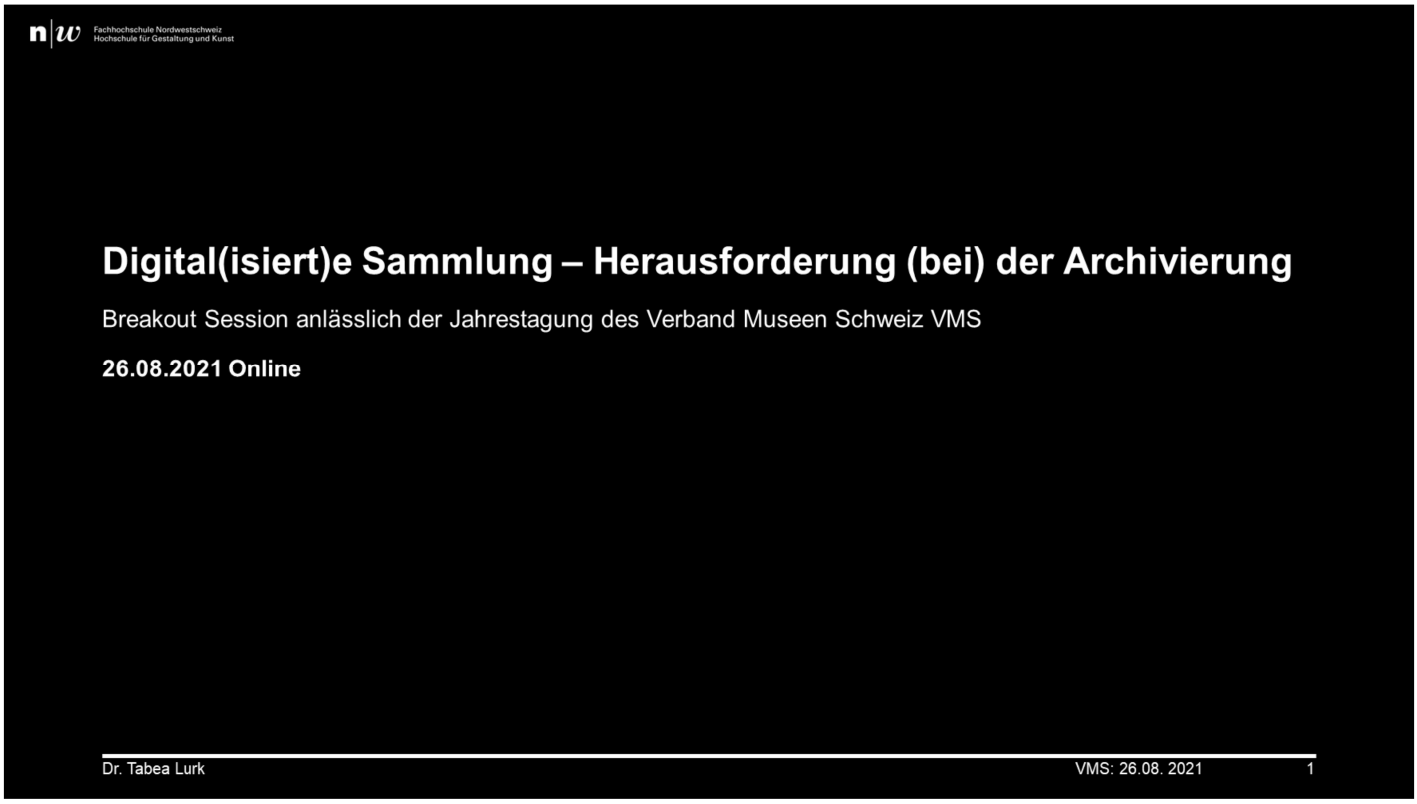
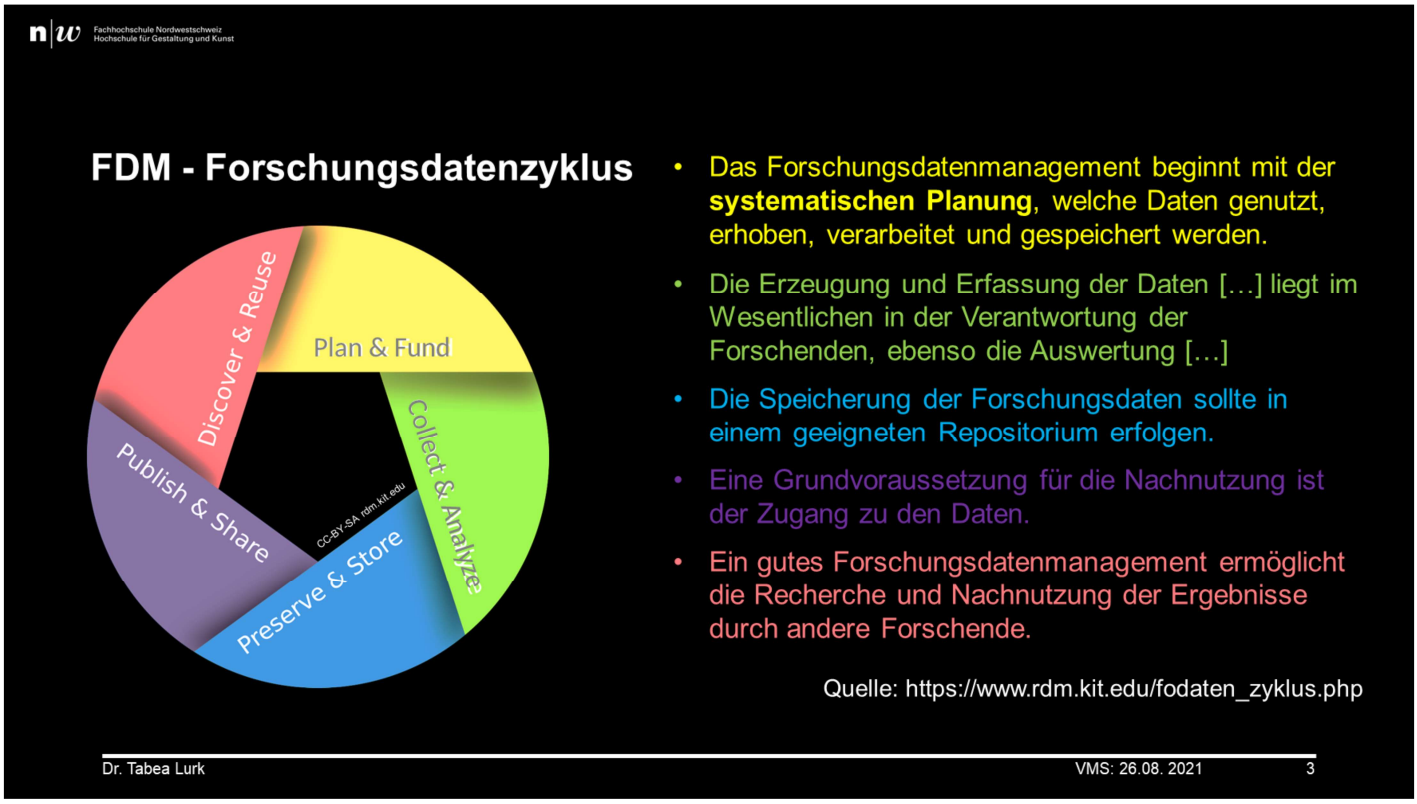


Gegenstand der Breakout Session zu digital(isiert)e Sammlung sind Herausforderung (bei) der Archivierung im musealen Kontext. Neben allgemeineren, konzeptionell-administrativen Fragestellungen wird der Fokus auf Aspekte der alltäglichen Archivierungspraxis gelegt. Schwerpunkte sind dabei u.A. Teilautomatisierungen bei der Dokumentation, der Konsistenzprüfung und weiteren Elemente des klassisch-digitalen Preservation Managements.

Das Format der Breakout Session legt es nahe, dass die Teilnehmenden sich mit eigenen Fragen und Anregungen einbringen. Gern dürfen Sie entsprechende Anliegen vorbereiten, um diese zur Diskussion zu stellen.

Im Rahmen des Wrap-Ups wird ein knapper Ausblick auf aktuelle Entwicklungen in der Schweiz gegeben. Das kann dazu beizutragen, besser einschätzen zu können, wo jeder/jede Sammlung für sich derzeit steht, wo Desiderate bestehen und was nächste Schritte zur Optimierung oder Verstetigung der eigenen Archivierungspraktiken sein können.

Im Zentrum der vorliegenden Präsentation steht die Entwicklung eines Modells für das Datenmanagement im musealen bzw. Sammlungszusammenhang.



Es gibt diverse unterschiedliche Modelle für das Forschungsdatenmanagement. Ich persönlich schätze den generischen Ansatz des KIT, der hier exemplarisch abgebildet ist.

Diese Managementmodelle werden dann je nach Einrichtung in digitalen Systemen abgebildet, wobei das Ziel eines dokumentierten Datenmanagements darin besteht, möglichst nachhaltig zu funktionieren. Im Rahmen der Forschung spielt z.B. neben der Dokumentation auch noch die Nachnutzung eine wichtige Rolle.

Basis-Forschungszyklus KIT: https://www.rdm.kit.edu/fodaten_zyklus.php.

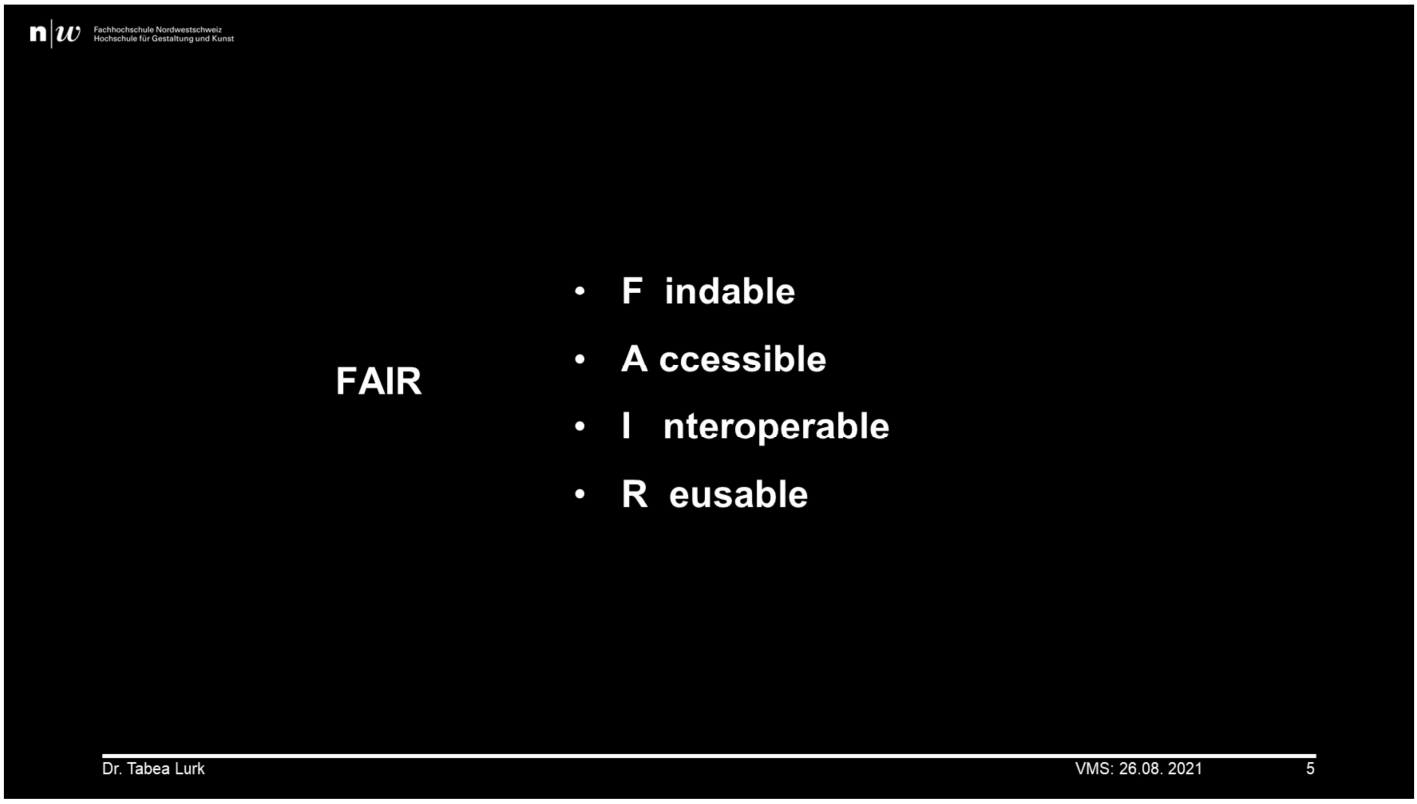
Vertrauen bzw. die Vertrauenswürdigkeit von Systemen lässt sich durch bestimmte Workflows erhöhen/sichern, wozu es mittlerweile zertifizierte Verfahren gibt.

Exemplarisch sei hier das CoreTrustSeal genannt

https://www.coretrustseal.org/about/. Das Zertifikat enthält die beiden im deutschsprachigen Kontext verbreiteten Zertifikate nestor-Seal DIN 31644 und ISO 16363, welche die Kriterien für Vertrauenswürdige Archive definiert.

Ein anderes Framework, auf das ich kürzlich gestossen bin und das im kulturellen Kontext von Sammlungen besonders wichtig erscheint, ist das TRUST-Framework von Lin et al. (2020), das die Verantwortlichkeiten starker in den Blick nimmt. Tatsächlich können Policy-basierte Veränderungen etwa in der Sammlungsstrategie zu ernsthaften Gefährdungen für kulturelle Sammlungen warden, sodass ich eine Auseinandersetzung mit diesem Framework nur empfehlen kann.

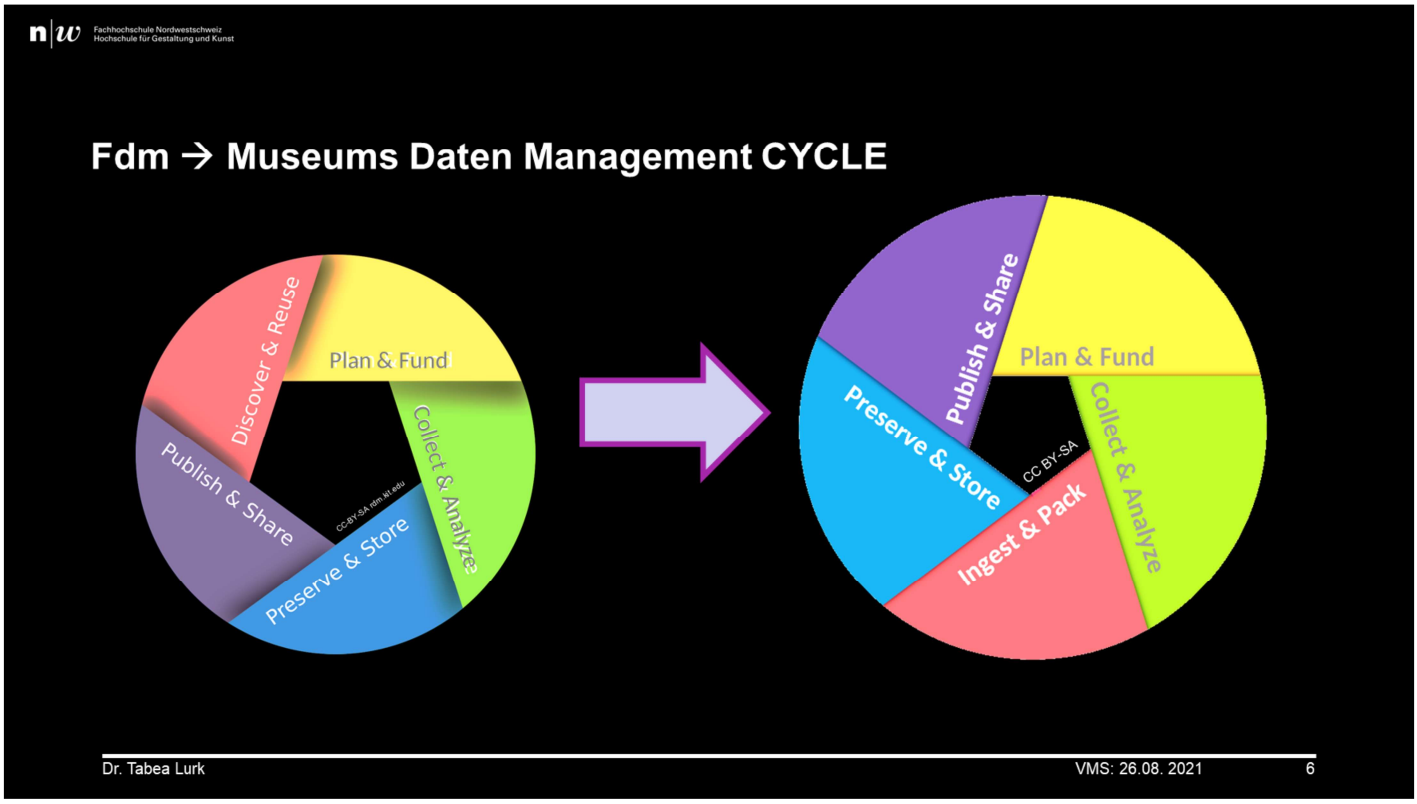
Lin, Dawei, Jonathan Crabtree, Ingrid Dillo, Robert R. Downs, Rorie Edmunds, David Giaretta, Marisa De Giusti, u. a. „The TRUST Principles for Digital Repositories“. *Scientific Data* 7, Nr. 1 (14. Mai 2020): 144. https://doi.org/10.1038/s41597-020-0486-7.



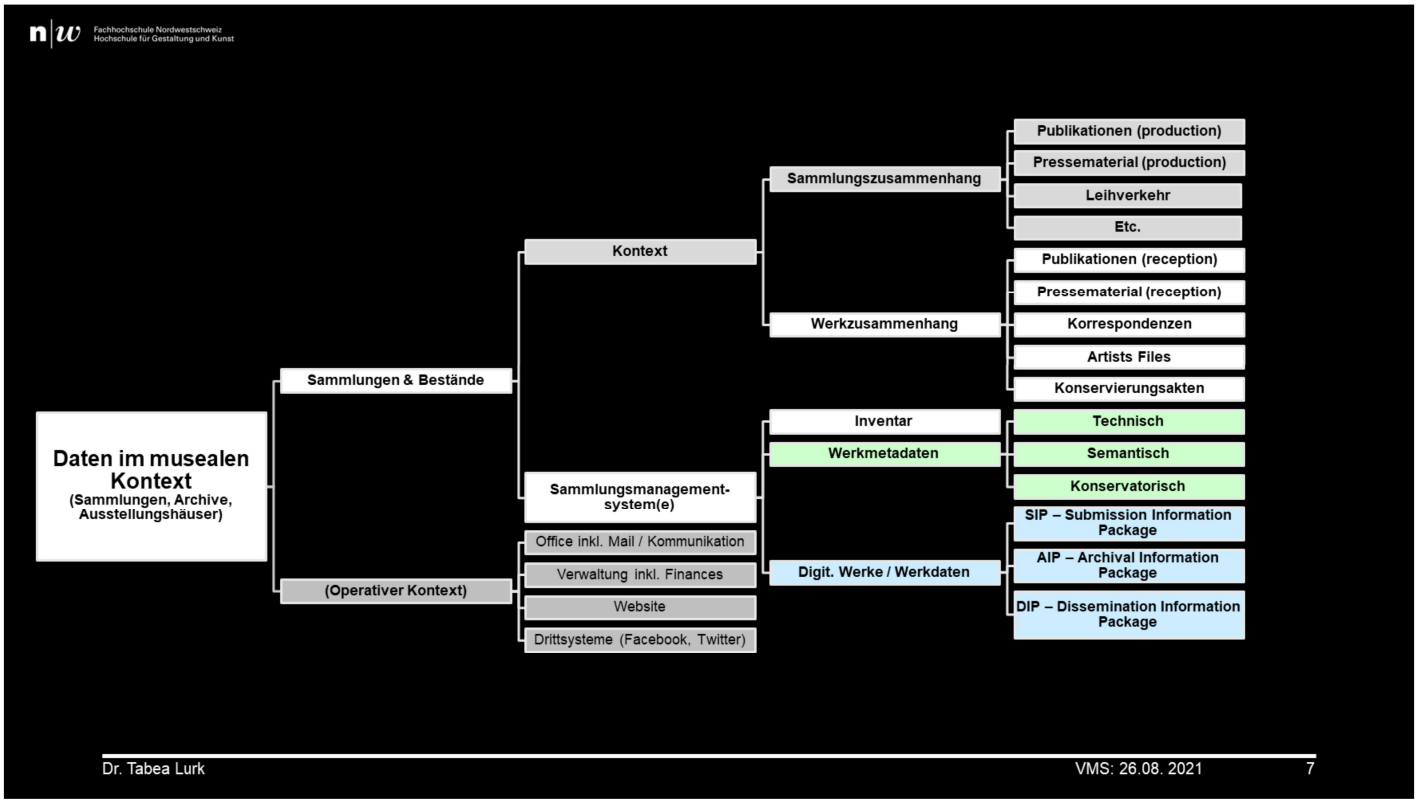
Ebenfalls aus dem Forschungskontext und der Nachnutzung stammen die FAIR-Principles, die ich aber ebenfalls im Sammlungszusammenhang als ausgesprochen fruchtbar empfinde um Prioritäten zu setzen.

Es besagt, dass Inhalte auffindbar, zugänglich (also man weiss, wo sie liegen ≠ frei im Internet verfügbar) und nachnutzbar sind (d.h. die Rechtesituation sollte geklärt sein). Hinzu kommt, dass der automatisierte Datenaustausch zwischen den Sammlungen, also die Interoperabilität, angestrebt wird – ich denke, da sind wir im musealen Kontext auch policy-basiert noch sehr weit von entfernt.

Go-Fair-Org. „FAIR Principles“. GO FAIR, 2016. https://www.go-fair.org/fair-principles/.

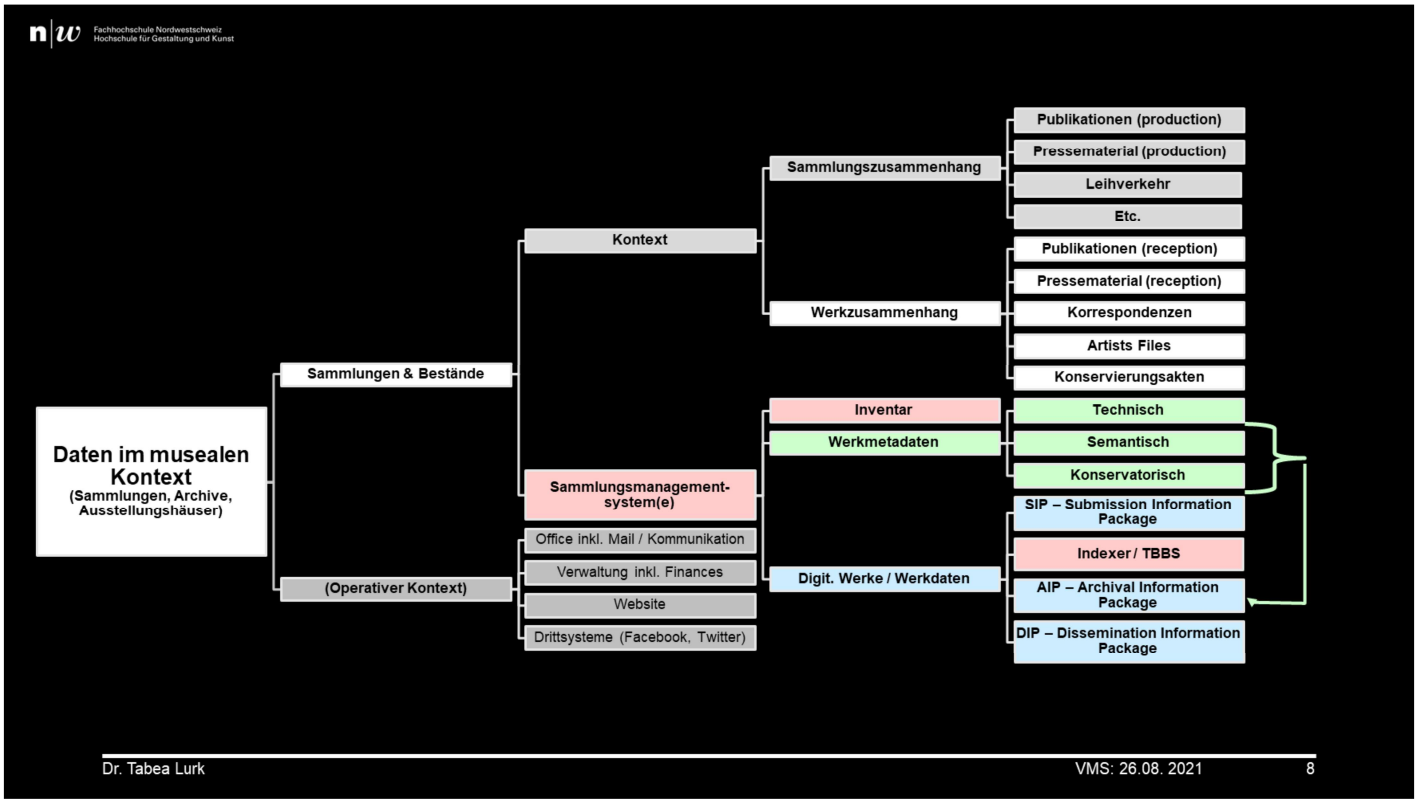


Überträgt man den FDM-Zyklus aus dem universitären Kontext in den Bereich musealer Sammlungen, verschieben sich Fokus und Zuständigkeiten.
Da die Recherche und die Nachnutzung zumindest bis dato bei musealen Sammlungsbeständen eine untergeordnete Rolle spielen bzw. in spezifischen (kuratorischen) Abteilungen erfolgt, wird dieser Aspekt hier in einem Zusammenhang mit dem Publizieren (d.h. auch Ausstellen) und Verteilen betrachtet.
Das frei gewordene Feld des «Discover und Reuse» kann damit für den Ingestprozess und das Erstellen von digitalen Verpackungen (AIPs) aufgewendet werden. Im Unterschied zu vollautomatisierten Archivsystemen, die beim Ingest Daten, die nicht (genug) harmonisiert sind, zurückweisen, besteht diese Option im musealen Kontext kaum. Es braucht also nachhaltige Verpackungen.
Bevor genauer auf dieses «Verpacken» eingegangen wird, zunächst ein Blick auf die Datenstrukturen.

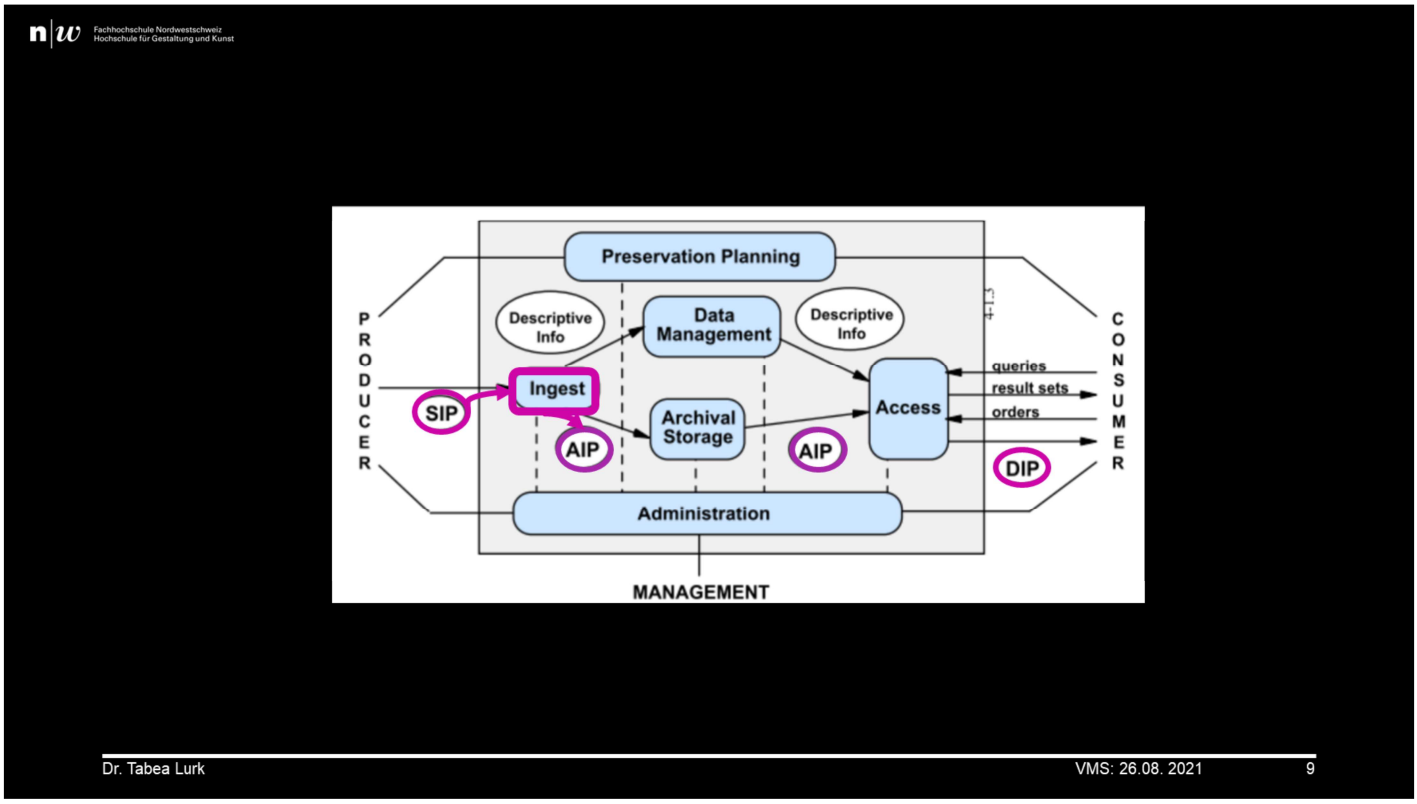


Betrachtet man alle im musealen Kontext anfallenden Daten, finden sich unterschiedliche Wertigkeiten.
Es empfiehlt sich, für jede Sammlung ein Datenprofil zu erstellen, in dem die Datenarten, ihre Wertigkeit und Ablagestruktur verzeichnet ist. Einige Anhaltspunkte zeigt das Schema auf. Es differenziert zwischen Daten, die beim operativen Geschäft anfallen, und Sammlungsrelevanten Daten, wobei mein besonderer Fokus im Folgenden dem gilt, was ich hier als «Werkdaten» bezeichne. Also jenen Daten, die auch als digitale Sammlungsgüter gelten (hier blau dargestellt). Zu ihnen kommen in aller Regel noch beschreibende Daten (Mintgrün) hinzu, die Werkmetadaten enthalten.

Digital(isiert)e Sammlungen - Herausforderungen (bei) der Archivierung



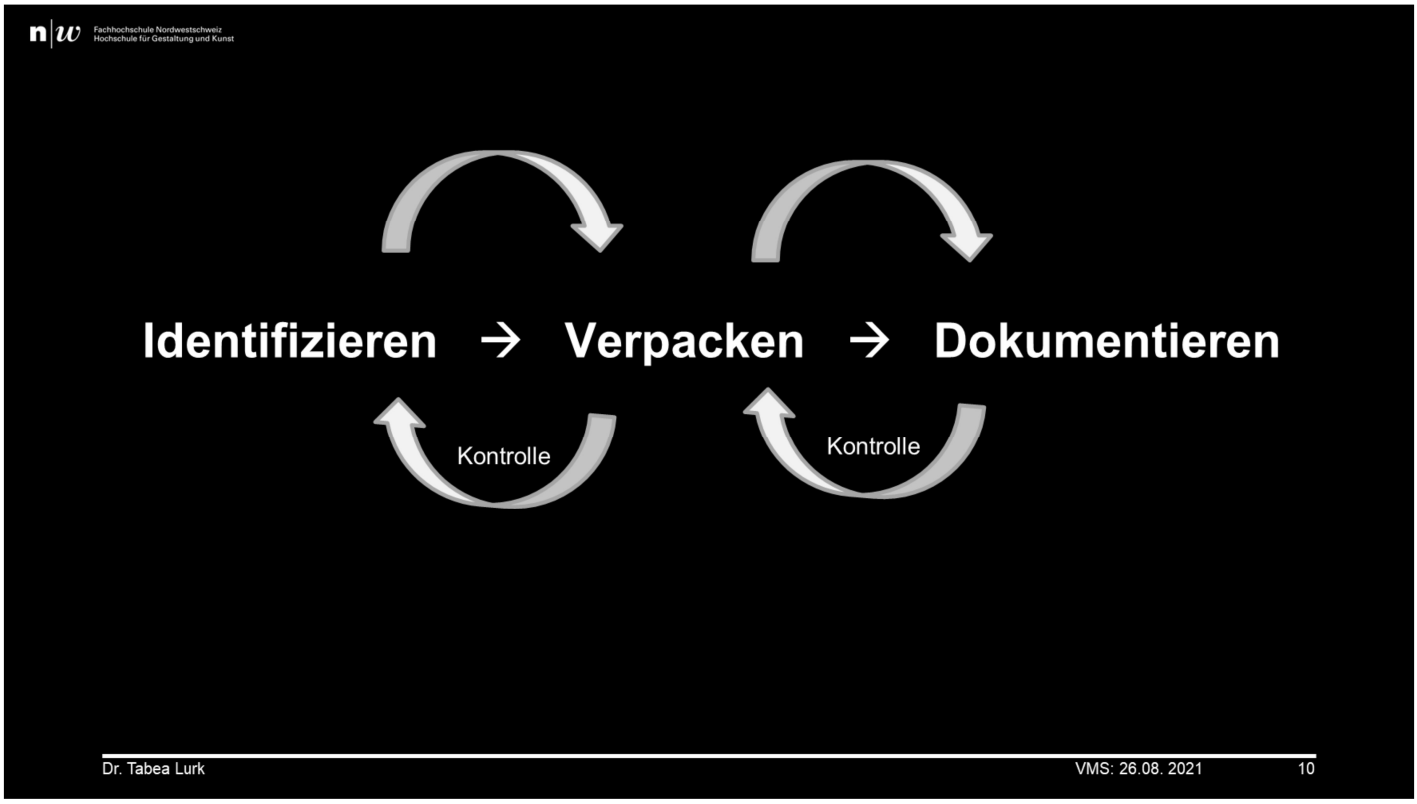
Auch wenn das Inventar einer digitalen Sammlung von einer einfachen Tabelle oder Liste bis hin zu komplexeren Sammlungsmanagementsystemen reichen kann, bietet es sich an, eine systematische Ablage spätestens dann anzustreben, wenn im Bestand auch digitale Kulturgüter verwaltet werden müssen. Denn so können Zusatzinformationen, die im Rahmen des Ingest erzeugt werden, dort abgelegt und gespeichert werden.
Die entsprechenden Schritte sind hier in Rot hervorgehoben.
Bevor also noch einmal genauer auf den Ingestprozess, also die Vereinnahmung der Daten eingegangen wird, hier ein kurzer Exkurs zum OAIS-Modell, einem der wichtigsten Referenzmodelle der digitalen Archivierung



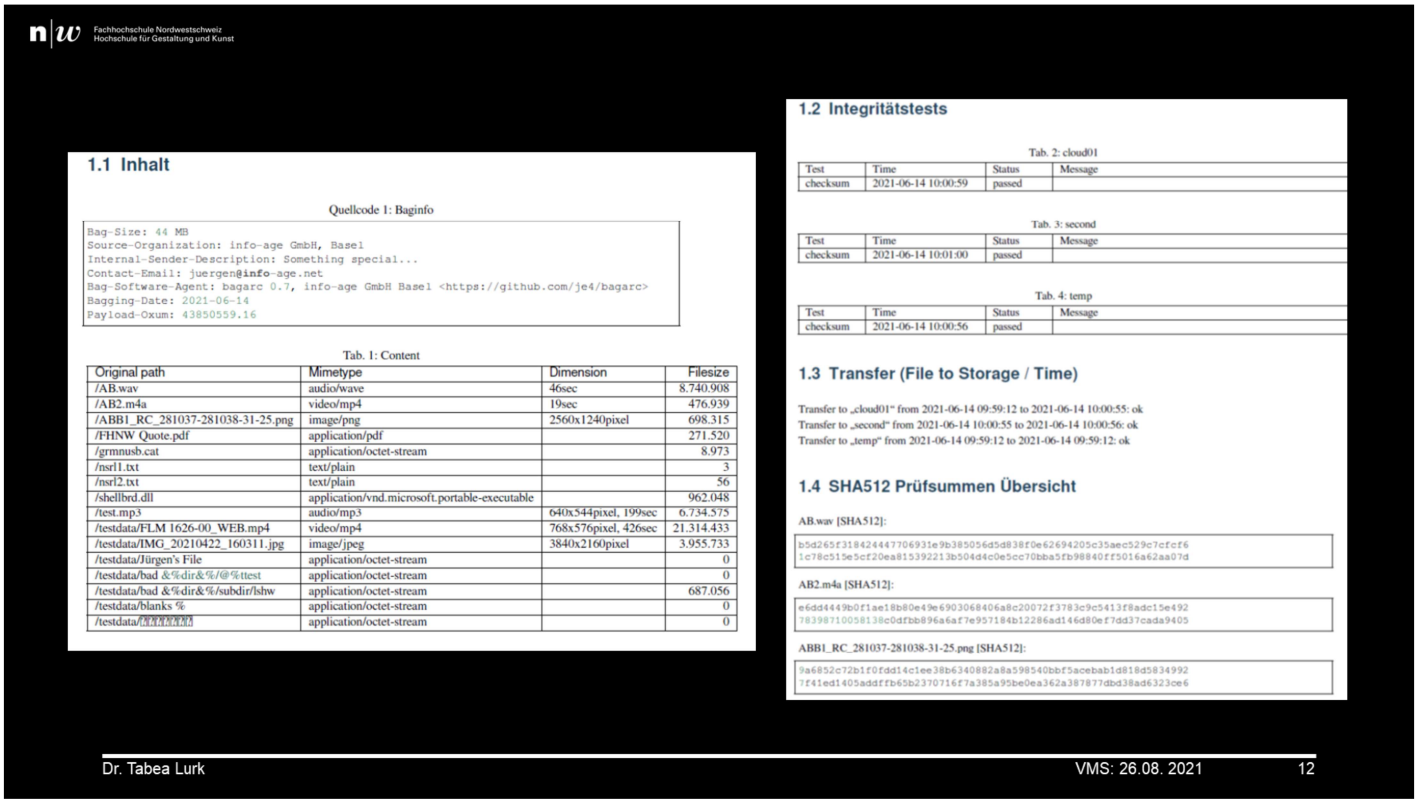
Im Unterschied zu analogen Objekten, «sehen» oder «erfahren» wir - vereinfacht gesagt – bei digitalen Kulturgütern nicht das, was in die Sammlung einmal eingebracht wurde, das Submission Information Package (SIP), dessen Inhalts vielleicht am ehesten als Rohdaten gedacht werden können, sondern wir sehen «erhaltene» Objekte in den jeweiligen Zugangsformaten, die Dissemination Information Packages (DIP). Dafür Sorge zu tragen, dass es sich dabei um eine logistische Differenz und keine spürbare Differenz handelt, ist Aufgabe des systeminternen Datenmanagements (hier Administration genannt) und je nach Kontext, also vor allem bei (schnell) alternden Dateiformaten, des Preservation Managements, denn hier gehört z.B. das Filemonitoring zum Standardprozess, der jedoch kollaborativ von der Community betrieben wird.

Mit geht es im folgenden um den Schritt der Erzeugung der AIPs, die im Rahmen des Ingestprozess initial «gepackt» werden – den rechten Part des AIP-Handlings lassen wir hier aus.
Da wir zu den Packroutinen demnächst eh Info-Workshops anbieten, wenn das Bundesamt für wirtschaftliche Landesversorgung seinen Minimalstandard zur Datensicherheim im kulturellen Sektor veröffentlich, kann man dabei auf einen späteren Zeitpunkt verweisen, für diejenigen, die das interessiert. Dabei werden dann auch die Inhalte der Folgefolien genauer exploriert. Hier geht es erst einmal nur um das konzeptionelle Framework.

Neuroth, Heike, Achim Oßwald, Stefan Strathmann, Matthias Jehn, Regine Scheffel, und nestor – Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland, Hrsg. „Kapitel 4: Das Referenzmodell OAIS – Open Archival Information System“. In *nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierunghgdes Projektes.*, 2009. http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_368.pdf.


n|w Fachhochschule Nordwestschweiz
Hochschule für Gestaltung und Kunst
Identifizieren → Verpacken → Dokumentieren
Kontrolle
Kontrolle
Dr. Tabea Lurk
VMS: 26.08. 2021
10

Exemplarisch für die Dokumentation des technischen Prozesses beim Erstellen der digitalen Verpackung in BagITs sei hier das Dokumentationsprotokoll eines Testdatensatzes gezeigt, das von dem BagIT-Modul von Jürgen Enges TBBS erzeugt wurde.
Im Inhaltsverzeichnis links sieht man die Tools der Identifikationskaskade.



n|w Fachhochschule Nordwestschweiz
Hochschule für Gestaltung und Kunst

1.1 Inhalt

Quellcode 1: Baginfo

```
Bag-Size: 44 MB
Source-Organization: info-age GmbH, Basel
Internal-Sender-Description: Something special...
Contact-Email: juergen@info-age.net
Bag-Software-Agent: bagarc 0.7, info-age GmbH Basel <https://github.com/je4/bagarc>
Bagging-Date: 2021-06-14
Payload-Oxum: 43850559.16
```

Tab. 1: Content

Original path	Mimetype	Dimension	Filesize
/AB.wav	audio/wave	46sec	8.740.908
/AB2.m4a	video/mp4	19sec	476.939
/ABB1_RC_281037-281038-31-25.png	image/png	2560x1240pixel	698.315
/FHNW Quote.pdf	application/pdf		271.520
/grmnusb.cat	application/octet-stream		8.973
/nsrl1.txt	text/plain		3
/nsrl2.txt	text/plain		56
/shellbrd.dll	application/vnd.microsoft.portable-executable		962.048
/test.mp3	audio/mp3	640x544pixel, 199sec	6.734.575
/testdata/FLM 1626-00_WEB.mp4	video/mp4	768x576pixel, 426sec	21.314.433
/testdata/IMG_20210422_160311.jpg	image/jpeg	3840x2160pixel	3.955.733
/testdata/Jürgen's File	application/octet-stream		0
/testdata/bad &%dir&%/@%ttest	application/octet-stream		0
/testdata/bad &%dir&%/subdir/lshw	application/octet-stream		687.056
/testdata/blanks %	application/octet-stream		0
/testdata/[redacted]	application/octet-stream		0

1.2 Integritätstests

Tab. 2: cloud01

Test	Time	Status	Message
checksum	2021-06-14 10:00:59	passed	

Tab. 3: second

Test	Time	Status	Message
checksum	2021-06-14 10:01:00	passed	

Tab. 4: temp

Test	Time	Status	Message
checksum	2021-06-14 10:00:56	passed	

1.3 Transfer (File to Storage / Time)

Transfer to „cloud01" from 2021-06-14 09:59:12 to 2021-06-14 10:00:55: ok
Transfer to „second" from 2021-06-14 10:00:55 to 2021-06-14 10:00:56: ok
Transfer to „temp" from 2021-06-14 09:59:12 to 2021-06-14 09:59:12: ok

1.4 SHA512 Prüfsummen Übersicht

AB.wav [SHA512]:

```
b5d265f31842444770693 1e9b385056d5d838f0e62694205c35aec529c7cfcf6
1c78c515e5cf20ea815392213b504d4c0e5cc70bba5fb98840ff5016a62aa07d
```

AB2.m4a [SHA512]:

```
e6dd4449b0f1ae18b80e49e6903068406a8c20072f3783c9c5413f8adc15e492
783987100581 38c0dfbb896a6af7e957184b12286ad146d80ef7dd37cada9405
```

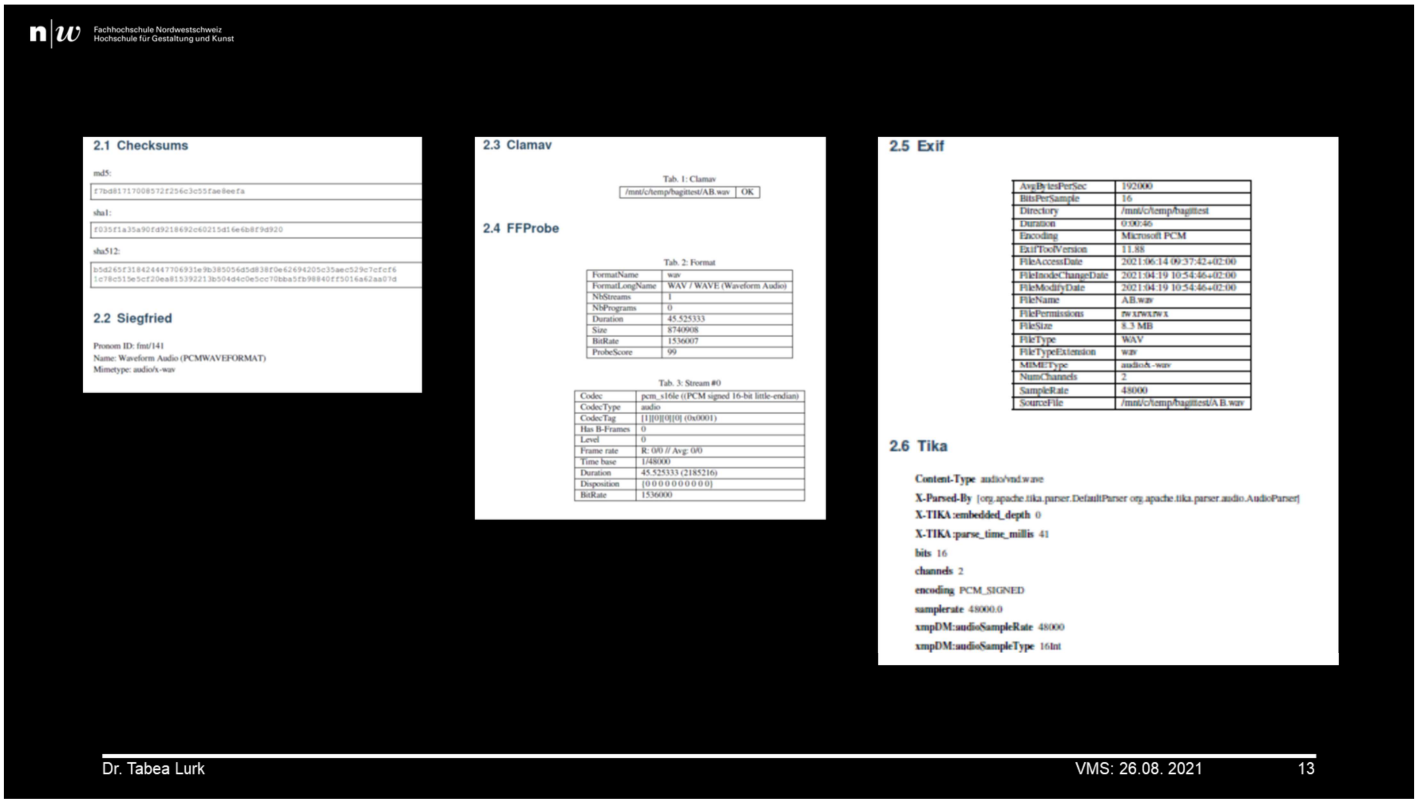
ABB1_RC_281037-281038-31-25.png [SHA512]:

```
9a6852c72b1f0fdd14c1ee38b6340882a8a598540bbf5acebab1d818d5834992
7f41ed1405addffb65b2370716f7a385a95be0ea362a387877dbd38ad6323ce6
```

Dr. Tabea Lurk VMS: 26.08. 2021 12

Es beginnt mit einer

- Inhaltsangabe, was im BagIT genau enthalten ist,
- gefolgt von Integritätschecks, einem Protokoll zu den erfolgten Transfers, wohin die Daten gespeichert wurden, hier einen Cloudspeicher
- Und den drei Checksummen: md5, sha1 und sha512.



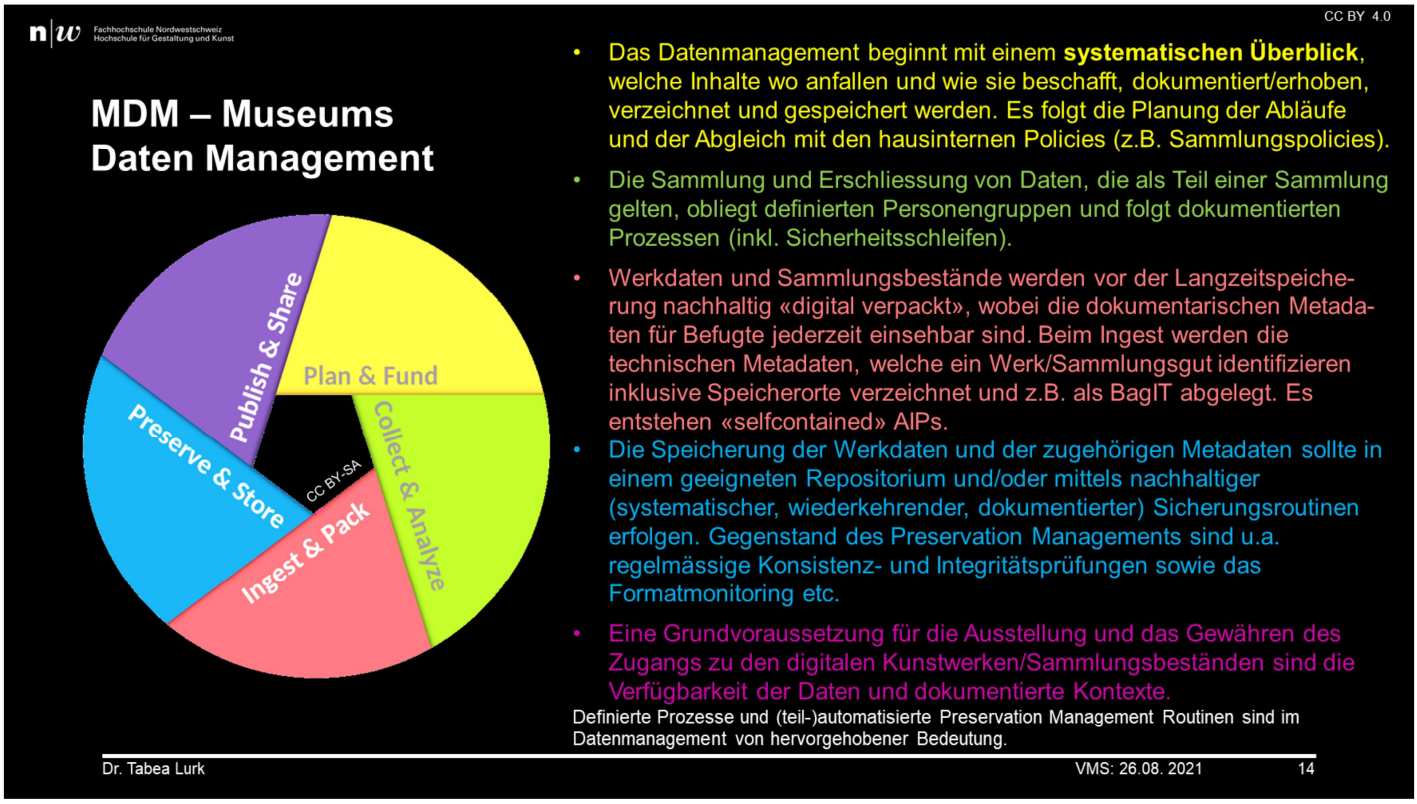
Dann folgt die Identifikationskaskade mit

- Siegfried, Clamav, FFProbe, Exif und Tika
- Sowie ggf. eine normalisierung der Dateinamen bei Sonderzeichen.

Was Sie erkennen sollten, ist hier lediglich, dass die unterschiedlichen Werkzeuge zu variierenden Ergebnissen mit Blick auf die Daten kommen. Das ist der Grund, warum Jürgen Enge bereits vor über 10 Jahren im Indexer für das Deutsche Literaturarchiv in Marbach und den Nachlass von Friedrich A. Kittler, die Identifikationskaskade eingeführt hat.

Spätestens beim Anblick der Checksummen wird deutlich, warum zuvor erwähnt wurde, dass es sich anbietet, für digitale Kulturgüter auch digitale Verwaltungssysteme einzusetzen: eine regelmässige Konsistenz- und Integritätsprüfung kann hier eigentlich nur automatisiert durchgeführt werden.

Enge, Jürgen, und Heinz Werner Kramski. „Exploring Friedrich Kittler's Digital Legacy on Different Levels: Tools to Equip the Future Archivist", 3. Oktober 2016. https://doi.org/10.26041/fhnw-782.



Das Datenmanagement beginnt mit einem systematischen Überblick, welche Inhalte wo anfallen und wie sie beschafft, dokumentiert/erhoben, verzeichnet und gespeichert werden. Es folgt die Planung der Abläufe und der Abgleich mit den hausinternen Policies (z.B. Sammlungspolicies).

Die Sammlung und Erschliessung von Daten, die als Teil einer Sammlung gelten, obliegt definierten Personengruppen und folgt dokumentierten Prozessen (inkl. Sicherheitsschleifen).

Werkdaten und Sammlungsbestände werden vor der Langzeitspeicherung nachhaltig «digital verpackt», wobei die dokumentarischen* Metadaten für Befugte jederzeit einsehbar sind. Beim Ingest werden die technischen Metadaten, welche ein Werk/Sammlungsgut identifizieren inklusive Speicherorte verzeichnet und z.B. als BagIT abgelegt. Es entstehen «selfcontained» AIPs.

Die Speicherung der Werkdaten und der zugehörigen Metadaten sollte in einem geeigneten Repositorium und/oder mittels nachhaltiger (systematischer, wiederkehrender, dokumentierter) Sicherungsroutinen erfolgen. Gegenstand des Preservation Managements sind u.a. regelmässige Konsistenz- und Integritätsprüfungen sowie das Formatmonitoring etc.

Eine Grundvoraussetzung für die Ausstellung und das Gewähren des Zugangs zu den digitalen Kunstwerken/Sammlungsbeständen sind die Verfügbarkeit der Daten und dokumentierte Kontexte.

Definierte Prozesse und (teil-)automatisierte Preservation Management Routinen sind im Datenmanagement von hervorgehobener Bedeutung.

* Unter dokumentarischen Metadaten sind primär die technischen, aber auch semantische und konservatorische Metadaten gemeint vgl. METS/PREMIS.


n|w Fachhochschule Nordwestschweiz
Hochschule für Gestaltung und Kunst
Ohne Datenmanagement
keine (digitale) Langzeitarchivierung !
Dr. Tabea Lurk
VMS: 26.08. 2021
15