutral formats like PDF/A or TIFF

Example “3D Document” Container Technical Data Packages based on PDF



Product Requirements
PDM Meta Data
3D Geometry
PMI
Related Documents
etc.

Archivierung (Name, Datum)	automation generiert	10.10.16
Erstellung (Name, Datum)	Anna Jacobs	
Änderungsfreigebe / Change Approval (Status, Name, Datum)	REL: Susan Wong	01.10.16
Erzset für		
Geprüft	10.06.16 Alexander Koch	
Gelesen	12.06.16 Emma West	

Inhaltsverzeichnis	
Bauteile Übersicht / Parts List Overview	2
Housing MM - Baugruppenstruktur / Assembly Structure	2
Housing MM - Stammdaten / Master Data	3
Housing MM - Änderungs Historie / Change history	4

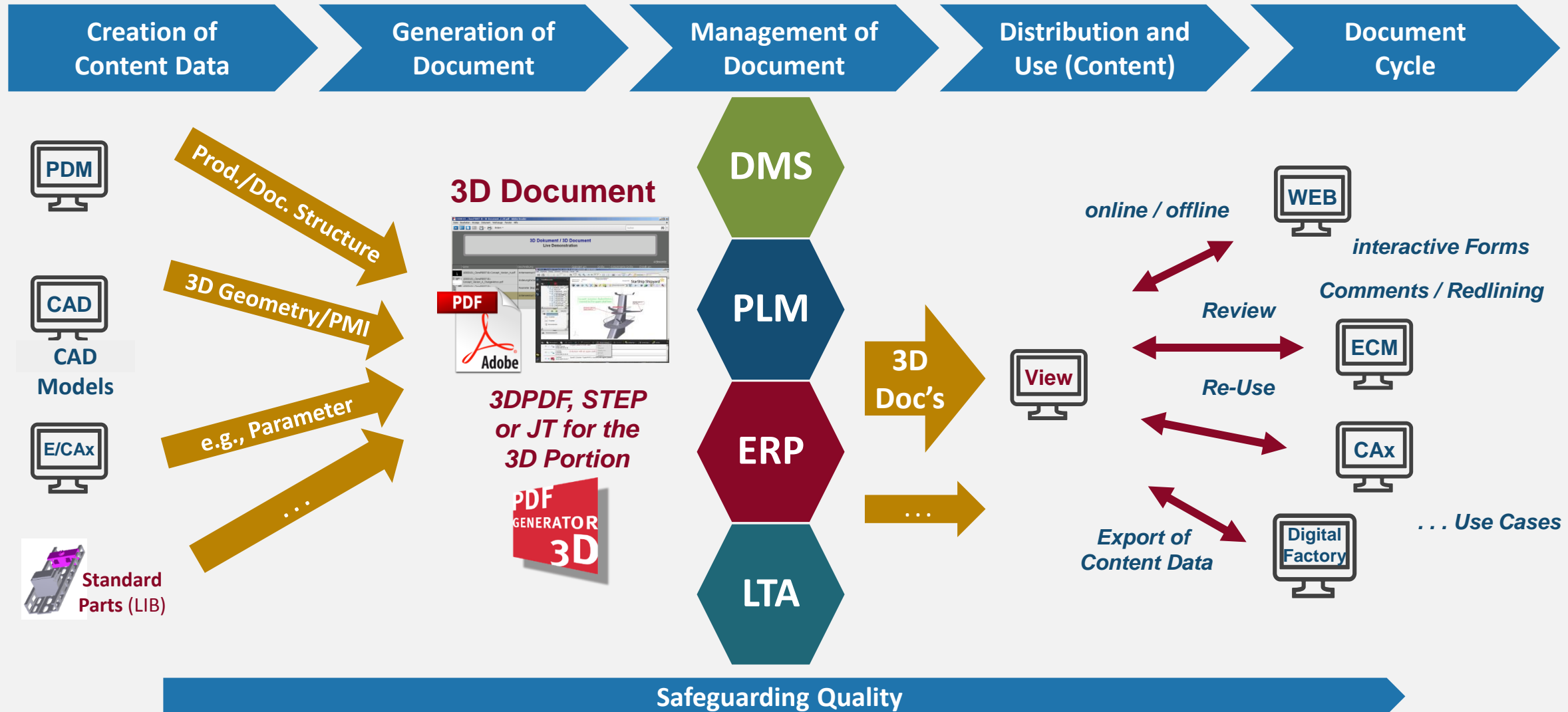
Anlagen	
Metadaten XML / meta data XML	
3D Geometrie / 3D geometry	
2D Zeichnung / 2D drawing	
Technisches Zusatzdokument 1 / Technical amendment 1	
Inspection Report	

Case No.	PMI Name	Text	Dimensional Accuracy	Type	Results	Inspection / Test Results	Comments	Approved
1	Linear Dimension (R)	52,3 ± 0,10	Dimension / Markup					
2	Linear Dimension (R)	17,2 ± 0,10	Dimension / Markup					
3	Linear Dimension (R)	30,1	Dimension / Markup					
4	Datum Feature Symbol E of IS		Datum / Markup					
5	Linear Dimension (R)	122,0	Dimension / Markup					
6	Linear Dimension (R)	52,2 ± 0,10	Dimension / Markup					

3D Geometry
PMI / GD&T / Annotations
... here an example use case with extension
of Measurement Features / Quality Criteria
“to-be “ ↔ “as-is”

Reference Aerospace: 3D ECAD (Harnesses)

Documentation and Supply for Downstream Processes (2007-2008)



Many thanks for your attention
... any questions?



Your contact:

Andreas Trautheim-Hofmann

Senior Manager & Principle Consultant

PLM Strategy & Processes

Mail andreas.trautheim@prostep.com

Phone +49 6151 9287 370

Cell +49 178 9509 380

Michelle Lindlar | Archivierung von Building Information Models (BIM) - Risiken und Möglichkeiten

Michelle Lindlar stellt zunächst das von 2013 bis 2016 durchgeführte DURAARK-Projekt vor, welches die Entwicklung von Methoden und Tools zur Langzeitarchivierung und Kuratierung von 3D-Building Data und –Metadaten zum Ziel hatte.

In der Folge geht sie im Hinblick auf Building Data auf die damit verbundene Datenmengen-Problematik ein. Pro Großprojekt (ab 10 Mio. US-Dollar) entstehen zum Beispiel unter anderem durchschnittlich 56.000 Seiten Dokumentation und 50 verschiedene digitale Objekttypen.

Building Information Models (BIM) decken den kompletten Bauprozess von der Design- bis zur Konstruktionsphase ab (inkl. Planung, Kostenrechnung etc.) und dienen auch nach Fertigstellung des Gebäudes als erweiterbare Dokumentation für das Gebäudemanagement. Als solches wird BIM auch oft als 7D-Datenmodell beschrieben, welches weit über die klassische dreidimensionale Modellierung in z.B. CAD hinausgeht und Dimensionen wie Produktdetails, Zeitplanung, Kostenplanung, energetische Optimierung und Bauunterhaltung beinhaltet.

Während 3D-Pläne aus klassischer Sicht als vorschreibende Darstellung gelten, sind 3D-Scandokumente beschreibende Dokumente, die den IST-Stand zum Erfassungszeitpunkt widerspiegeln. Die Lücke zwischen vorschreibenden und beschreibenden Dokumentationen hat das DURAARK-Projekt durch die semi-automatische Generierung von IFC-Modellen basierend auf Punktwolken geschlossen. Durch diese Verknüpfung vor- und beschreibender Informationsmodelle wird BIM auch für historische Gebäude voll einsetzbar, für die bisher CAD-Pläne nur aufwendig erstellbar waren.

Im Gegensatz zu allen anderen Wirtschaftszweigen hat die Bauindustrie in den letzten 40 Jahren kaum Produktivitätszuwächse verbuchen können. Steigende Lohnkosten wirken sich daher besonders auf Bauvorhaben aus. Als Gegenmittel wird eine stringenter Organisation von Bauprojekten mittels BIM empfohlen, da so eine effiziente Übergabe aller vorliegenden Dokumentation zwischen den einzelnen Gewerken und insbesondere zwischen Planungs-, Konstruktions- und Betriebsphase möglich wird.

Unterschieden wird zwischen offenen und geschlossenen BIM. Bei geschlossenen BIM handelt es sich in der Regel um eine komplette Softwaresuite eines Anwenders, in welche verschiedene weitere Anwendungen integriert werden können. Offene BIM ermöglichen eine modulare Herangehensweise. Hierdurch erfährt man zwar einen Funktionalitätsverlust, jedoch erhöht sich die Flexibilität durch die Möglichkeit der Datenaustauschverfahren.

Während geschlossene BIM-Systeme wie ‚AutoCAD‘ auf proprietäre Dateiformate bauen, existiert mit dem ‚STEP Application Profil Industry Foundation Classes‘ (IFC), welches als ISO 16739 standardisiert ist, ein offenes Datenaustauschformat. Im Rahmen des

DURAARK-Projekts wurden für verschiedene IFC-Schemata und Serialisierungen Dateisignaturpattern geschrieben, welche die automatische Erkennung via PRONOM-basierter Tools wie ‚DROID‘ und ‚Siegfried‘ ermöglichen, sowie technische Metadaten definiert und Extraktoren entwickelt. Das für IFC spezifizierte technische Metadaten-Schemata ‚ifcm‘ beinhaltet extrahierbare Informationen über den Erzeugungsprozess des digitalen Objekts, verwendete Parametrisierungen sowie Informationsbreite und –tiefe.

Auf kuratorischer Ebene wurde im DURAARK-Projekt (2013-2016) der Ansatz verfolgt, zur Lösung der übergreifenden Bestandserhaltungsfrage ein semantisches Netz aus Metadaten zu spannen, das sich das Linked Data-Konzept zunutze macht. Hierzu wurden im ersten Schritt aus Linked Open Data Quellen Informationen zur deskriptiven Anreicherung genutzt.

Die zur Anreicherung genutzten Quellen wurden geharvested und in einem Metadatenspeicher namens „Semantic Digital Archive“ neben den erfassten Metadaten archiviert. Ein zweites System, das „Semantic Digital Observatory“, überwachte die zur Anreicherung eingesetzte Quellen auf Veränderungen und löst auf Policy-Ebene definierten Schwellwerten ein re-harvesting und re-enrichment aus.

Mit dieser ersten prototypischen Entwicklung eines Systems zur automatischen semantischen Migration beschriftet das Projekt neue Wege der Langzeitarchivierung. In den letzten Jahren kann eine zunehmende Verbreitung von PDF/E (ISO 24517-1:2008), insbesondere für Konstruktionsdaten, beobachtet werden, allerdings bleiben aktuell noch viele zusätzlichen Dimensionen eines BIM außen vor. Eine aus Sicht der klassischen Archive faszinierende Folge von PDF/E ist, dass sich aus dem Inhalt der PDF-Datei zwar eine definierte Anzahl Normansichten, aber praktisch unendlich viele druckbare Seiten ergeben. Sollte für PDF/E eine langfristig haltbare Abart analog zu PDF/A definiert werden, so wäre dieser Umstand mit zu bedenken.

Zum Abschluss stellte Michelle Lindlar die offene Frage, ob langfristig ein Beharren auf proprietären Herstellerformaten, ein Umwandeln in offene CAD-Standards oder ein Vereinfachen auf PDF/E die gängige Lösung wird. All diese Lösungen haben ihr Für und Wider und hängen letztlich von den Erwartungen der Hersteller und der Kunden ab.

Informationen zum im DURAARK-Projekt entwickelten ‚techMD‘-Schemata:

‚ifcm‘-Schemata:

https://github.com/DURAARK/Schemas/blob/master/xml/2015_03_30_duraark_ifcm_1_1.xsd

Beschreibung der technischen Metadaten-Schemata ‚ifcm‘, ‚e57m‘ und des deskriptiven Metadatenschemata ‚buildm‘:

Lindlar, Michelle, Panitz, Michael, & Gadiraju, Ujwal. (2015). DURAARK Deliverable: Ingest and Storage of 3D Objects in a Digital Preservation System.

<https://doi.org/10.5281/zenodo.1115511>

Weitere Informationen:

Lindlar, Michelle, Saemann, Hedda, Ochmann, Sebastian, Jonsson, Östen, & Gadiraju, Ujwal. (2014). DURAARK Deliverable: Current state of 3D object digital preservation and gap-analysis report. <https://doi.org/10.5281/zenodo.1115503>

Ball, Alex. DPC Technology Watch Report: Preserving Computer Aided Design (CAD). (2013). <http://dx.doi.org/10.7207/twr13-02>

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK



Archivierung von Building Information Models (BIM) – Risiken und Möglichkeiten

M. Lindlar

Koblenz, 8. November 2018

Workshop: „Von der Theorie zur Praxis:
Bestandserhaltung digitaler Unterlagen“

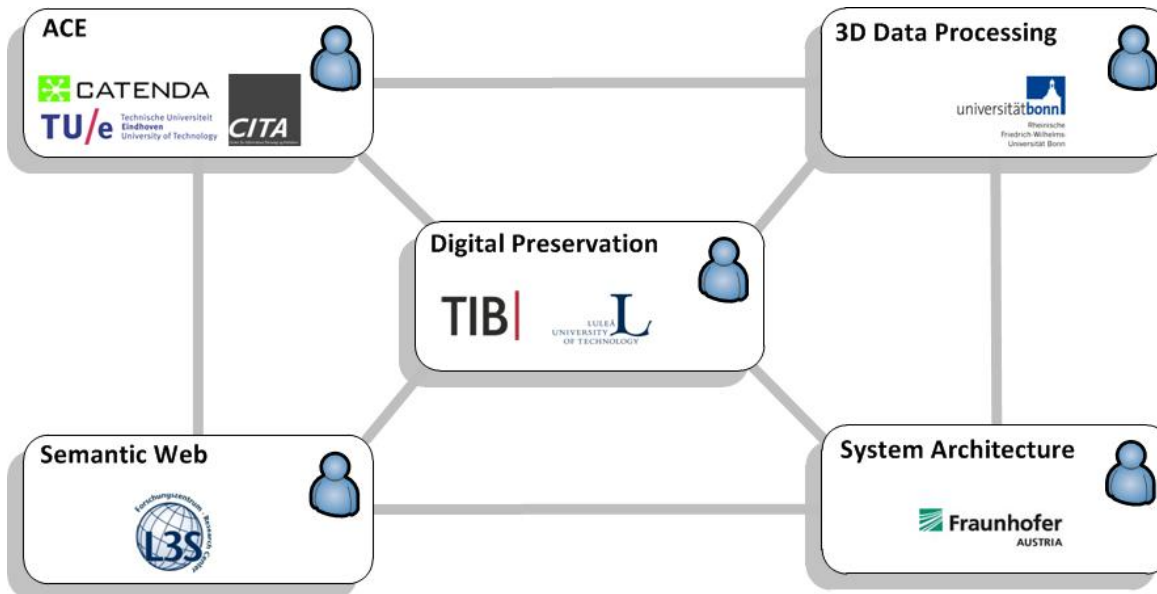
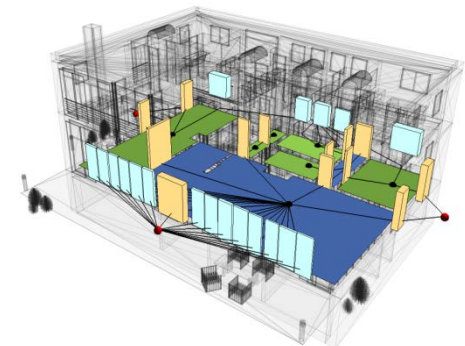
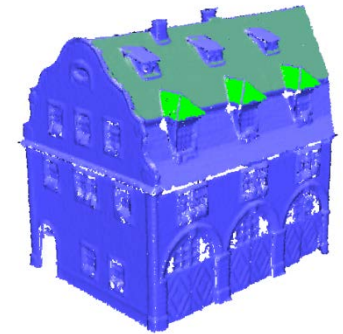


DURAARK Projekt 2013 - 2016

DURAARK (DURable Architectural Knowledge)
 FP7 – ICT – Digital Preservation (STReP)

Ziel:

Entwicklung von Methoden und Tools zur Kuratierung und Langzeitarchivierung von 3D Building Data & Metadaten.



Building Data – das Problem

Durchschnittliche Zahlen für Großprojekte (> 10 Mio USD):

56.000 Seiten Dokumentation

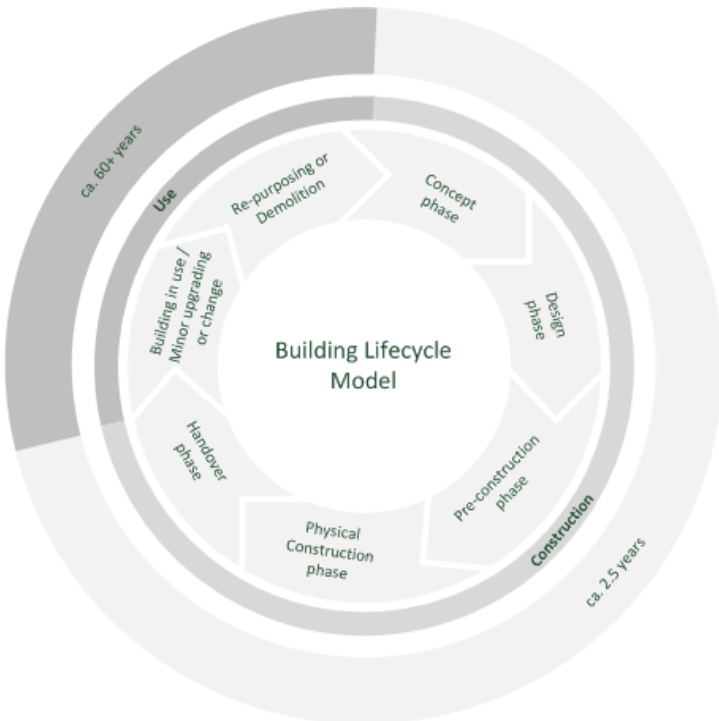
850 Beteiligte Personen

420 Beteiligte Firmen (inkl. Zulieferer und Subsubunternehmer)

50 verschiedene digitale Objekttypen

25 Bankschließfächer mit Projektdokumentationen

[Quelle: Hendrickson, C. 2008. Project Management for Construction.]



... typische
Projektübergabe:

[Quelle: Jakob Beetz.]



Was ist BIM?

Building Information Models (BIM)

- Weiter als „nur CAD“ durch Abdeckung des kompletten Design-bis-Konstruktionsprozess (inkl. Projekt Planung, Kosten, Spezifikationen einzelner Bauteile, ...)
- Klassischer Weise eine vorschreibende Darstellung („as-planned“) in Abgrenzung zu z.B. Punktwolkenscans als beschreibende Darstellung („as-is“) → Trend geht allerdings zu Einsatz im „as-is“ Bereich z.B. durch starke Verbreitung im Gebäudemanagement

3D CAD

Geometrie entlang X-Y-Z Axen



4D CAD

Projektplan / -laufzeit



5D CAD

Informationen zu Kosten



6D CAD

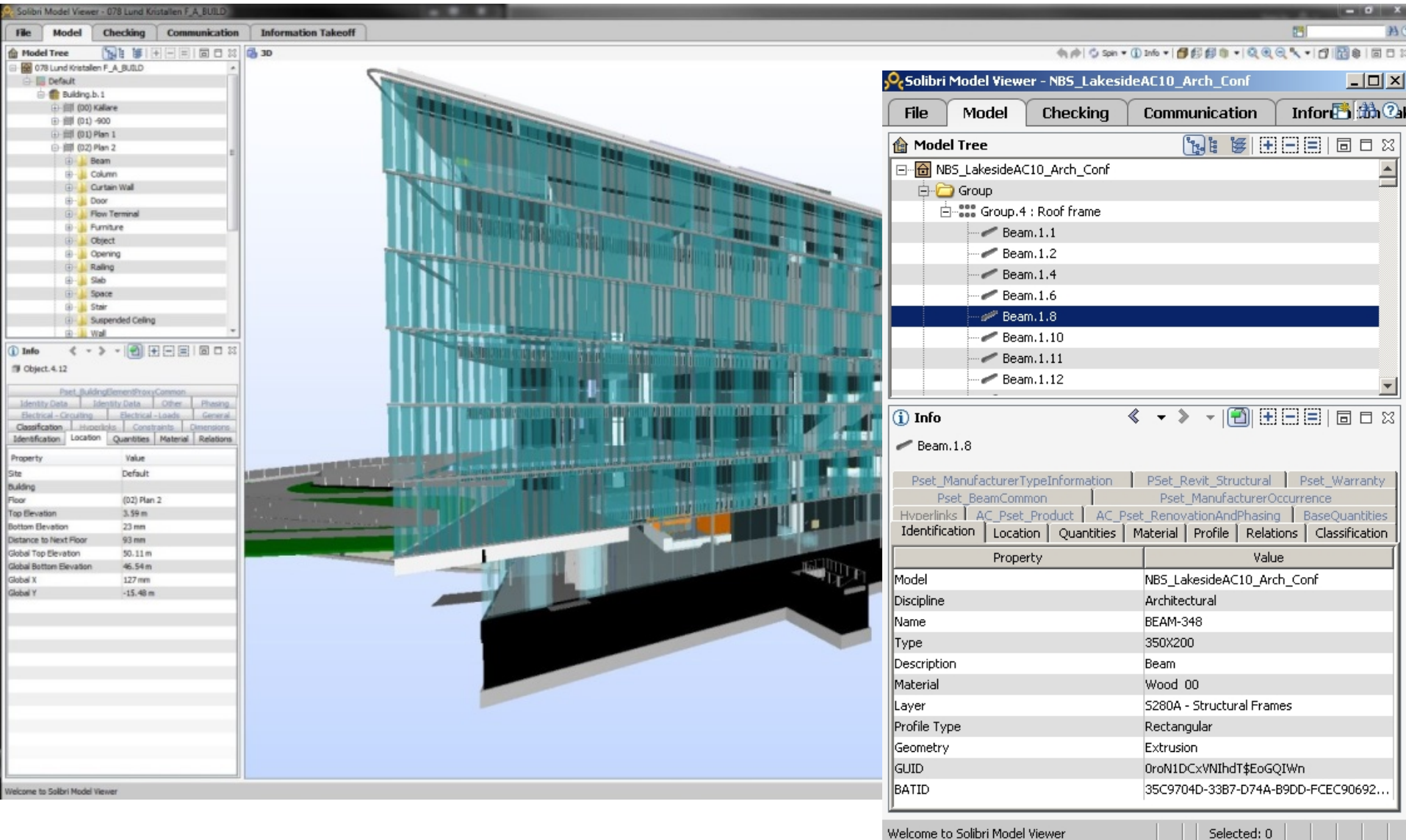
Energie und Nachhaltigkeit



7D CAD

Gebäudemanagement

BIM Model Ansicht



Solibri Model Viewer - 078 Lund Kristallen F_A_BUILD

Model Tree

- 078 Lund Kristallen F_A_BUILD
 - Default
 - Building.b.1
 - (00) Kallare
 - (01) -900
 - (01) Plan 1
 - (02) Plan 2
 - Beam
 - Column
 - Curtain Wall
 - Door
 - Flow Terminal
 - Furniture
 - Object
 - Opening
 - Railing
 - Slab
 - Star
 - Suspended Ceiling
 - Wall

Info

Object 4.12

Pset_BuildingElementProxyCommon

Property	Value
Site	Default
Building	
Floor	(02) Plan 2
Top Elevation	3.59 m
Bottom Elevation	23 mm
Distance to Next Floor	93 mm
Global Top Elevation	50.11 m
Global Bottom Elevation	46.54 m
Global X	127 mm
Global Y	-15.40 m

Solibri Model Viewer - NBS_LakesideAC10_Arch_Conf

Model Tree

- NBS_LakesideAC10_Arch_Conf
 - Group
 - Group.4 : Roof frame
 - Beam.1.1
 - Beam.1.2
 - Beam.1.4
 - Beam.1.6
 - Beam.1.8**
 - Beam.1.10
 - Beam.1.11
 - Beam.1.12

Info

Beam.1.8

Pset_ManufacturerTypeInformation | Pset_Revit_Structural | Pset_Warranty

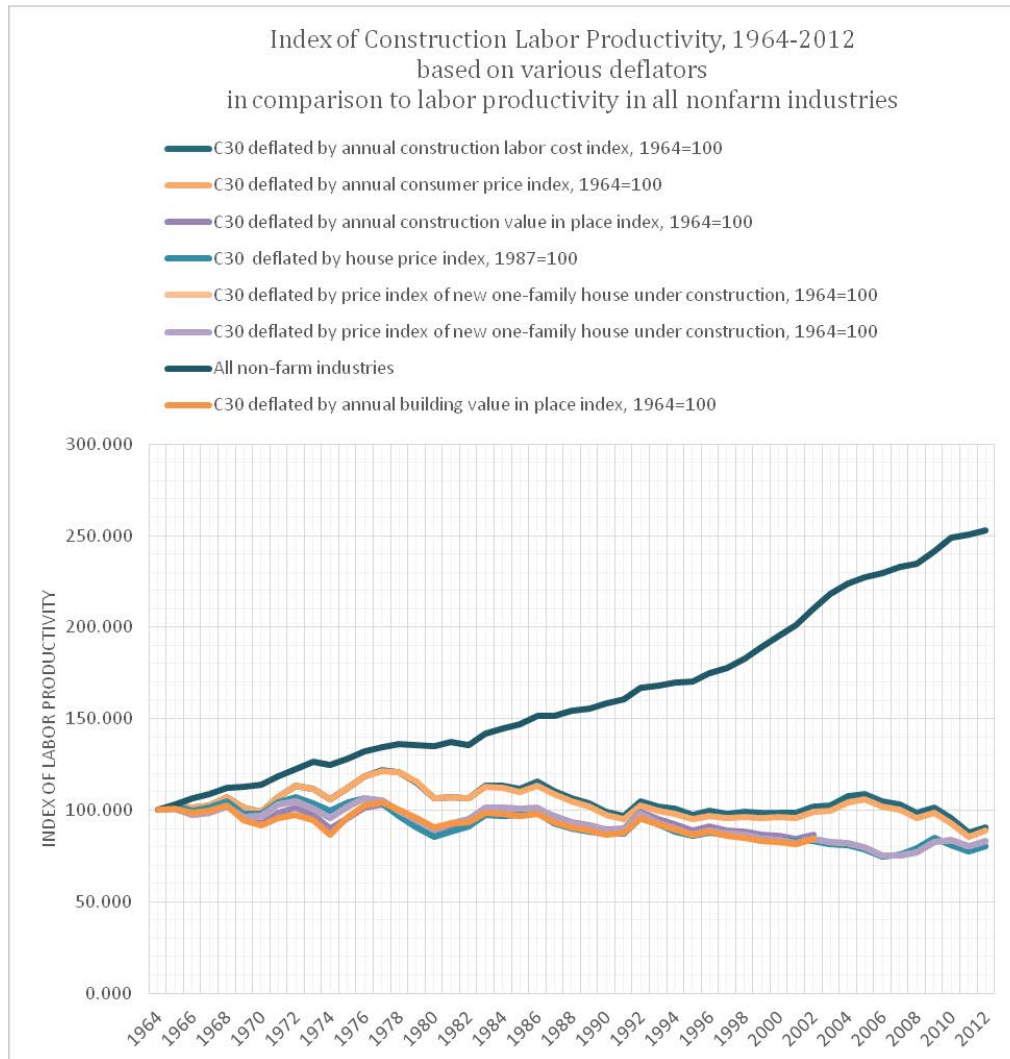
Pset_BeamCommon | Pset_ManufacturerOccurrence

Hyperlinks | AC_Pset_Product | AC_Pset_RenovationAndPhasing | BaseQuantities

Property	Value
Model	NBS_LakesideAC10_Arch_Conf
Discipline	Architectural
Name	BEAM-348
Type	350X200
Description	Beam
Material	Wood 00
Layer	S280A - Structural Frames
Profile Type	Rectangular
Geometry	Extrusion
GUID	0roN1DCxVNIhdT\$EoGQIWh
BATID	35C9704D-33B7-D74A-B9DD-FCEC90692...

Welcome to Solibri Model Viewer | Selected: 0

Produktivitätsverfall in der Bauindustrie – Gründe und Lösungen



“[...] major opportunities for positive change include better use of BIM enabled processes and procurement contracts [...] to support enhanced understanding, and efficiency. The industry needs to reduce its use of low-tech “stick built” approach and learn to take advantage of IT enabled practices closer to those used in the manufacturing industry.”

Nationale BIM Richtlinien (Stand: 2016)

Country	BIM Status	IFC Status	Driver
Australia	Not mandatory	Not mandatory	Association driven
Denmark	Mandatory (partially)	Mandatory (partially)	Government driven
Finland	Mandatory	Mandatory	Government driven
Germany	Not mandatory	Not mandatory	Association driven
Hong Kong	Mandatory	Not mandatory	Government driven
Netherlands	Mandatory (partially)	Mandatory (partially)	Government driven
Norway	Mandatory	Mandatory	Government driven
Singapore	Mandatory (partially)	Mandatory (partially)	Government driven
United Kingdom	Mandatory	Mandatory	Government driven
USA	Mandatory	Mandatory	Government driven

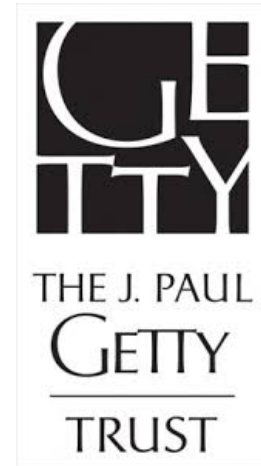
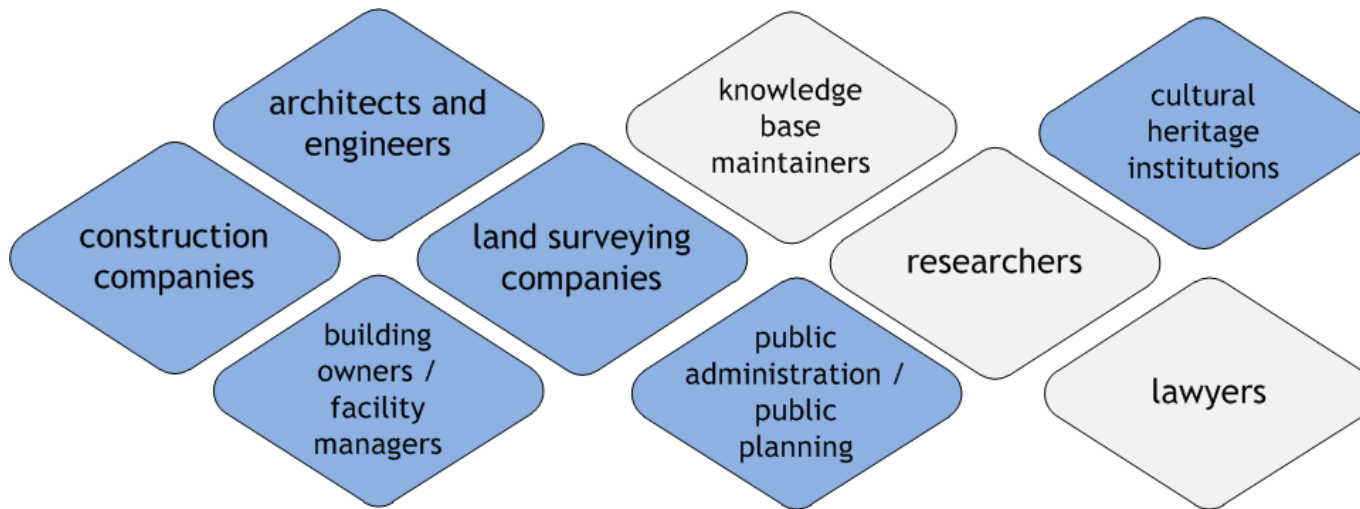


2015

2013



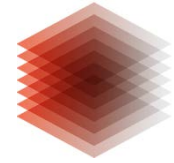
Stakeholder – wer arbeitet mit BIM ?



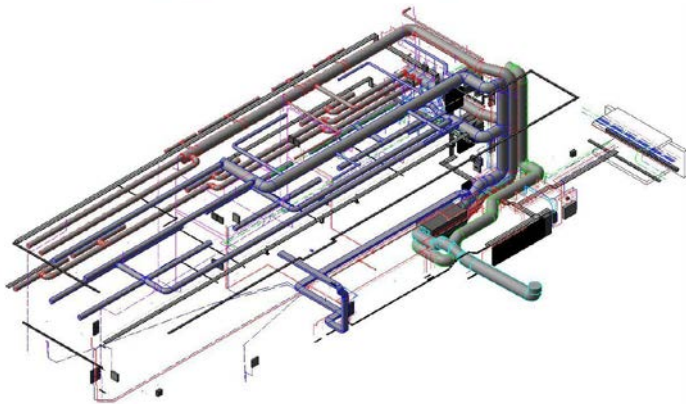
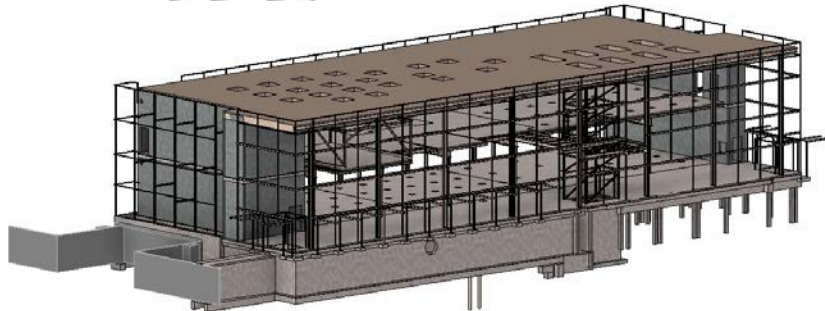
Technical University of Denmark



Massachusetts Institute of Technology



Closed BIM, Open BIM



Closed BIM

- komplette Softwaresuite eines Anbieters mit nahtloser Integration verschiedener Anwendungen / Gewerke (z.B. AutoDesk BIM Solution)

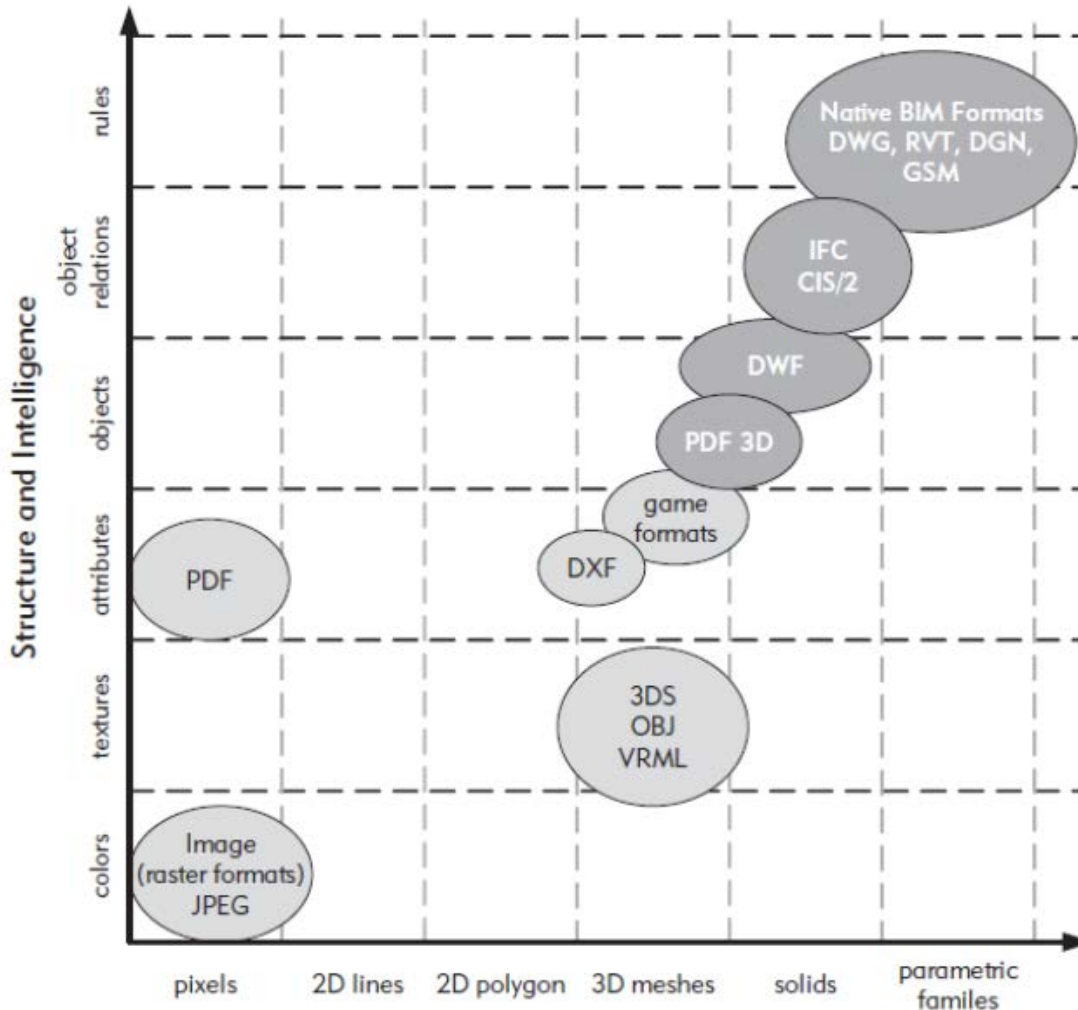
→ Enge Integration und beste Features, aber wenig Interesse in Datenaustauschroutinen

Open BIM

- modulare Herangehensweise mit verschiedenen Systemen (z.B. Catenda's bimsync als kleinster gemeinsamer Nenner)

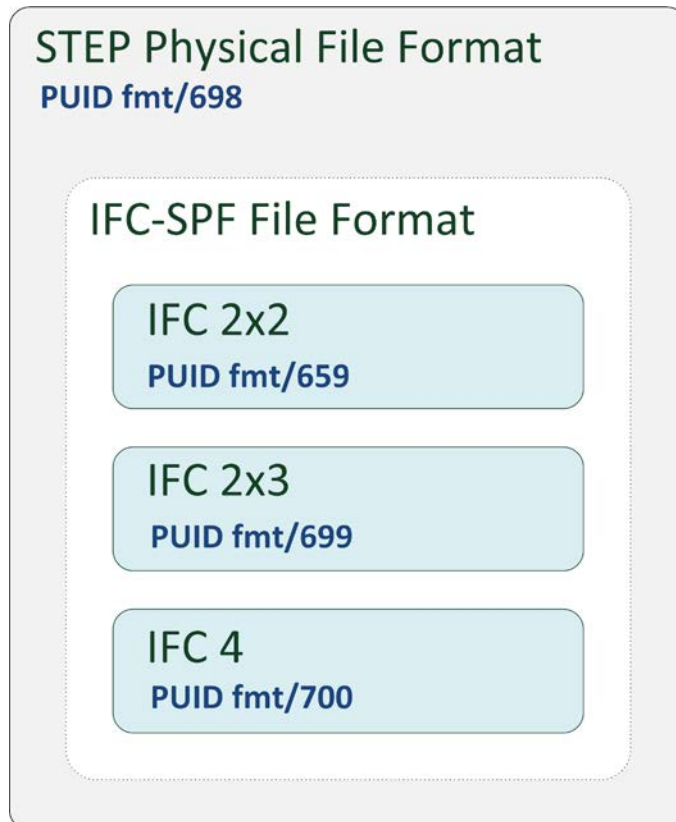
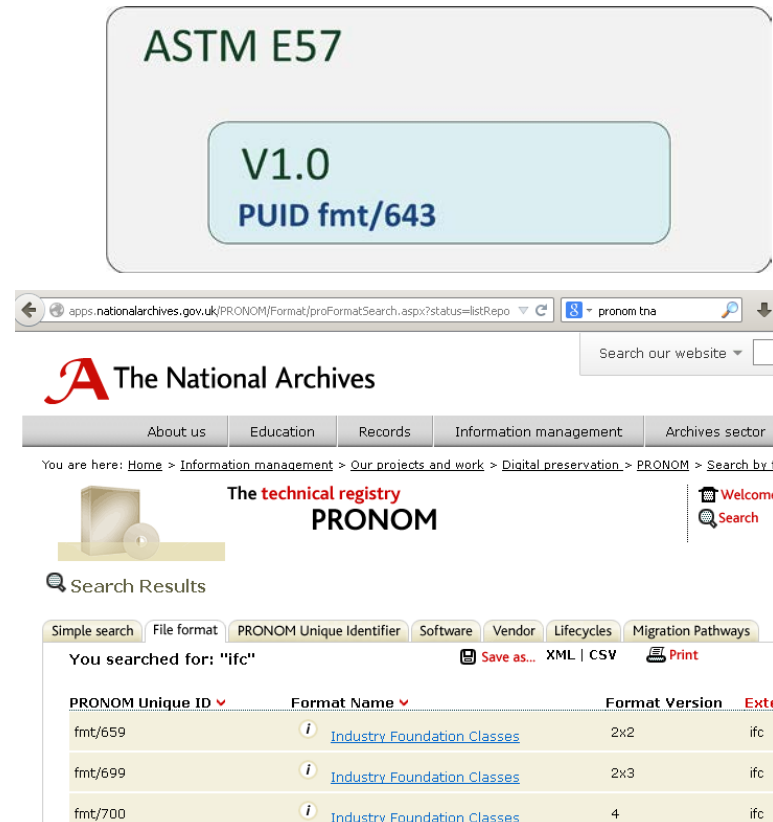
→ Funktionalitätsverlust, aber höhere Flexibilität durch Datenaustauschverfahren

Datenaustausch und Archivformate ?



Nähere Betrachtung von IFC – ISO16739:2013, basiert auf STEP Standard ISO 10303)

Bestandserhaltung: Identifizierung

ASTM E57

V1.0
PUID **fmt/643**

apps.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=lstRepo

The National Archives

About us | Education | Records | Information management | Archives sector

You are here: Home > Information management > Our projects and work > Digital preservation > PRONOM > Search by fr

The technical registry
PRONOM

Welcome
Search

Search Results

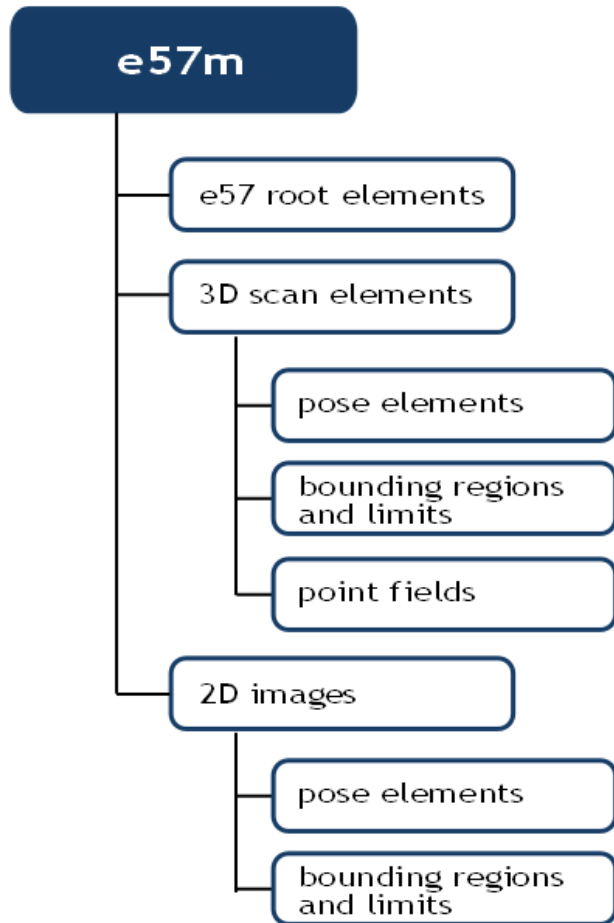
Simple search | File format | PRONOM Unique Identifier | Software | Vendor | Lifecycles | Migration Pathways

You searched for: "ifc" | Save as... | XML | CSV | Print

PRONOM Unique ID	Format Name	Format Version	Ext
fmt/659	Industry Foundation Classes	2x2	ifc
fmt/699	Industry Foundation Classes	2x3	ifc
fmt/700	Industry Foundation Classes	4	ifc

→ Beziehung von Formaten: jedes IFC (z.B. fmt/700) ist auch ein STEP File (fmt/698)

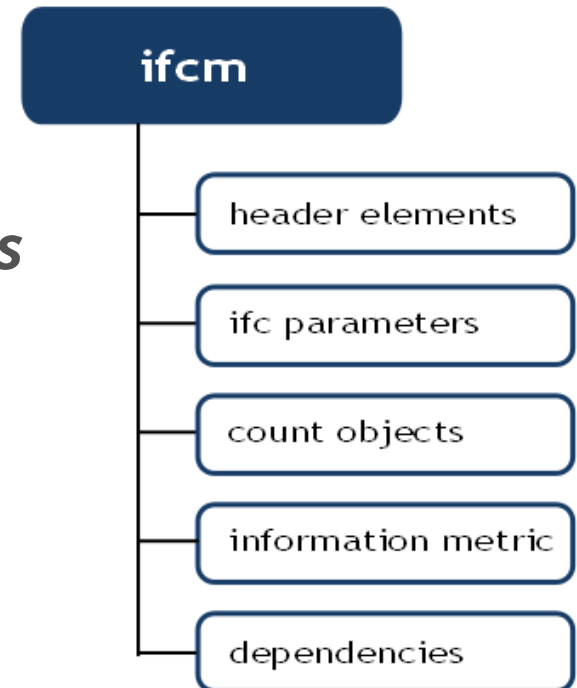
Technische MD Schemata: buildm, ifcm, e57m



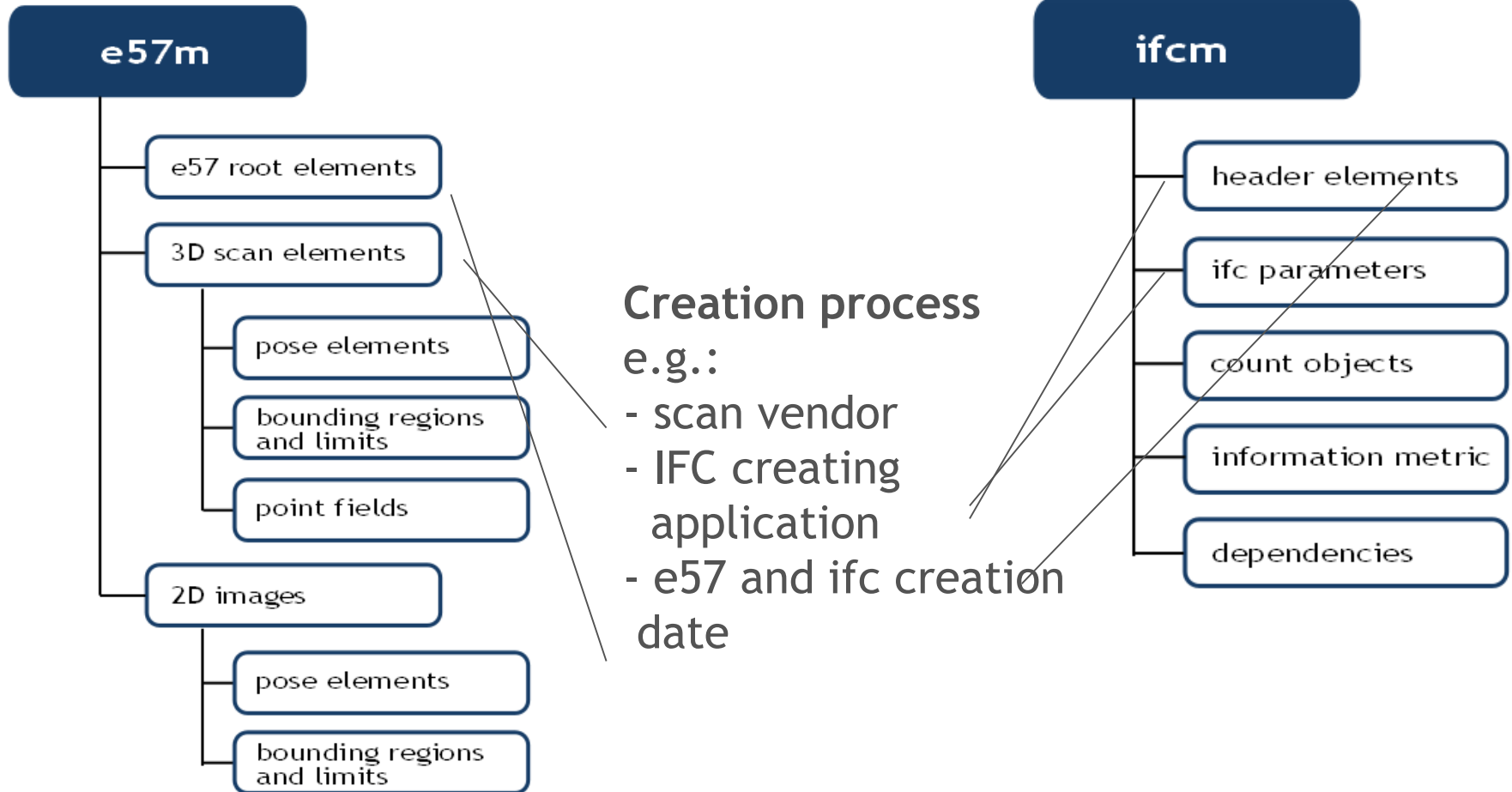
Creation process

Parametric information

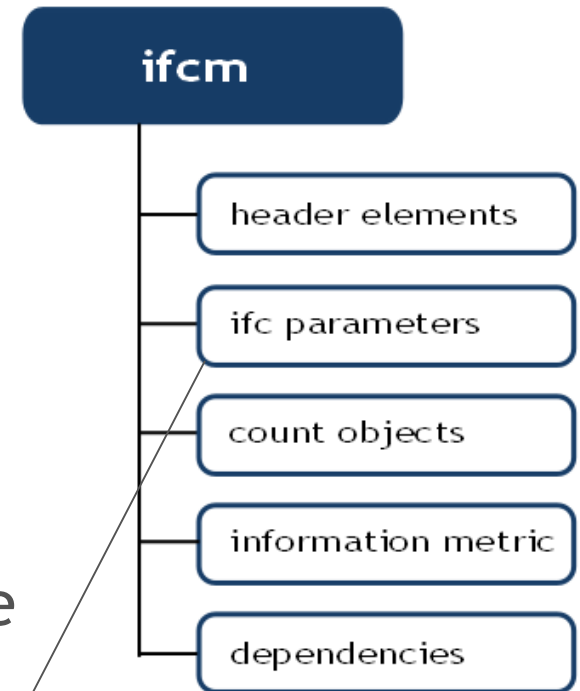
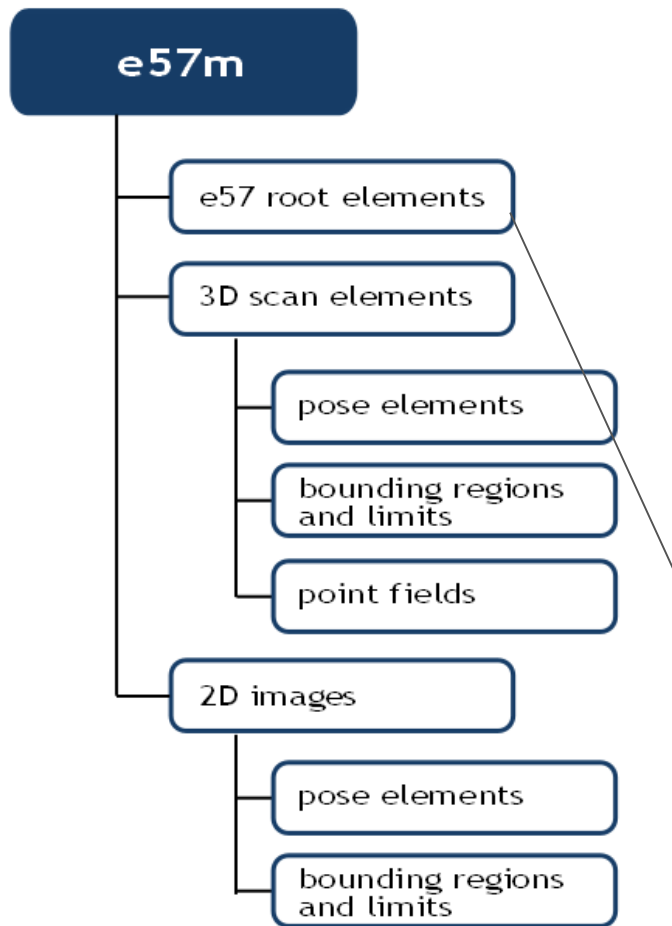
Content extent



techMd - Dateierzeugung



techMD - Maßeinheiten

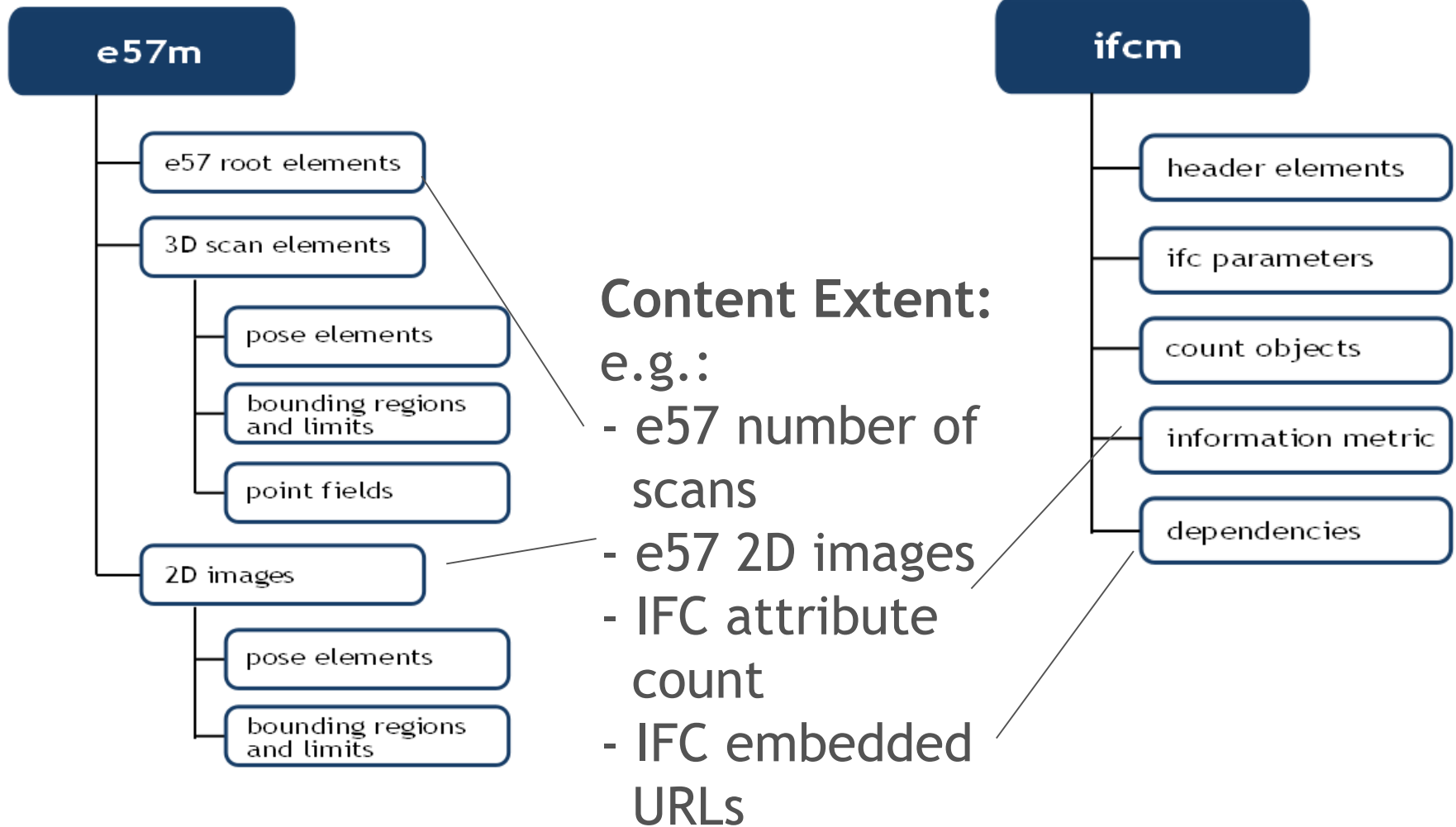


Parametric Information:

e.g.:

- e57 coordinate system
- IFC measuring units

techMD - Informationsdichte



Projektergebnisse Dateiformatcharakterisierung



für techMD Extraction:

- Schemata für buildm, ifc5, e57m:
<http://duraark.eu/data-repository/#schemas>
- Dokumentation als Data Dictionary in:
D.6.2 Ingest and Storage of 3D Objects in a Digital Preservation System (http://duraark.eu/wp-content/uploads/2015/02/DURAARK_D6.2.pdf)
- Extraktoren:
<https://github.com/DURAARK/duraark-metadata>



für Validierung:

- Gute Ergebnisse für Pointclouds via libe57-Validator
- Schlechte Ergebnisse für IFC ob der Komplexität – aktuell kein gut nutzbarer Validator

Semantische Anreicherung mit Linked Open Data



<BPIECode:Heizkostenverordnung 2009>



<dbpedia:library, office space>

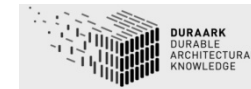
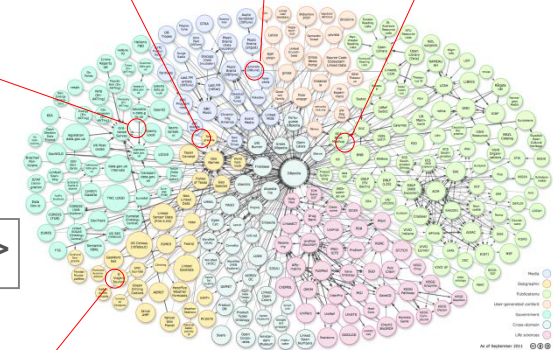
<WikiData:Marstall beim Welfenschloss>



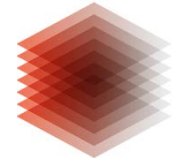
<GeoName: German National Library of Science and Technology >



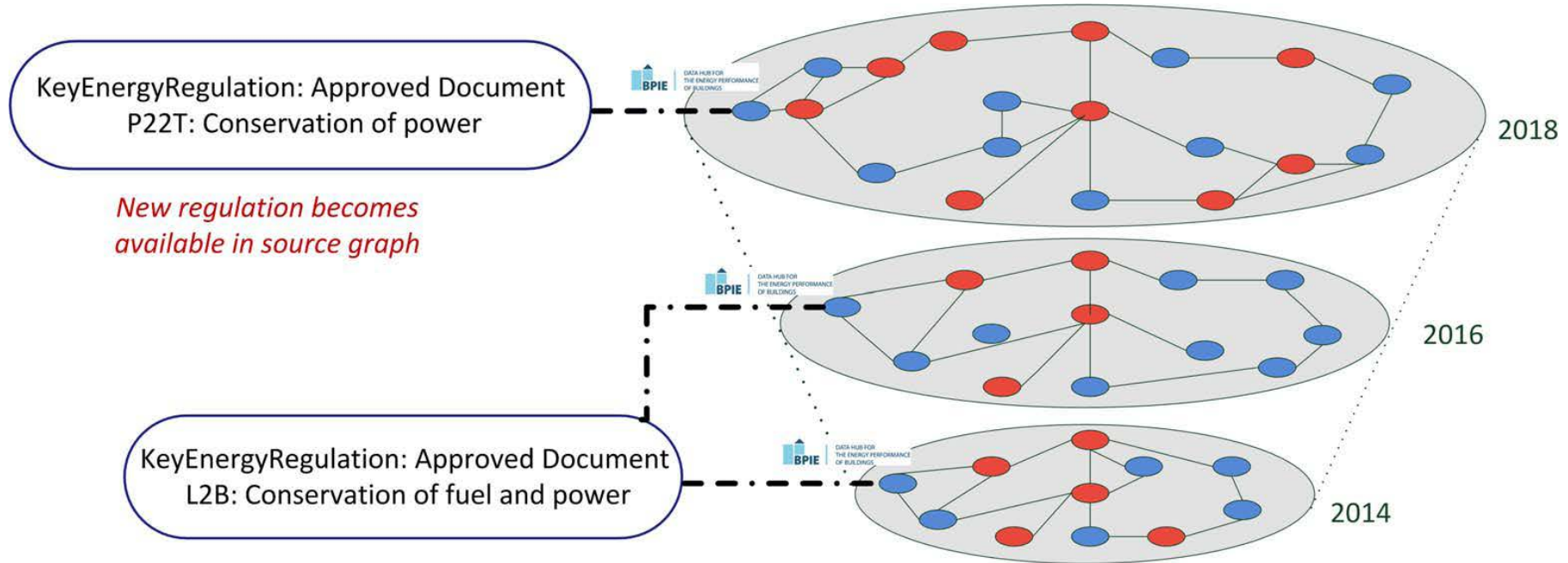
<GeoLatLong: 52.3814 / 9.72 >



Starting seed: Adresse



Von Anreicherung zur (semantischen) Archivierung



DURAARK SDO – semantic digital observatory layer

- Workbench zur Anreicherung aus LOD Quellen
- Beobachtet die genutzte Quelle auf Veränderungen
- Triggered (re-)harvest des Graphen bei Veränderungen

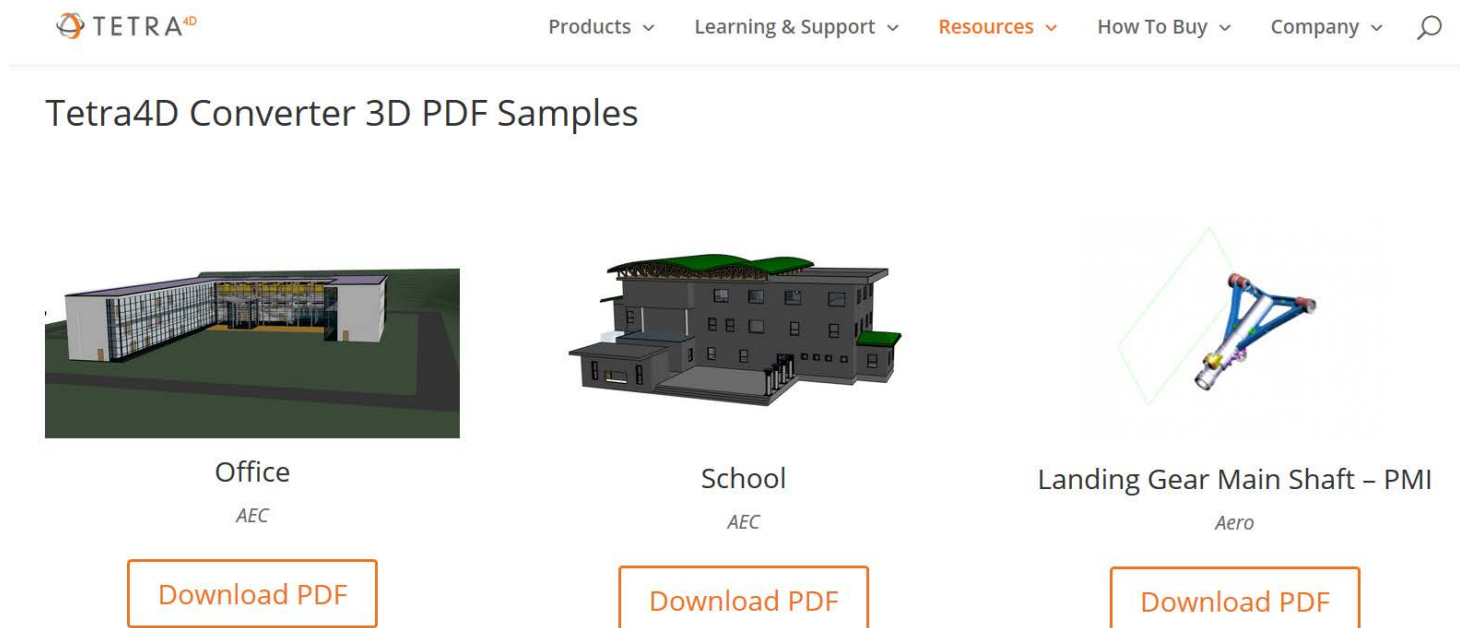
DURAARK SDA – semantic digital archive

- triple store as Zwischenspeicher für geharvestete Graphen
- kontrollierter Enrichment Pool

DURAARK Projektende: Januar 2016

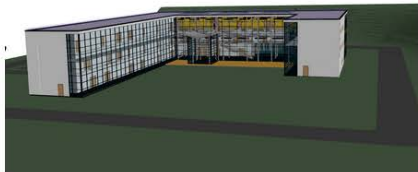


Gibt's was Neues?
Exkurs: PDF 3D (PDF/E)

Beispieldateien: <https://tetra4d.com/pdf-samples/>

A screenshot of the Tetra4D website's '3D PDF Samples' page. The page features a navigation bar with the Tetra4D logo and menu items: 'Products', 'Learning & Support', 'Resources', 'How To Buy', and 'Company'. Below the navigation bar, the heading 'Tetra4D Converter 3D PDF Samples' is displayed. Three sample categories are shown: 'Office' (AEC), 'School' (AEC), and 'Landing Gear Main Shaft - PMI' (Aero). Each category includes a 3D model image, a title, a category label, and a 'Download PDF' button.

TETRA^{4D} Products ▾ Learning & Support ▾ Resources ▾ How To Buy ▾ Company ▾

Tetra4D Converter 3D PDF Samples

Sample Name	Category	Action
	Office AEC	Download PDF
	School AEC	Download PDF
	Landing Gear Main Shaft - PMI Aero	Download PDF

PDF 3D – View 1

3dpdf-AEC-OfficeBuilding.pdf - Adobe Acrobat Pro

File Edit View Window Help

Create [Icons] Customize

1 / 1 [Navigation] 47% [Zoom] [Tools] Fill & Sign Comment

Model Tree

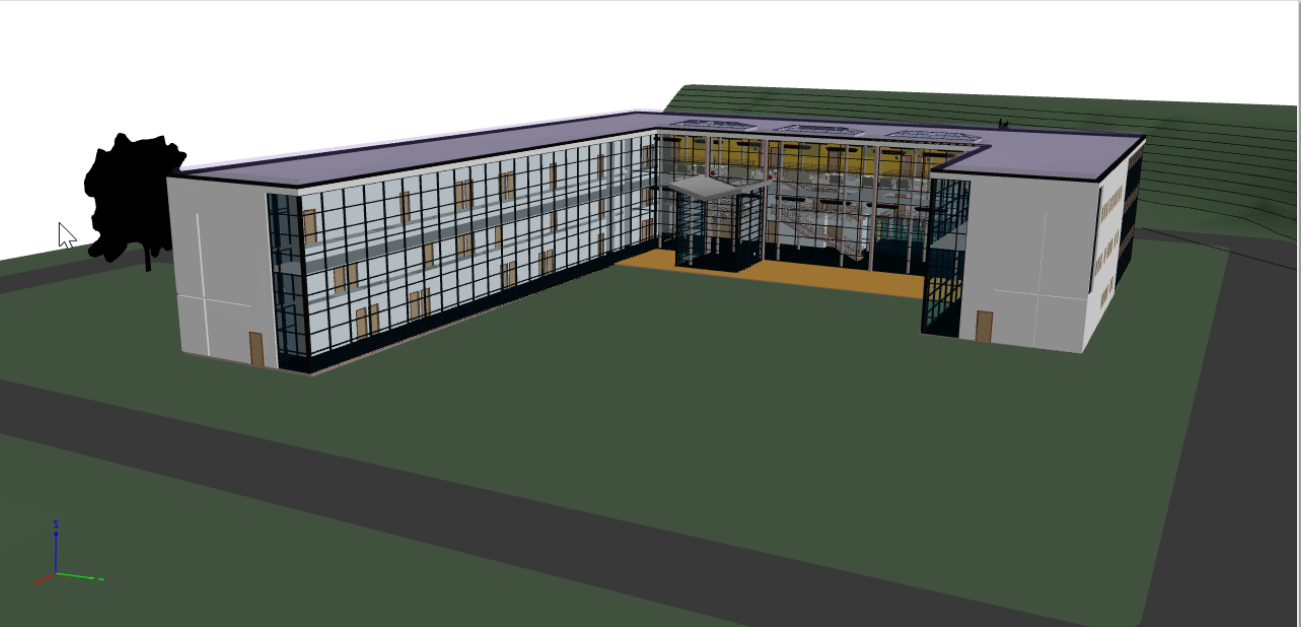
- model
 - node
 - Topography / Surface
 - Topography / Surface
 - Topography / Surface
 - Topography / Surface

Options

- Home
- Right
- Left

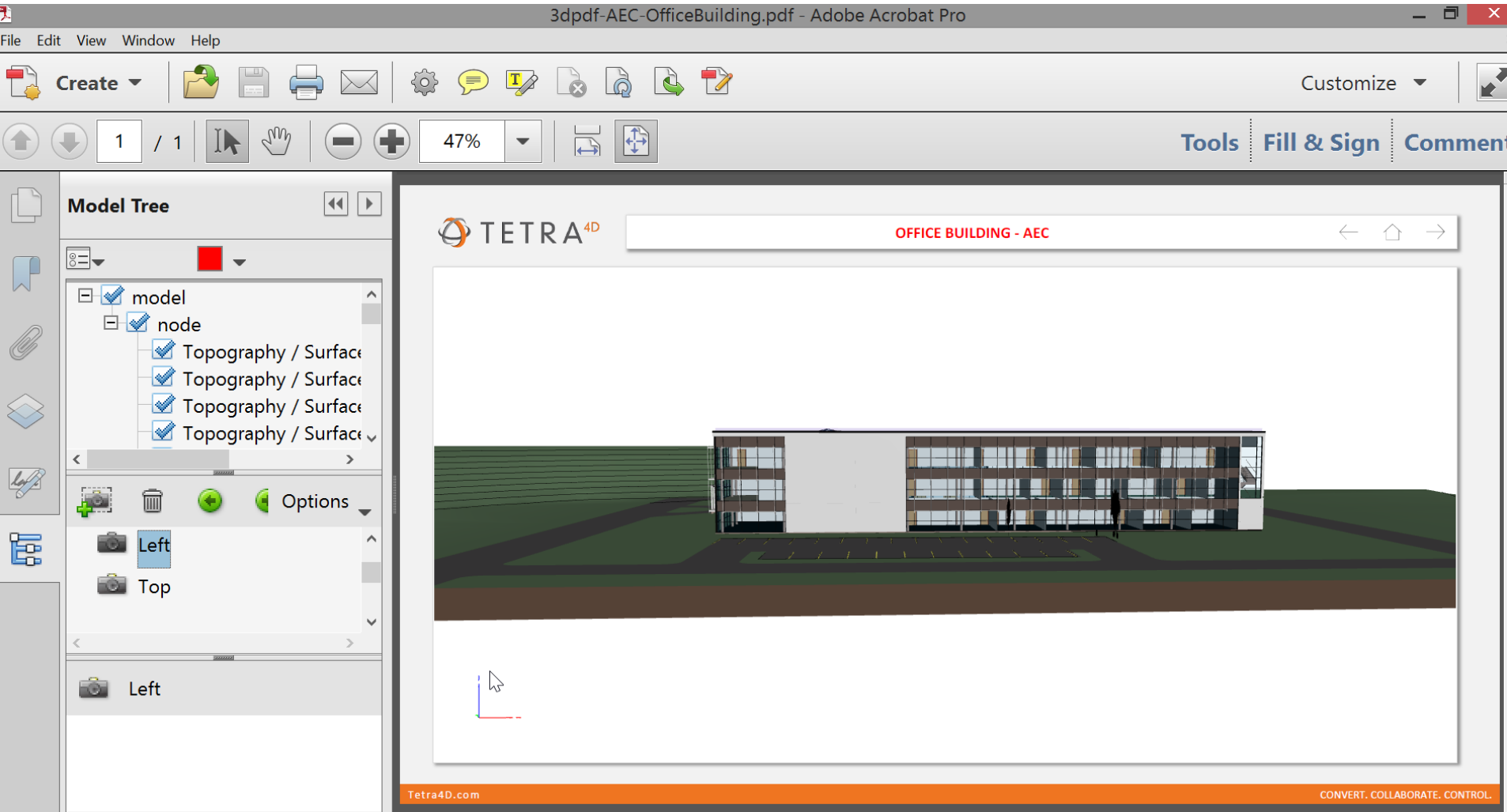
<No available information>

TETRA^{4D} OFFICE BUILDING - AEC



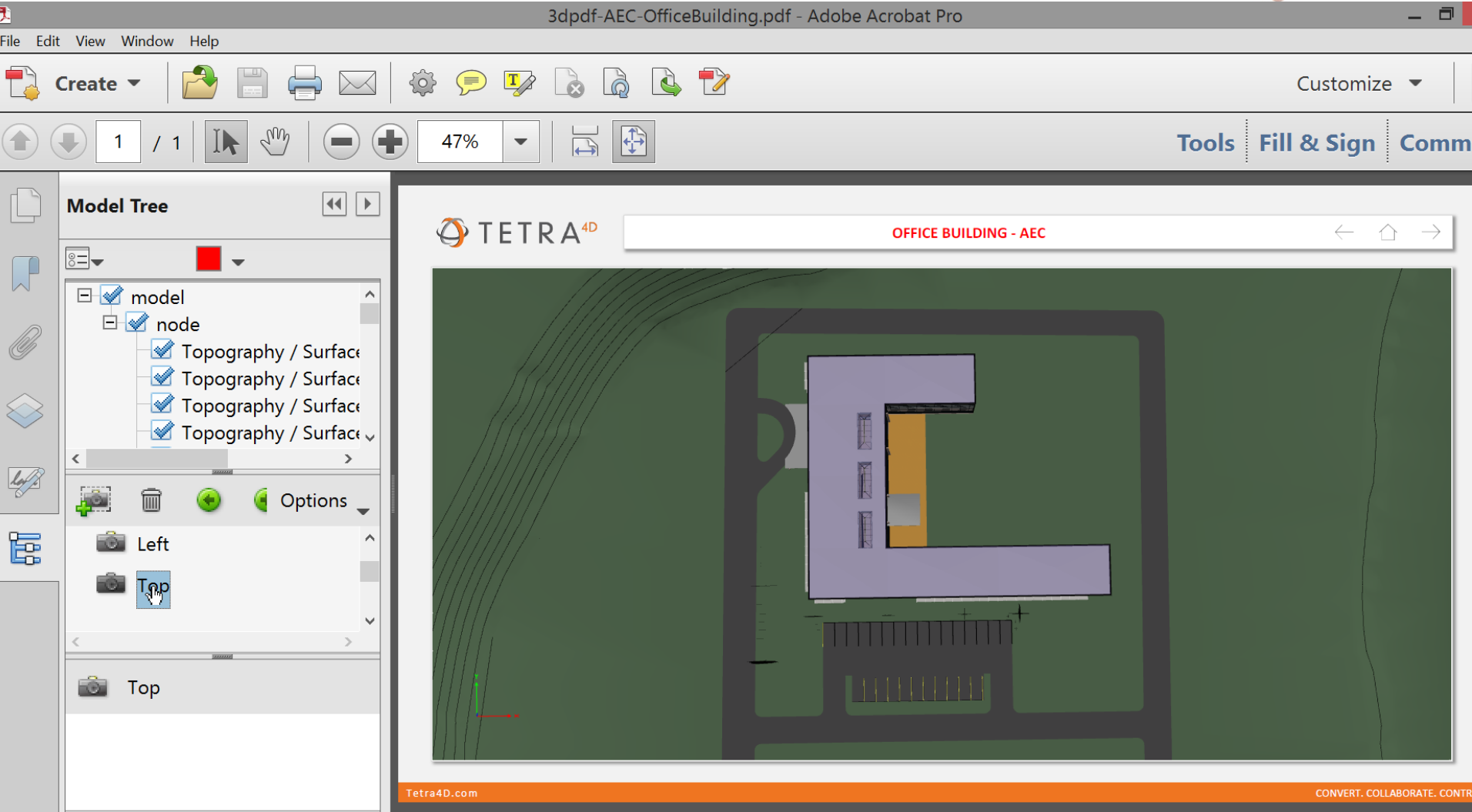
Tetra4D.com CONVERT. COLLABORATE. CONTROL.

PDF 3D – View Left



The screenshot displays the Adobe Acrobat Pro interface with a 3D model of an office building. The main window shows a left-side perspective view of the building, which has a modern design with large glass windows and a central entrance. The building is set on a green landscape with a parking lot in front. The interface includes a top menu bar (File, Edit, View, Window, Help), a toolbar with various icons, and a navigation pane on the left. The navigation pane is titled "Model Tree" and shows a hierarchical structure of the 3D model, including "model" and "node" folders, with several "Topography / Surface" items listed. Below the model tree, there are camera controls, including a "Left" view button. The main window also features a "TETRA 4D" logo and the text "OFFICE BUILDING - AEC" at the top. The bottom of the window has a footer with "Tetra4D.com" and the slogan "CONVERT. COLLABORATE. CONTROL."

PDF 3D – View Top



The screenshot displays the Adobe Acrobat Pro interface with a 3D model of an office building. The main window shows the model in a top-down view, with a purple building structure on a green terrain. The model is titled "OFFICE BUILDING - AEC".

The interface includes a menu bar (File, Edit, View, Window, Help) and a toolbar with various icons. The navigation pane on the left shows the "Model Tree" with a hierarchy of objects:

- model
 - node
 - Topography / Surface
 - Topography / Surface
 - Topography / Surface
 - Topography / Surface

Below the model tree, there are camera controls including "Options", "Left", and "Top" views. The "Top" view is currently selected. The status bar at the bottom of the model viewer shows "Tetra4D.com" and "CONVERT. COLLABORATE. CONTR".

PDF 3D – frei gedreht

3dpdf-AEC-OfficeBuilding.pdf - Adobe Acrobat Pro

File Edit View Window Help

Create [Icons] Customize

1 / 1 [Navigation Icons] 47% [Zoom In/Out] [Previous/Next Page] [Tools] Fill & Sign Comm

Model Tree

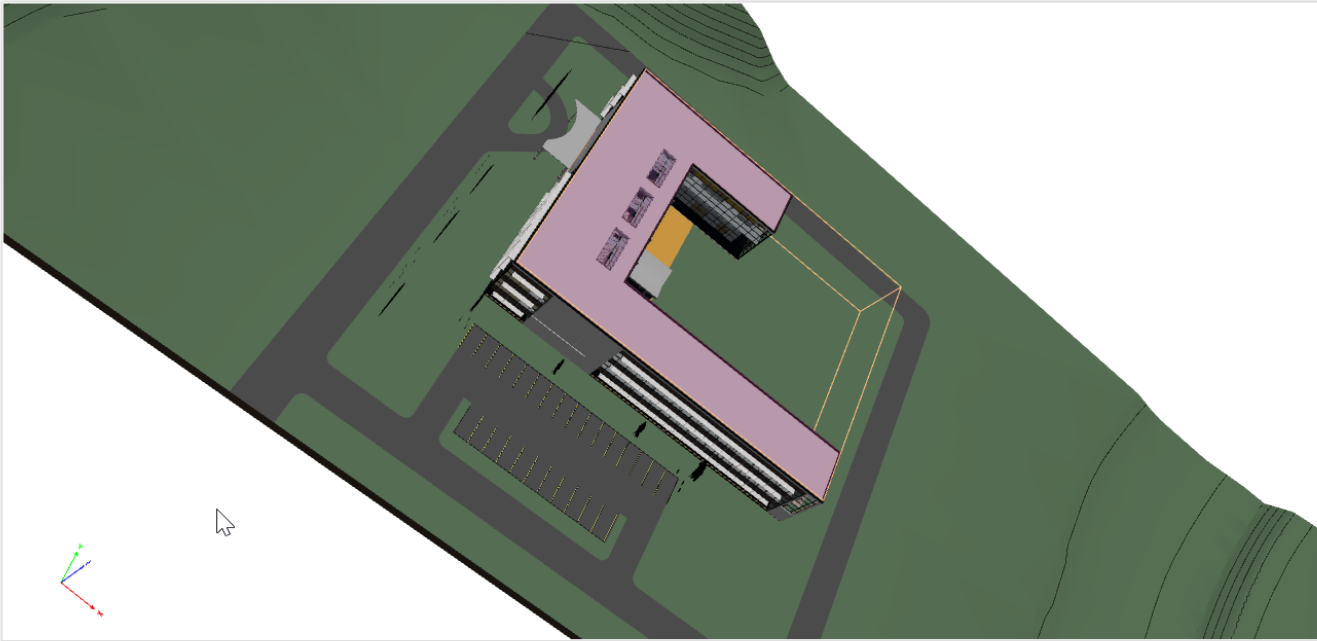
- Planting / Golden Fla
- Planting / Golden Fla
- Planting / Golden Fla
- Planting / Golden Fla
- Planting / Golden Fla
- Mass / Main Buildir**

Options

- Left
- Top

Property	Value
-1/-1	
Category/Category	Mass
Gross Floor Area/Gross Fl...	
Gross Surface Area/Gross	

TETRA^{4D} OFFICE BUILDING - AEC



Tetra4D.com CONVERT. COLLABORATE. CONTR

PDF 3D – deaktivierte Ebenen im Model Tree

3dpdf-AEC-OfficeBuilding.pdf - Adobe Acrobat Pro

File Edit View Window Help

Create [Icons] Customize

1 / 1 [Navigation Icons] 47% [Zoom In/Out] [Tools] Fill & Sign Comm

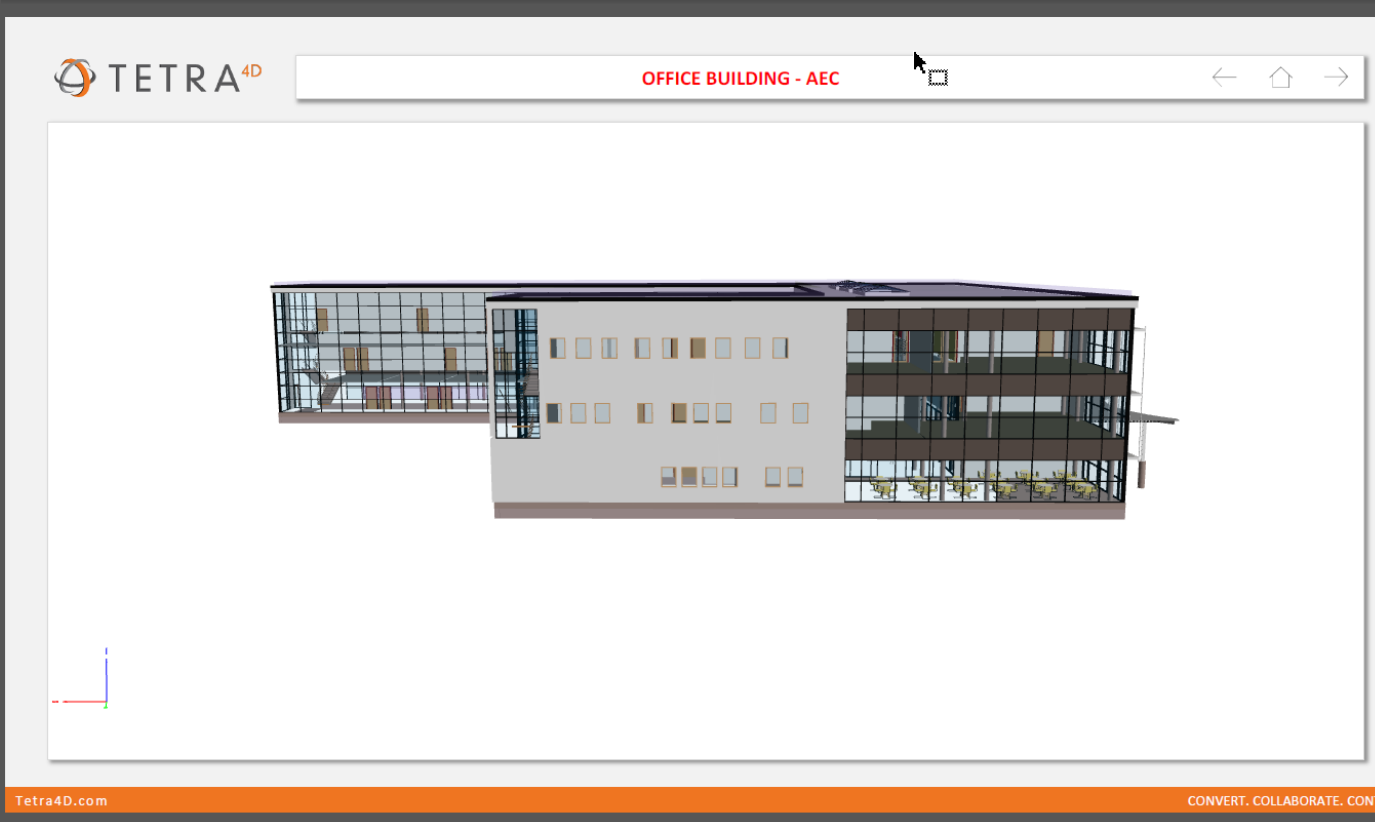
Model Tree

- Doors / M_Curtain W ^
- Curtain Wall Mullions
- Curtain Wall Mullions
- Curtain Wall Mullions
- Curtain Wall Mullions
- Curtain Wall Mullions
- Curtain Wall Mullions

Options

- Home
- Right
- Left

Property	Value
Node Type	Faceted
Node Info	Tessella
Node Tessellation	Read
Number of Lights	



Tetra4D OFFICE BUILDING - AEC

Tetra4D.com CONVERT. COLLABORATE. CONTR

PDF 3D – PDF/E

PDF/E - ISO 24517-1:2008
 → vor ISO 32000:2008, also basierend auf PDF 1.6

Informationsverlust für BIM
 → z.B. Projektplanung, Kosten

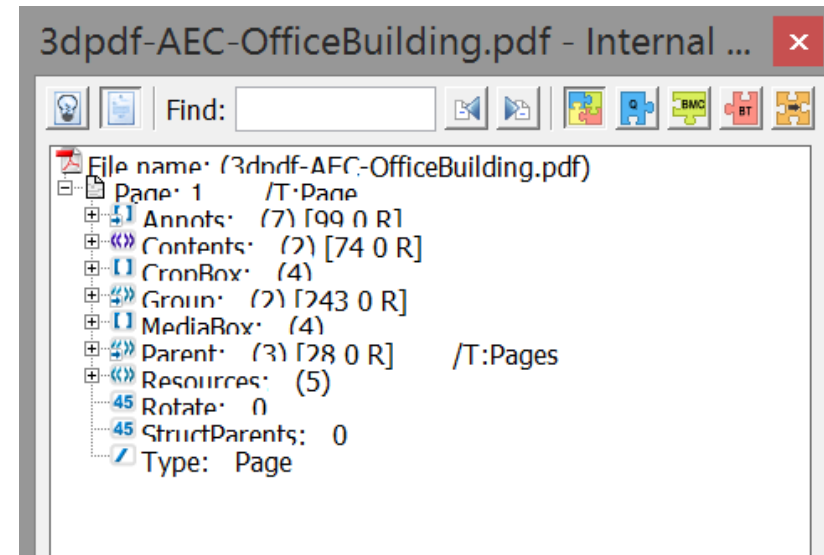
Significant properties

- Page objects als klassischer Counter für Anzahl Seiten in PDFs
- PDF/E:
 1 page object
 n Seiten

→ Wir brauchen eine profilbezogenere Herangehensweise an PDF

```

1 %PDF-1.7
2 %ääİÓ
3 1309 0 obj
4 <</Filter/FlateDecode/First 9/Length 271/N 2/Type
5 hPLijf0DCZIÄ_ä>Á'ESCÖyU(©)EÆ~h?HHDŠ³;tL#š,{ÜY~ÖöC'
6 š""æt~FÄŠö³f) °ESCACKmf^m+ÑACKEACK*1tö¼CINAKIEOI
7 jÖTmšlQqDLLEÄ±+EONIDOLly)@ •R²-Äd•iugdSUI]´Eiaâ=GÊii
8 NUL&^ACKEESC/CAÖEBSIEONw×Ñ...¥Û!âyUS-ä~NULIEIXINULy
9 endstream
  
```



Also, liebe Langzeitarchivierung, welches Format soll es denn nun sein?

Native CAD Files – AutoCAD, CATIA et al:

- + Abdeckung aller Informationsebenen (...innerhalb der gleichen SW Version)
- proprietäre, geschlossene Systeme / Formate

Open Data Exchange Formats – IFC:

- + sehr geringer Informationsverlust für BIM Objekte
- (noch?) angewiesen auf Qualität der Import / Export Schnittstellen der CAD Anbieter → oftmals sogar bei Roundtripping Verlust auf Geometrieebene

View Formats – PDF/E:

- + stabiles und verbreitetes Dateiformat
- vergleichsweise hoher Informationsverlust

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK




**Und jetzt – Diskussion!
Fragen? Anmerkungen!**

Kontakt:

M. Lindlar – TIB Hannover

 Michelle.lindlar@tib.eu

 0511 762 19826

 Lindlarm

 mickylindlar



**Welttag Digitale
Erhaltung**

29. November 2018

Wolfgang Kresse | Erhaltung von Geodaten (ISO 19165-1 und ISO/TC 211)

Prof. Dr. -Ing. Wolfgang Kresse berichtet über den Standard ISO 19165-1:2018 „Preservation of digital data and metadata – Part 1: Fundamentals“, an dessen Entstehung er maßgeblich beteiligt war. Der Standard entstand 2015-2017 im Technical Committee 211 der ISO auf Betreiben der nationalen Vermessungsbehörden und des Open Geospatial Consortium. Vorarbeiten waren von einer europäischen Arbeitsgruppe der EuroSDR geleistet worden. Die deutschen KLA-AdV Leitlinien zur bundesweit einheitlichen Archivierung von Geobasisdaten können als nationale Ausprägung dieser Norm verstanden werden.

Geodaten werden in Form von Vektordaten (Kataster, Landschaftsmodelle) oder von Rasterdaten (Laserscan, Luftbild, sonstige Fernerkundungsdaten) hergestellt. Die ISO-Norm macht sich die grundlegenden Normen der Geoinformatik zunutze und ergänzt neue Metadatenklassen, die sich an OAIS (ISO 14721) orientieren. Sie definiert geospezifische Ausprägungen der Konzepte von AIP, SIP und DIP und schreibt dafür die Anwendung der ISO 29500-2 (Open Packaging Conventions) vor. Sie definiert die Eigenschaften von Geodaten und Strategien wie Historisierung und das Bilden von Zeitschnitten. Ein Konzept, das einen erheblichen Verlust an Eigenschaften in Kauf nimmt, ist der Ansatz der „Gold Copy“, eine Archivierung in einfachsten Formaten, bei der Abhängigkeiten von technischen Umgebungen wie Datenbanken möglichst vermieden werden.

Als nächstes entsteht Teil 2 der Norm (ISO 19165-2 Content specifications for earth observation data). In diesem Abschnitt soll definiert werden, welche Formate für eine dauerhafte Aufbewahrung der verschiedenen Geodatentypen zu erzeugen sind.

Diskussion und Fragen:

1. Peter Worm: Wann ist mit der Fertigstellung des ISO-Standards zu rechnen?

A: Die Veröffentlichung des Standards ist 2018 erfolgt.

2. Eszter Kiss: Am 19. März 2019 findet eine Infoveranstaltung zur Geodatenlangzeitspeicherung statt, welche vom Staatsarchiv Hamburg gesteuert wird. Adresse und weitere Informationen werden an die Teilnehmerinnen und Teilnehmer des Workshops weitergeleitet. Am 13./14. März findet an der Universität Regensburg ein Workshop zur Digitalisierung von Karten in Bibliotheken statt.

3. Kai Naumann: Gibt es Tools zur Validierung des Inhalts eines Shapefiles?

A: Im Nachgang zur Tagung konnte Wolfgang Kresse einen Kontakt zu Esri herstellen, dem Erfinder des Esri Shapefiles. Auch dort sind keine Mechanismen zur Validierung des Inhalts eines Shapefiles bekannt.



Erhaltung von Geodaten (ISO 19165-1)

Wolfgang Kresse
kresse@hs-nb.de



Gliederung

Motivation für die Entwicklung der ISO 19165-1

ISO/TC 211 *Geographic information / Geomatics*

ISO 19165-1 *Preservation of digital data and metadata*

Nächste Schritte

Motivation 1: Rahmenbedingungen heutiger Datenspeicherung

Speicherung von fast nur noch digitalen Daten

Haltbarkeit digitaler Speichermedien sehr begrenzt

Digitale Daten meist nicht selbsterklärend

(Metadaten, Leseprogramme usw. erforderlich)

Drohender künftiger Rückblick auf heute:

“Digital Dark Age”

Motivation 2:

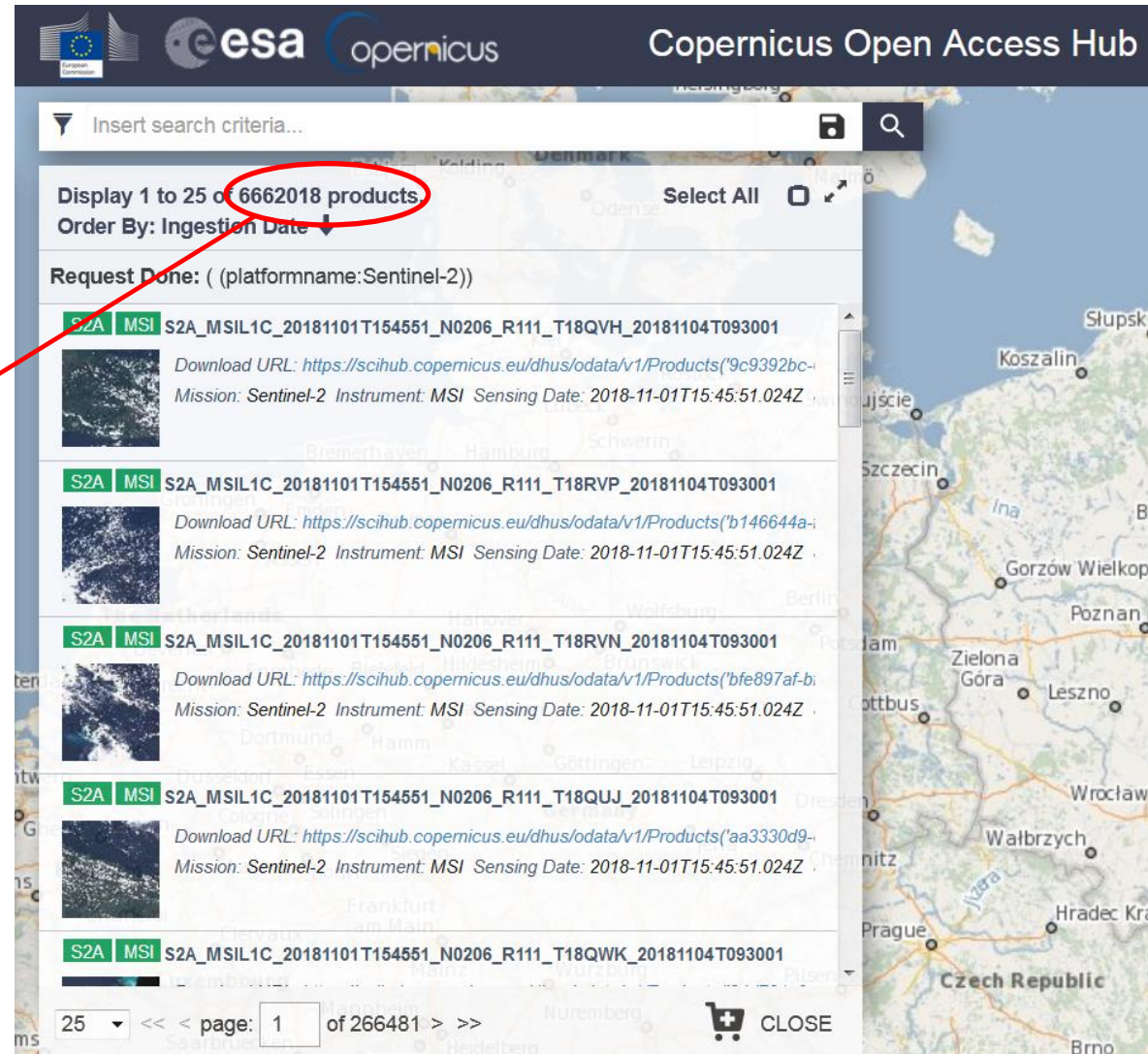
Derzeit über 100 abbildende
Satellitensysteme im Orbit

Hier: Sentinel 2A (2015)
und 2B (2017)

Freies Datenangebot:
6.662.018 Szenen mit je
knapp 1 GByte Daten

Überschlägige Summe:

5,4 PByte



esa opencopernicus Copernicus Open Access Hub

Insert search criteria...

Display 1 to 25 of **6662018** products

Order By: Ingestion Date

Select All

Request Done: ((platformname:Sentinel-2))

S2A	MSI	Product ID	Download URL	Mission	Instrument	Sensing Date
		S2A_MSIL1C_20181101T154551_N0206_R111_T18QVH_20181104T093001	https://scihub.copernicus.eu/dhus/odata/v1/Products('9c9392bc-...	Sentinel-2	MSI	2018-11-01T15:45:51.024Z
		S2A_MSIL1C_20181101T154551_N0206_R111_T18RVP_20181104T093001	https://scihub.copernicus.eu/dhus/odata/v1/Products('b146644a-...	Sentinel-2	MSI	2018-11-01T15:45:51.024Z
		S2A_MSIL1C_20181101T154551_N0206_R111_T18RVN_20181104T093001	https://scihub.copernicus.eu/dhus/odata/v1/Products('bfe897af-...	Sentinel-2	MSI	2018-11-01T15:45:51.024Z
		S2A_MSIL1C_20181101T154551_N0206_R111_T18QUJ_20181104T093001	https://scihub.copernicus.eu/dhus/odata/v1/Products('aa3330d9-...	Sentinel-2	MSI	2018-11-01T15:45:51.024Z
		S2A_MSIL1C_20181101T154551_N0206_R111_T18QWK_20181104T093001		Sentinel-2	MSI	

25 << < page: 1 of 266481 > >>

CLOSE

Grundlagen der ISO 19165-1

Preservation of digital data and metadata – Part 1: Fundamentals (1/2)

Wunsch der AdV

Ende 2013



Markus Seifert, AdV, München

Entwicklungen des Open Geospatial Consortium (OGC)

Internationale Standardisierung für Geoinformation (nicht Normung)
über 500 Mitglieder (Softwarehersteller, Anwender, Universitäten)
Leitung in den USA



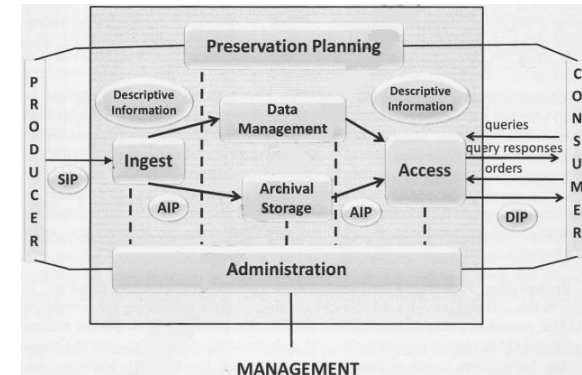
Joan Masó, OGC, Barcelona

Grundlagen der ISO 19165-1

Preservation of digital data and metadata – Part 1: Fundamentals (2/2)

ISO 14721:2012 *Space data and information transfer systems – Open archival information system (OAIS) – Reference model*

Consultative Committee for Space Data Systems:
NASA, ESA,
Arbeiten z.B. SCID (spacecraft identification)
SIP, AIP, DIP
(Submission / Archival / Dissimination
Information Package)



GI+100: Long term preservation of digital Geographic Information – 16 fundamental principles agreed by National Mapping Agencies and Archives

UK, D, CH, S, N

Leitlinien zur bundesweit einheitlichen Archivierung von Geobasisdaten

ISO/TC (Technical Committee) 211 *Geographic information / Geomatics*



Olaf Østensen, Alan Morrison, Bjørnhild Saeterøy
Lawrence D. Eicher Leadership Award, 2010



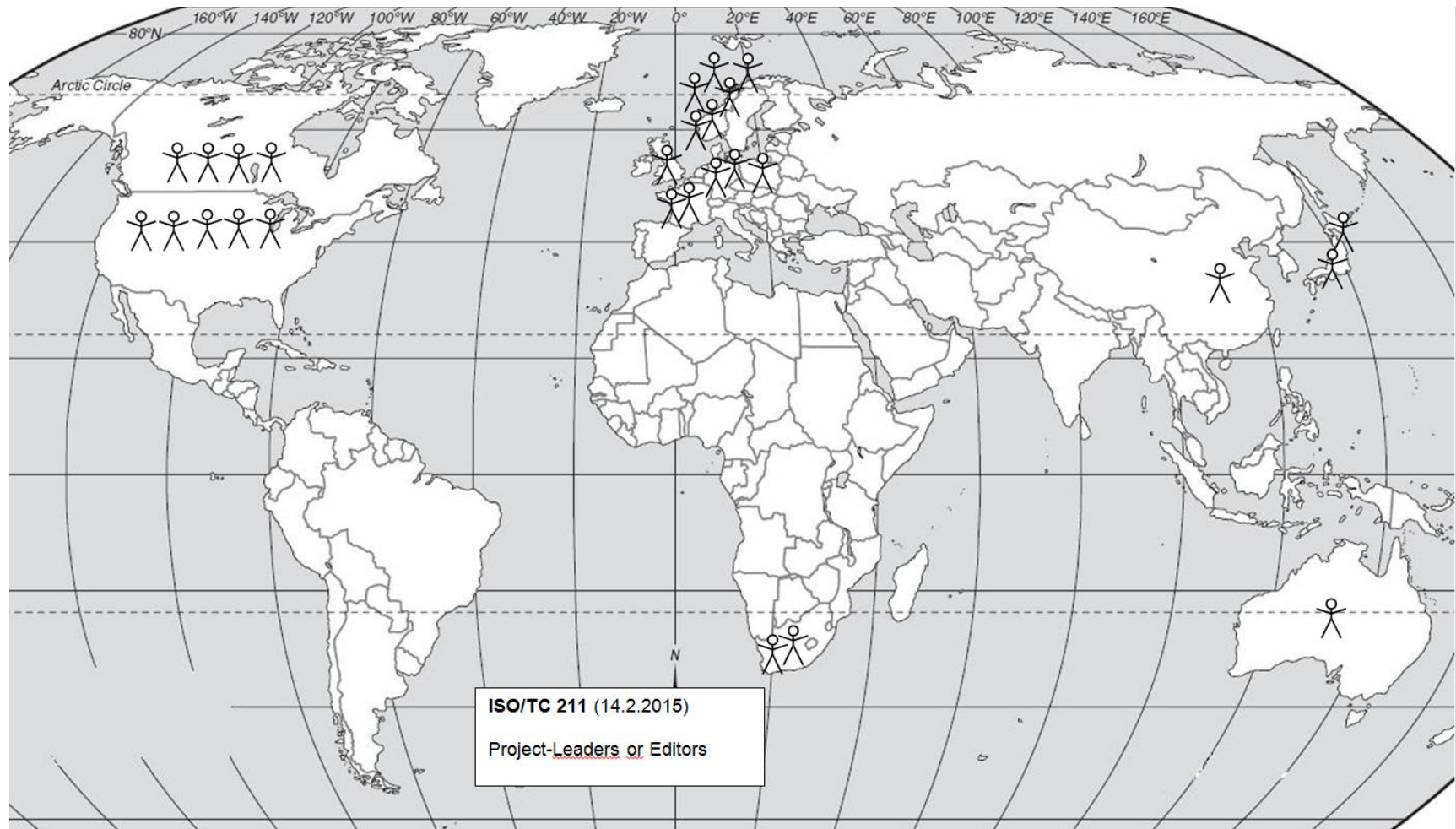
Christina Wasström, Mats Åhlin
2018



Sitzung des ISO/TC 211 im Juni 2016 in Tromsø, Norwegen

69,6494°N 18,9587°E

Projektleiter und Editoren beim ISO/TC 211 (2015)



ISO/TC 211: Grundlegende Normen (Auswahl)

Conceptual schema language (Objektorientierte Modellierung von Geodaten)

Spatial schema (Vektorgeometrie)

Temporal schema (zeitlich veränderliche Geodaten)

Methodology for feature cataloguing (Objektartenkataloge)

Referencing by coordinates (Koordinatenreferenzsysteme)

Metadata

Portrayal (Kartographie)

Web Map Server und *Web Feature Service*

Imagery sensor models for geopositioning (Georeferenzierung von Bildern)

Geography Markup Language (GML, Vektordatenformat)

Data quality

ISO/TC 211: Ergänzende Normen (Auswahl)

Geospatial Digital Rights Management Reference Model (Rechte an Geodaten)

Addressing (Adressverwaltung, z.B. für Postverkehr)

Ontologies (Formalisierung von Inhalten, z.B. vor Aufbau einer Datenbank)

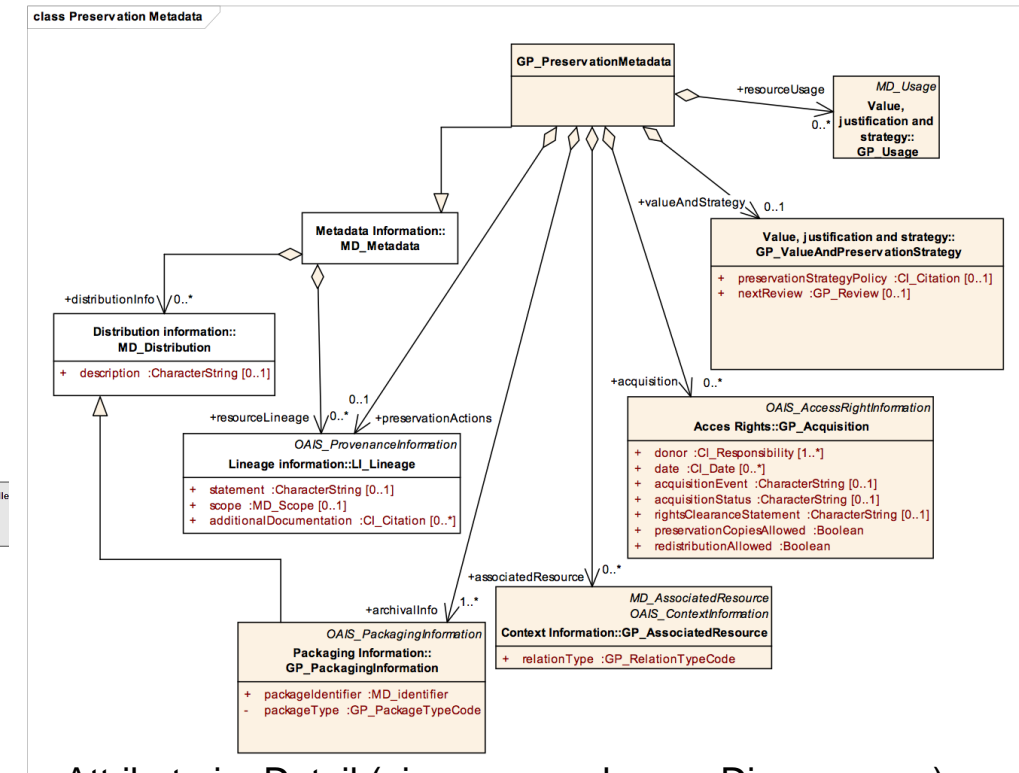
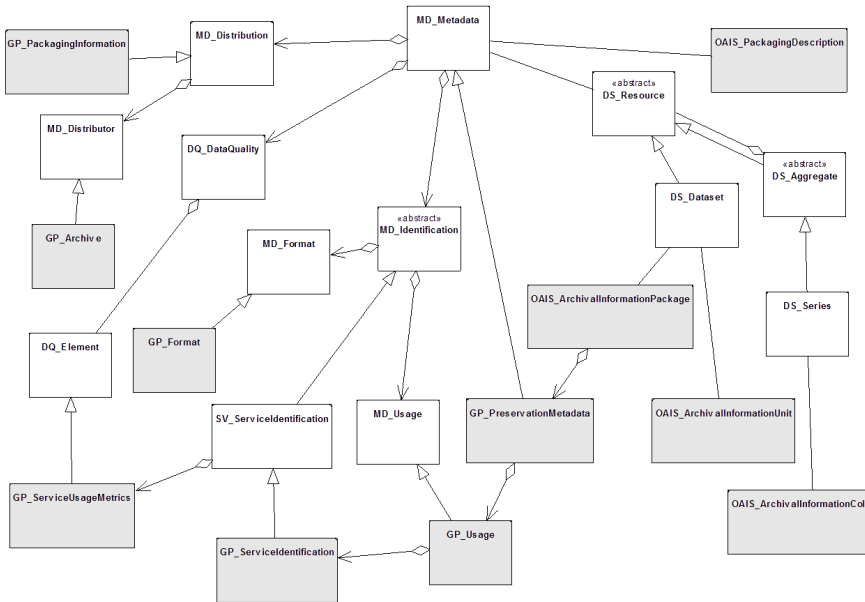
Preservation of digital data and metadata (Langzeitarchivierung)

ISO 19165-1: Lösungen für Geodaten (1/5)

Integration ins Modell der Geoinformatik-Normen

Archivierung ist Spezialfall von Metadaten

Weiterverwendung von vorhandenen Klassen (Attributen) soweit wie möglich



Attribute im Detail (eines von mehreren Diagrammen)

weiß: vorhandene Klassen
grau: neue Klassen der ISO 19165-1

ISO 19165-1: Lösungen für Geodaten (2/5)

Übernahme und Weiterentwicklung der Konzepte von ISO 14721:2012

Spezialisierte Formen der Informationspakete: GeoSIP, GeoAIP, GeoDIP, aber getrennte Speicherung von Daten und Metadaten empfohlen

Getrennte Speicherung oft notwendig bei Arbeiten übers Internet

ISO/IEC 29500-2:2012 Open Packaging Conventions

Daten mit zip-Kompression und Beschreibung im xml-Format
(ähnlich dem bisherigen tar-Format)

ISO 19165-1: Lösungen für Geodaten (3/5)

Besonderheiten von „Geo“-Daten:

Koordinaten Karten	Vielfältige Koordinatenreferenzsysteme Kartenwerke (Blattschnitt, Nummerierung der Kartenblätter) Kartographische Ausgabe von Geodaten
Sachdaten Rechte	Vielfältige Verknüpfung zu Sachdaten Verknüpfung von Geodaten und Rechten
Redundanz, Vektor/Raster Topologie	Redundante Datenhaltung, z.B. Level-of-Detail, Bildpyramiden, Maßstabsreihen Gemeinsame Speicherung von Vektor- und Rasterdaten (1) Topologische Verknüpfungen (2)
Datenformate	Datenformate: Struktur, Eigenschaften, Metadaten, Zugriffsmöglichkeiten
Zeit	Dauer der Archivierung: short-term, long-term Viele Geodaten werden nie obsolet, z.B. Kataster. Alternative Verfahren: - Historienverwaltung - Zeitschnitt: Archivierung des gesamten Archivs
Gold-copy	Gold-copy: Separate Archivierung, einfache Speicherung (keine Datenbank)

ISO 19165-1: Lösungen für Geodaten (4/5)

(Ref. 1) Gemeinsame Speicherung von Vektor- und Rasterdaten (Beispiele)

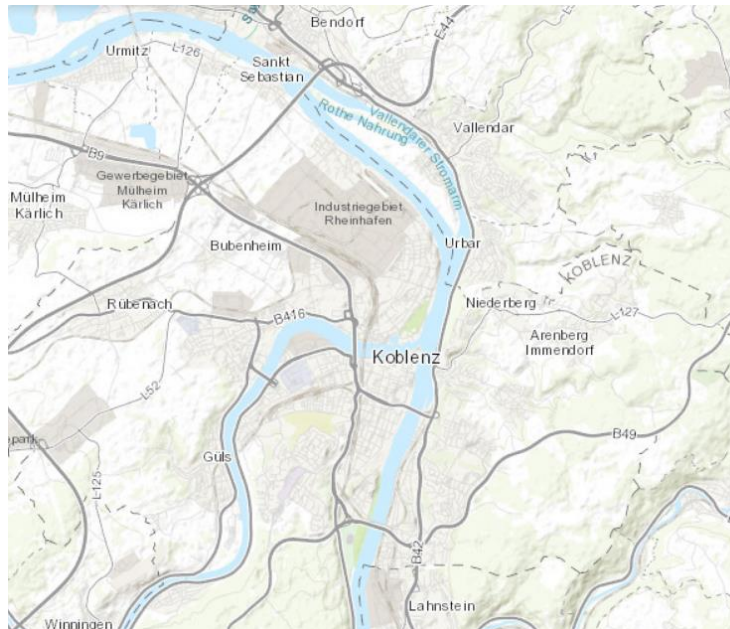
GeoTIFF:

Orthophoto mit Daten zur Georeferenzierung im Datei-Header

GML in JPEG 2000 (GMLJP2):

Orthophoto im Format JPEG 2000 (Wavelet-Komprimierung)

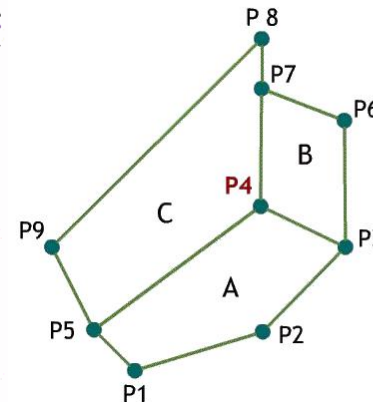
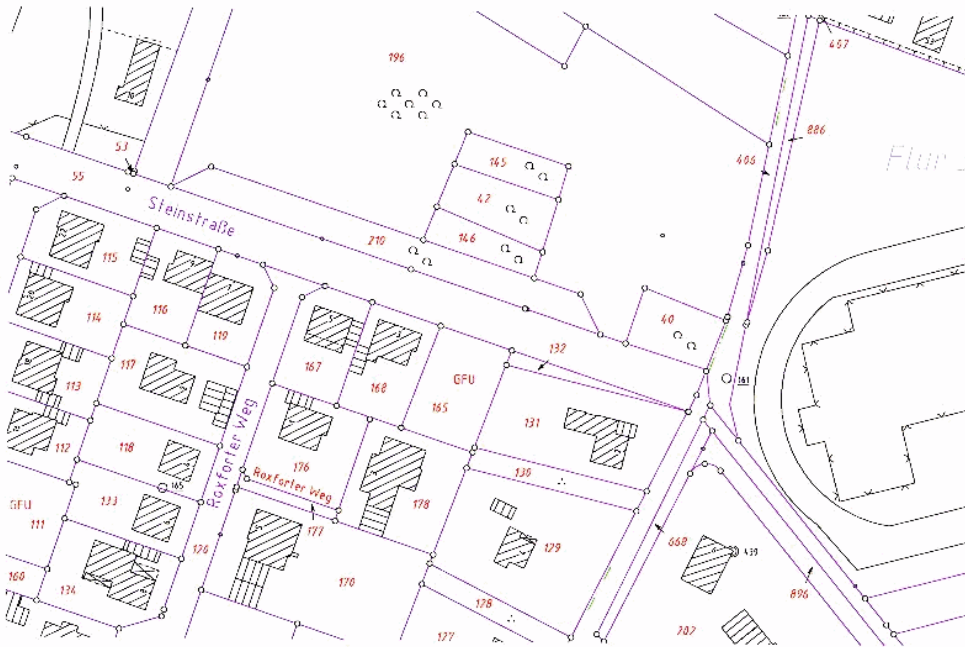
Vektor für Überlagerung im Format Geography Markup Language (GML)



ISO 19165-1: Lösungen für Geodaten (5/5)

(Ref. 2) Topologische Verknüpfungen

Beispiel: Katasterkarte (ALKIS) mit nicht-redundanter Punktstorage



Flächen:

A: P1 P2 P3 **P4** P5

B: **P4** P3 P6 P7

C: **P4** P7 P8 P9 P5

Punkte: x - Koord: y - Koord:

P1	2.0	0.0
P2	5.0	1.0
P3	7.0	3.0
P4	5.0	4.0
P5	1.0	1.0
P6	7.0	6.0
...		

Nächste Schritte

Implementierung

*ISO 19165-2 Preservation of digital data and metadata –
Part 2: Content specifications for earth observation data and derived digital products*

Detaillierte Attributlisten für Fernerkundungsdaten

Kalibrierung

Daten-Historie

Nutzerhinweise

EuroSDR

Vortrag am 15./16.11.2018 in Berlin

EuroGeographics

bisher keine Antwort

Nächste Schritte



Dave Danko
Esri
Redlands, CA
Vienna, VA



Clemens Portele
Interactive Instruments
Bonn



Brohme, ca. 1985



Vielen Dank für Ihre Aufmerksamkeit!

Rainer Jacobs | Extrahieren von Dateien aus Containern

Rainer Jacobs (Bundesarchiv) beschreibt einen Weg aus einem bekannten Dilemma, das Archivare im praktischen Umgang mit digitalen Unterlagen befällt. Dateien erweisen sich nicht nur als Träger von Texten und sonstigen Inhalten, sondern auch als Container, die wie russische Puppen weitere Container enthalten. Er stieß bei der Bewertung einer Dateisammlung einer Bundesbehörde auf zahlreiche Microsoft-Office-Dokumente, die ihrerseits wiederum Dokumente enthielten. Da letztere bei der Formatwandlung nach PDF/A keine Berücksichtigung fanden, sondern von der Konversionsplattform einfach ignoriert wurden, hat Herr Jacobs in Python ein Extraktionswerkzeug für verschachtelte Dateien programmiert. Mit dem Tool können nun im Bundesarchiv automatisiert verschachtelte Office-Dokumente zerlegt und für eine PDF/A-gerechte Bestandserhaltung hergerichtet werden. Darüber hinaus ist das Werkzeug in der Lage, Dateien hinsichtlich weiterer problematischer Eigenschaften, wie z. B. sich aktualisierender Datumfelder, zu analysieren.

Michael Tobegen | Entwicklung eines DIMAG-Bestandserhaltungsmoduls: Werkstattbericht

Michael Tobegen (Staatsarchiv Hamburg) berichtet von der Konzeption des DIMAG Bestandserhaltungsmoduls (BEM), das einen kontrollierten Rahmen für die Konversion von Dateien nach Formatklassen bieten wird. Das BEM ging aus einem Projekt zur PDF/A-Erstellung aus dem Hamburgischen E-Akten-System ELDORADO hervor, das aber in dieser Form nicht realisiert wurde. Stattdessen wurden die angestrebten Funktionen in eine übergreifende, für alle Partner des DIMAG-Verbands benutzbare Konversionsplattform integriert, die nicht nur PDF/A als Zielformat ermöglicht, sondern sich als Framework grundsätzlich dazu eignet alle denkbaren Informationstypen (z.B. Textdokumente, Töne, Bewegtbilder, Geodaten) zu verarbeiten.

Für eine Formatmigration wird zunächst das Format der zu bearbeitenden Dateiobjekte festgestellt (Formaterkennung). Diese durchlaufen in definierten Gruppen den Konversionsprozess, an dessen Ende eine Formatvalidierung erfolgt, um den Erfolg der Konversionen messen zu können.

Die Konzeptionierung ist bereits abgeschlossen, realisiert wird das BEM bis Sommer 2019 von ITEOS, einem öffentlich-rechtlichen Rechenzentrum mit Hauptsitz in Stuttgart. Die erste Version des BEM wird kostenfreie Drittanwendungen implementiert haben. Individuelle Bedarfe der nutzenden Archive müssen durch diese selbst umgesetzt werden.

DIGITALE ARCHIVIERUNG DAN

Entwicklung eines DIMAG- Bestandserhaltungsmoduls

Ein Werkstattbericht

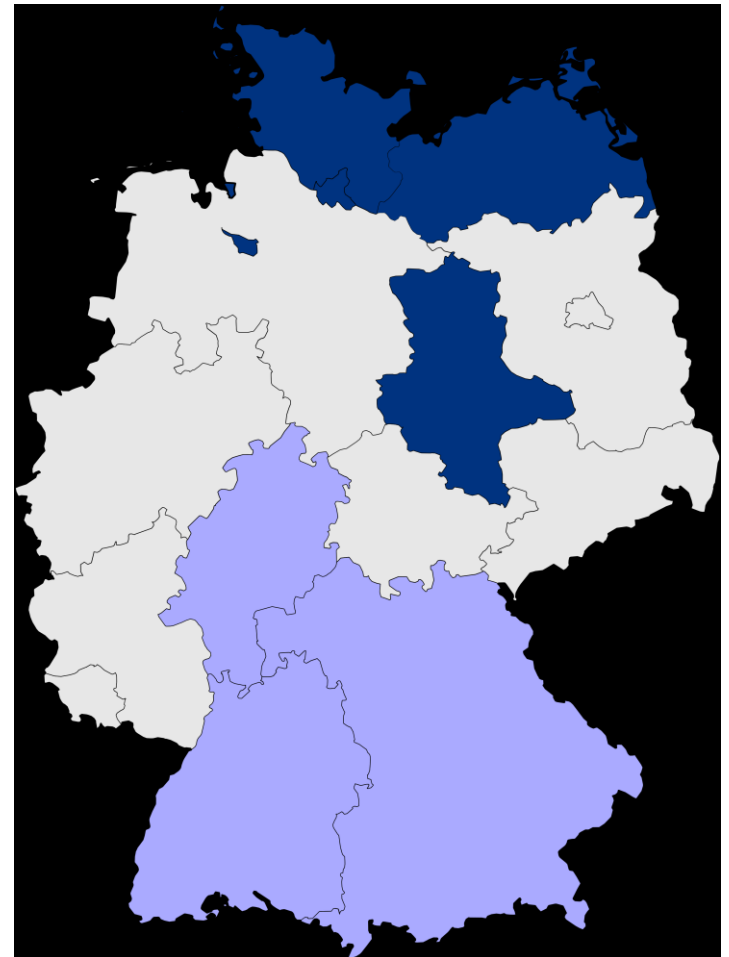
Michael Tobegen

AGENDA

- 01 Entstehungskontext
- 02 Funktionen des Bestandserhaltungsmoduls
- 03 Nachnutzung
- 04 Zeitplan

1. ENTSTEHUNGSKONTEXT

- Ursprüngliche Anforderung: Migrationslösung für das Hamburger DMS
- Vorstellung und Diskussion dieser Idee im DIMAG-Verbund
- Heute: BEM ist ein weiterer Beitrag des DAN-Verbundes in den DIMAG-Verbund
- Entwicklung findet innerhalb einer kleinen Arbeitsgruppe statt
- Verantwortlich für die Umsetzung: ITEOS GmbH



2. FUNKTIONEN: GRUNDSÄTZLICHES

- BEM als erweiterbares Framework
- Wird als Serveranwendung entwickelt
- „Grundausstattung“: Migration von Office-Dateien, PDF-Dateien und Bilddateien nach PDF/A-1b und PDF/A-2b
- Manuelle und automatisierte Abläufe vorgesehen
- Drei Hauptfunktionen
 - Formaterkennung
 - Formatmigration
 - Formatvalidierung

2. FUNKTIONEN: FORMATERKENNUNG

- Migration anhand einer Format Policy
- Einbindung der DIMAG-Komponente IngestList
- Dadurch Verwendung von DROID
- Schnittstelle noch zu definieren

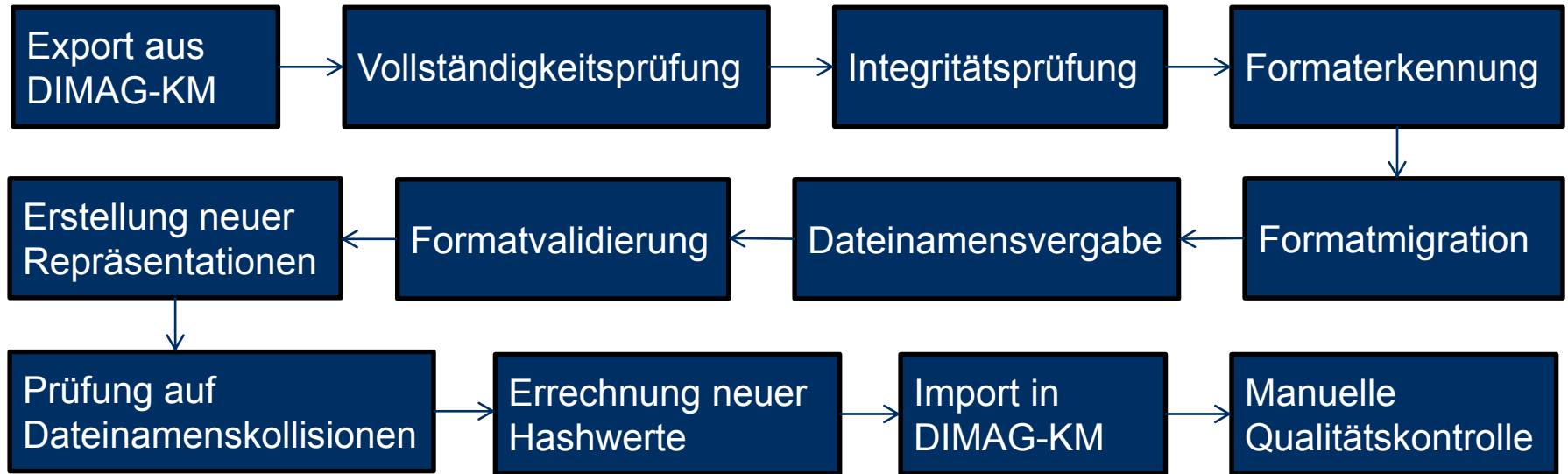
2. FUNKTIONEN: FORMATMIGRATION

- Besondere Herausforderungen:
 - Art der An- bzw. Einbindung
 - Kommerzielle vs. kostenlose Anwendungen
 - Qualität der Anwendungen
- Hier wird es innerhalb der Anwenderschaft den größten Spielraum für individuelle Anpassungen geben

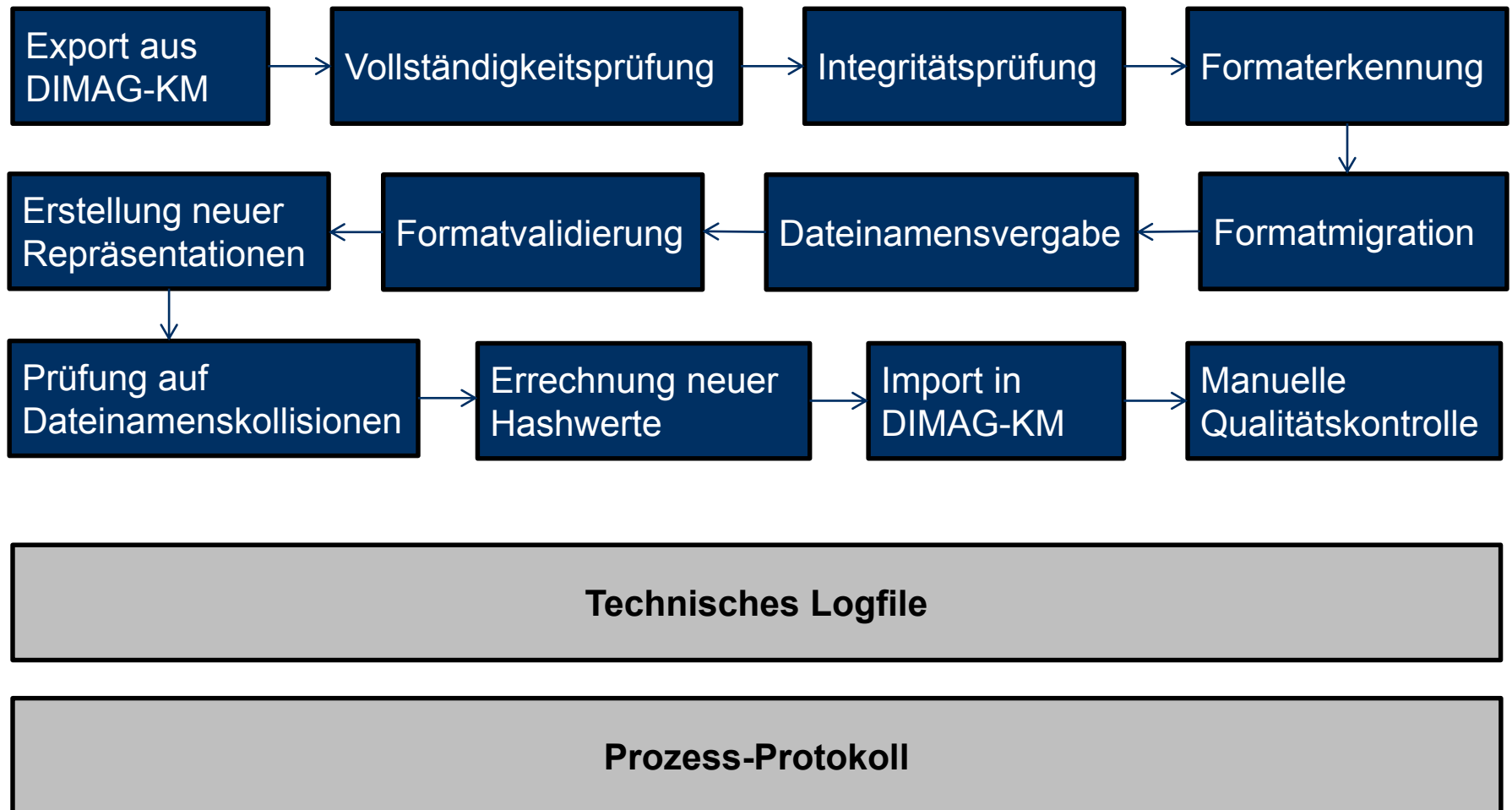
2. FUNKTIONEN: FORMATVALIDIERUNG

- Perspektivisch Einbindung der DIMAG-Komponente IngestList
- Zunächst individuelle Lösung für Validierung von PDF/A

2. FUNKTIONEN: WORKFLOW

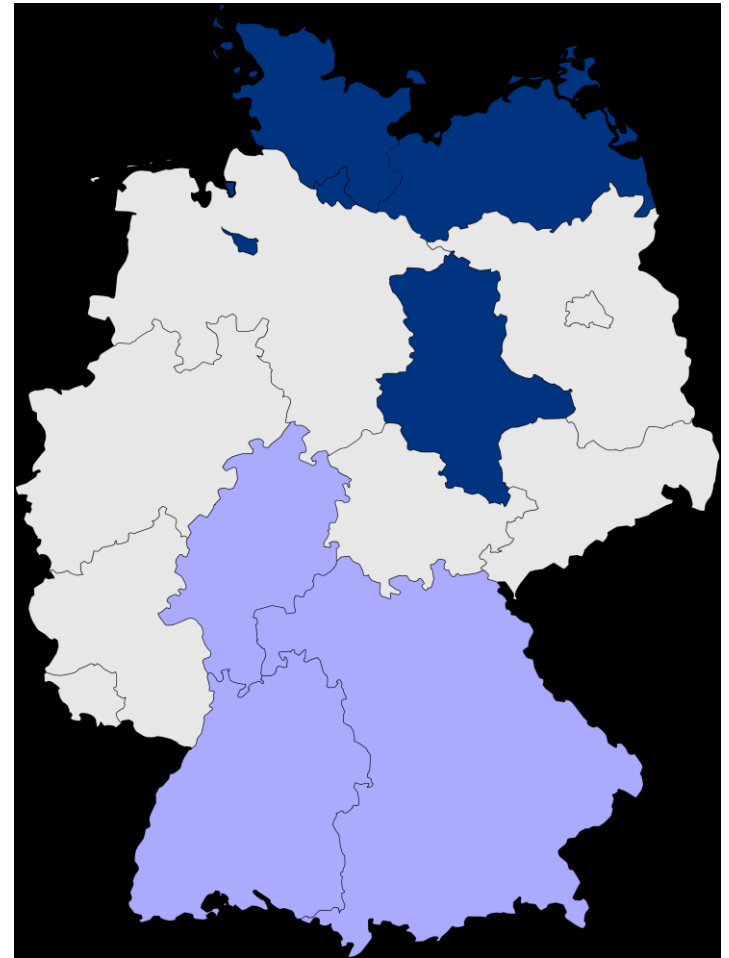


2. FUNKTIONEN: WORKFLOW

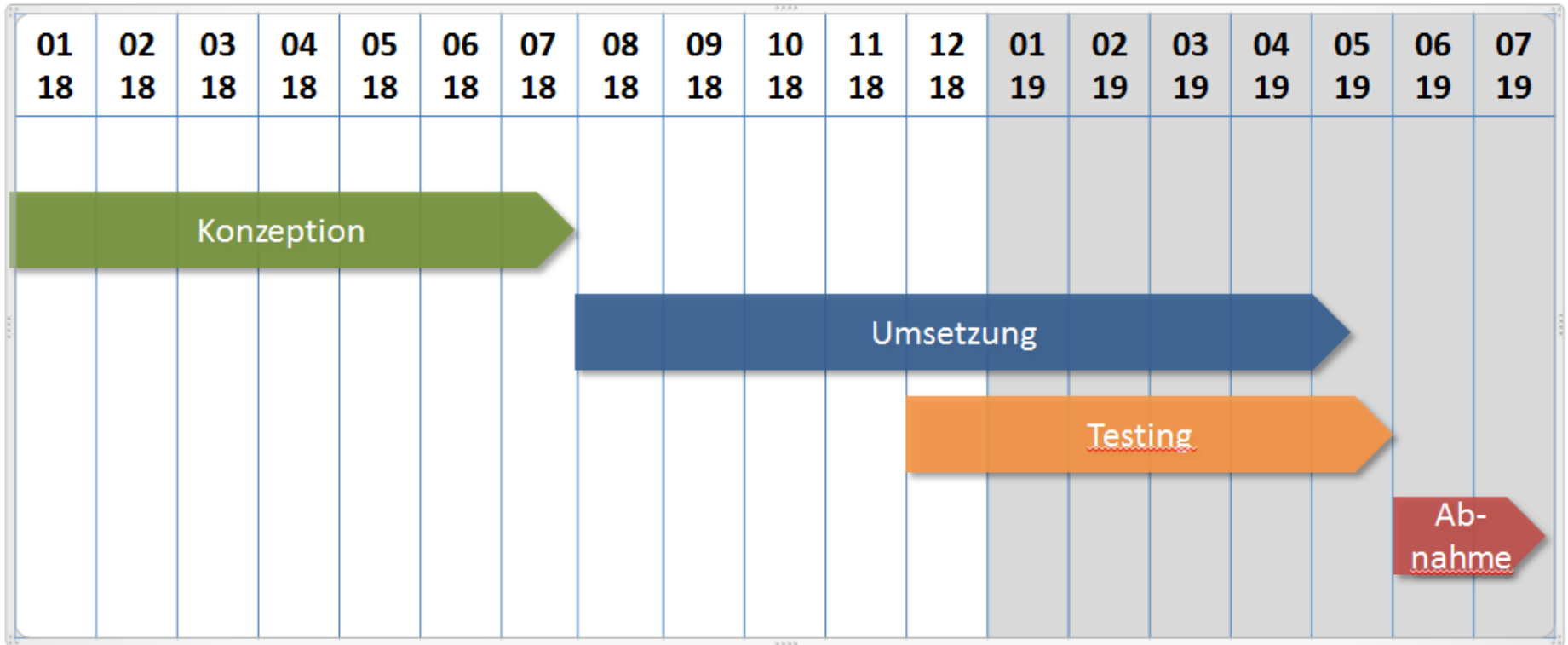


3. NACHNUTZUNG

- BEM wird Bestandteil des DIMAG-Software-Pakets
- Somit ist jeder DIMAG-Anwender berechtigt, auch das BEM zu nutzen
- Kosten für Betrieb, Pflege und Erweiterungen tragen die Anwender selbst



4. ZEITPLAN



How Projects Really Work (version 1.0)

Create your own cartoon at www.projectcartoon.com

**VIELEN DANK
FÜR IHRE
AUFMERKSAMKEIT.**



How the customer explained it



How the project leader understood it



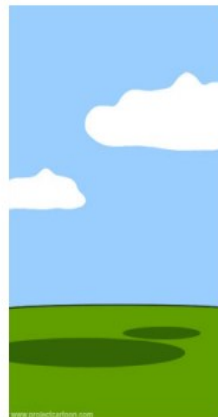
How the analyst designed it



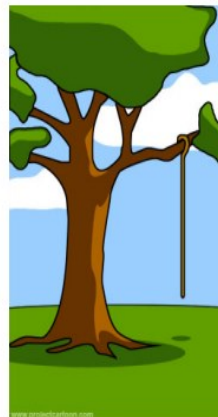
How the programmer wrote it



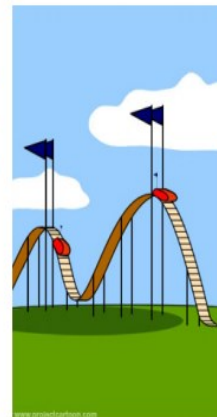
How the business consultant described it



How the project was documented



What operations installed



How the customer was billed



How it was supported



What the customer really needed

<http://www.projectcartoon.com/cartoon/3>