

**One-stop publishing and archiving:
Forschungsdaten für Promotionsvorhaben über Repositorien publizieren und archivieren: Eine landesweite Initiative im Rahmen des Projekts bwDataDiss am Beispiel des Karlsruher Instituts für Technologie (KIT)**

Tobias Kurze & Regine Tobias, Deutscher Bibliothekartag 2017, 02.06.2017

KIT-BIBLIOTHEK



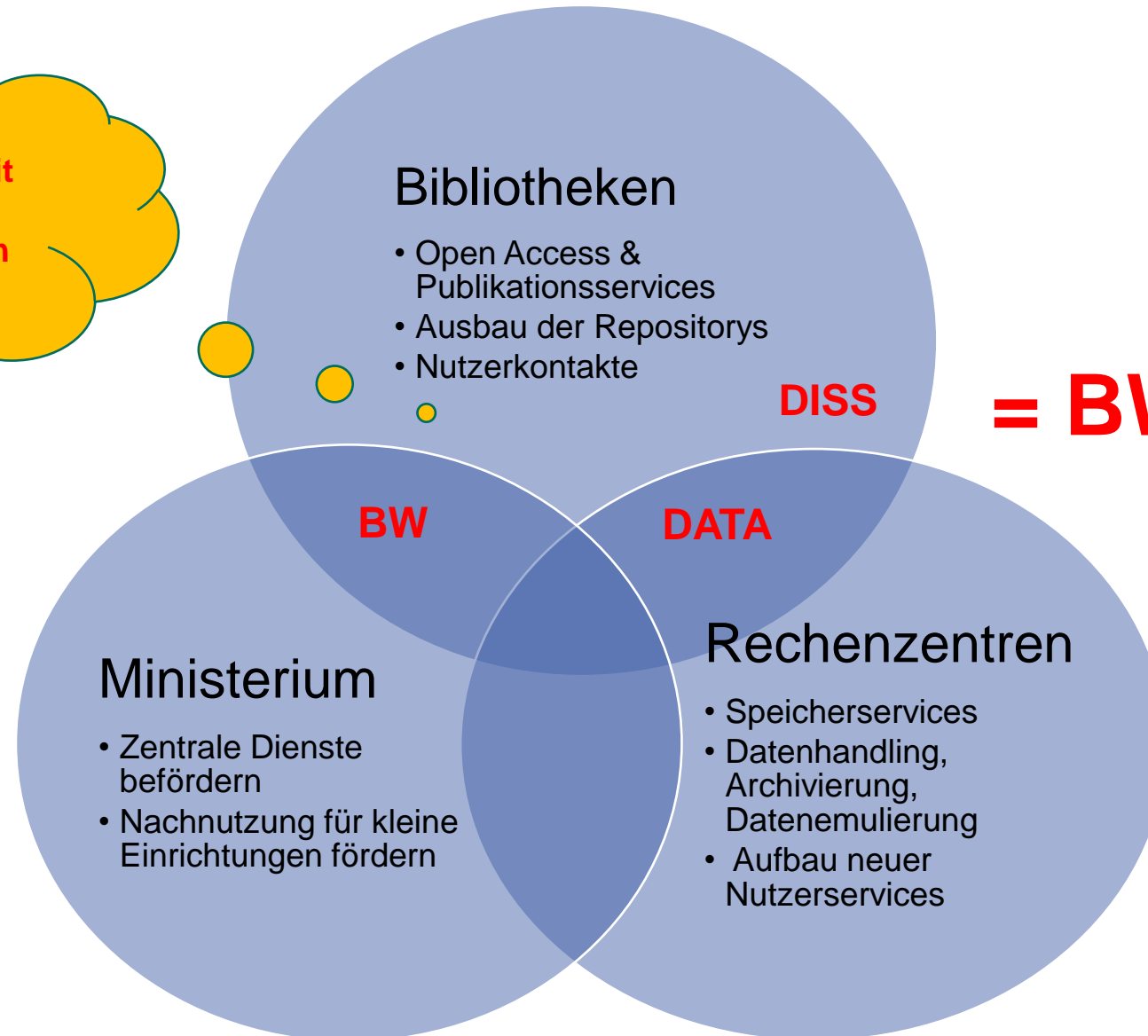
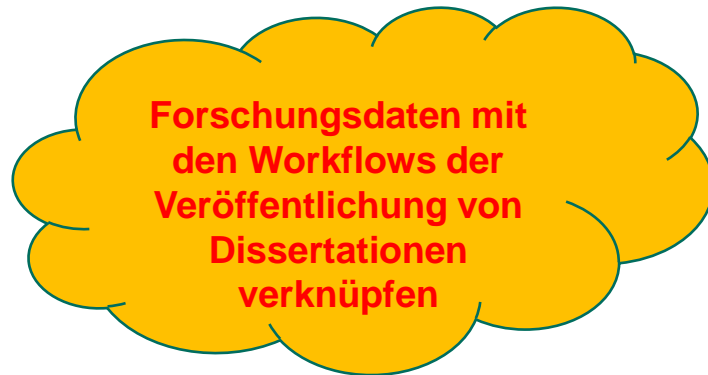
Gliederung

- Projektvorstellung
- Vom Projekt zum Dienst am KIT

- Technische Aspekte
- Ausblick



Projektidee



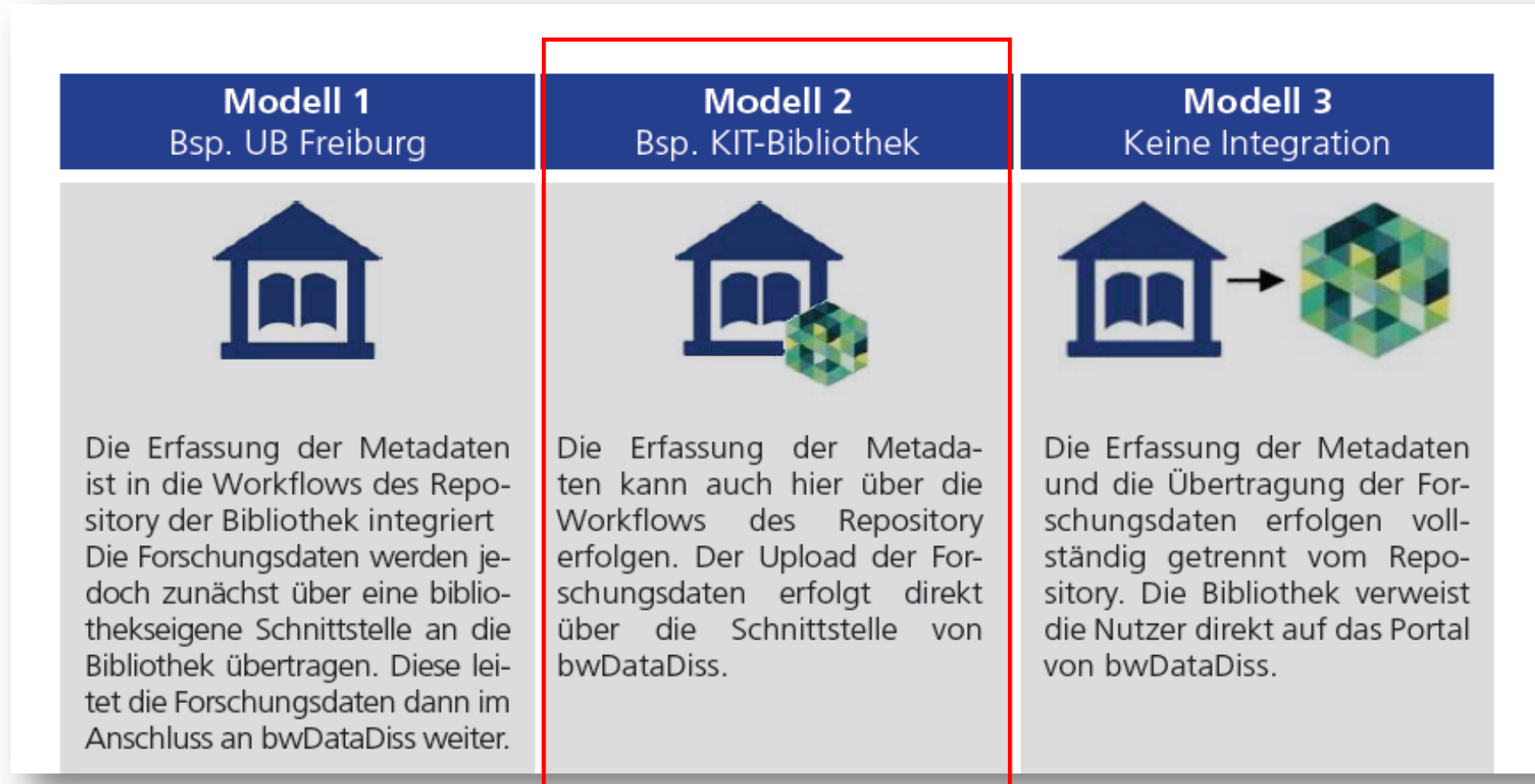
= BW DATA DISS

bwDataDiss – Was und wer ist das?

- Außerdem:
 - Vor der Archivierung führt bwDataDiss eine Charakterisierung der erhaltenen Forschungsdaten durch
 - Die Ergebnisse der Charakterisierung werden den Bibliotheken als Grundlage zur Bewertung der Daten zur Verfügung gestellt
 - ➡ Unterstützt die Qualitätssicherung der Bibliotheken
 - ➡ Basis für neue Services
- Projekt & Partner: Juli 2014 – August 2017
 - Universität Freiburg
 - Bibliothek
 - Rechenzentrum (Datenemulierung)
 - KIT
 - Bibliothek
 - SCC (Datenarchivierung)

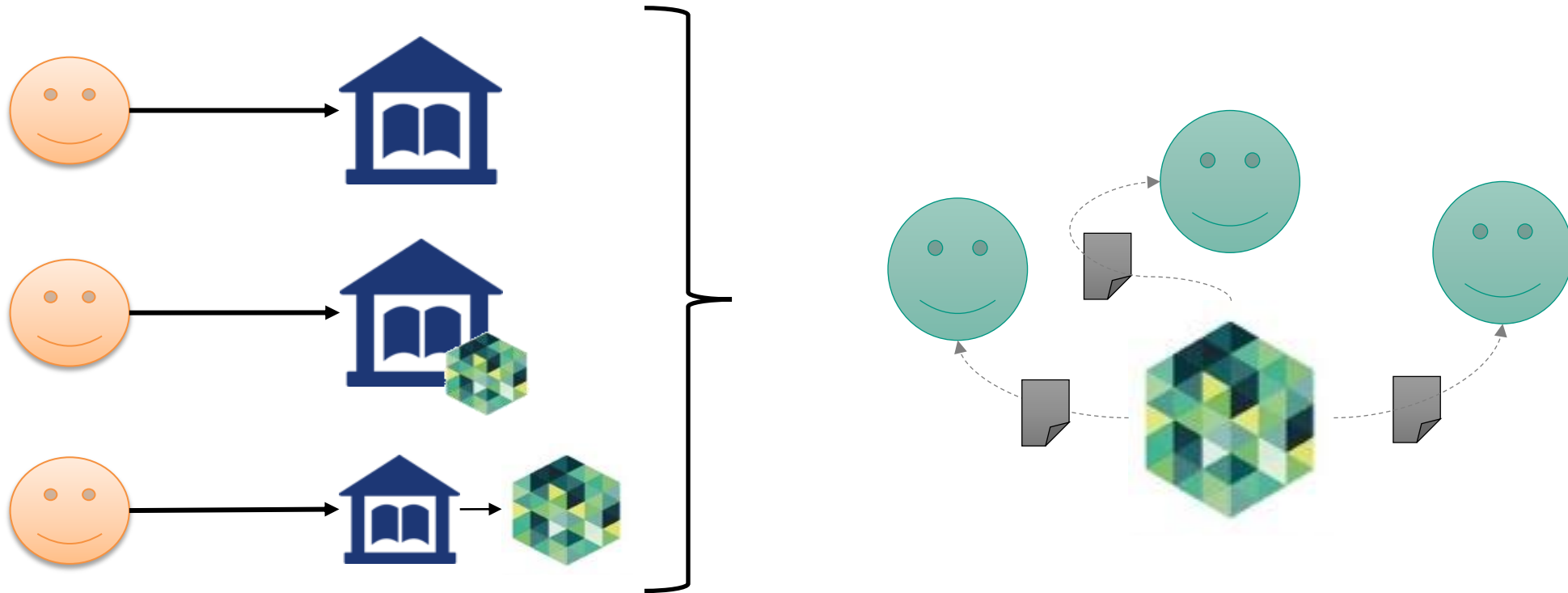
Ergebnisse (1) – Nachnutzung von bwDataDiss

- Erarbeitung von verschiedenen Modellen für die Verknüpfung mit den lokalen Repositorys:





Integrationsmodelle



- bwDataDiss ist sehr flexibel bzgl. Nutzung und Integration
- Alle Varianten und Modelle nutzen das bwDataDiss-Portal für den Zugriff auf die Daten.



Ergebnisse (2) - Aufbau eines neuen Forschungsservices am KIT

Willkommen bei KITopen-Erfassung

 Login Bibliotheksausweis  Login bwDataDiss

| | |
|---|---|
|  | Kontonummer des Bibliotheksausweises / @Alias |
|  | Passwort |
| Anmelden | |

Forschende und Promovierende des KIT können über KITopen ihre Publikationen (bibliographische Angaben, Volltexte, Dissertationen) und weitere Daten (Forschungsdaten, Indikatoren des Großforschungsbereichs) erfassen. Vor der Veröffentlichung erfolgt eine Validierung durch die KIT-Bibliothek. Weitere, allgemeine Informationen zu KITopen


KITopen für Forschende ▼
In 3 Schritten zum Ziel


1. Registrierung (einmalig) ▼
2. Anmeldung ▼
3. Publikationen und weitere Daten erfassen ▼


KITopen für Promovierende ▼
In 3 Schritten zum Ziel


1. Registrierung (einmalig)/ Anmeldung ▼
2. Dissertation publizieren ▼
3. Forschungsdaten publizieren (fakultativ) ▼


Integration des Workflows der Pflichtabgabe der Dissertationen


 Über die „Suche“ im linken Menübereich können Sie prüfen, ob Ihre Publikationen bereits erfasst sind.


Was möchten Sie tun? 



Publikationen in KITopen erfassen und publizieren


Meine Dissertation in KITopen veröffentlichen (Pflichtabgabe)

 Login Bibliotheksausweis

 Login bwDataDiss





Anmelden

Interaktion der Dienste KITopen und bwDataDiss



- Forschungsdatenaupload
- Charakterisierungsergebnisse



Repository KITopen:

- Erfassung der Metadaten der Forschungsdaten
- Kommunikation mit Nutzer




KIT-Katalog: Verweis auf Forschungsdaten in bwDataDiss

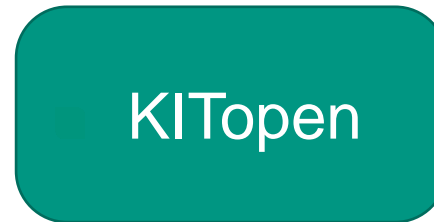
Vom Projekt zum Dienst: Forschungsdaten in KITopen

Dienst für
Doktoranden

Alle Fächer &
Disziplinen

Qualitätssicherung
der Metadaten

Verknüpfung mit
Publikationen



Forschungsdaten
publizieren

Embargooption

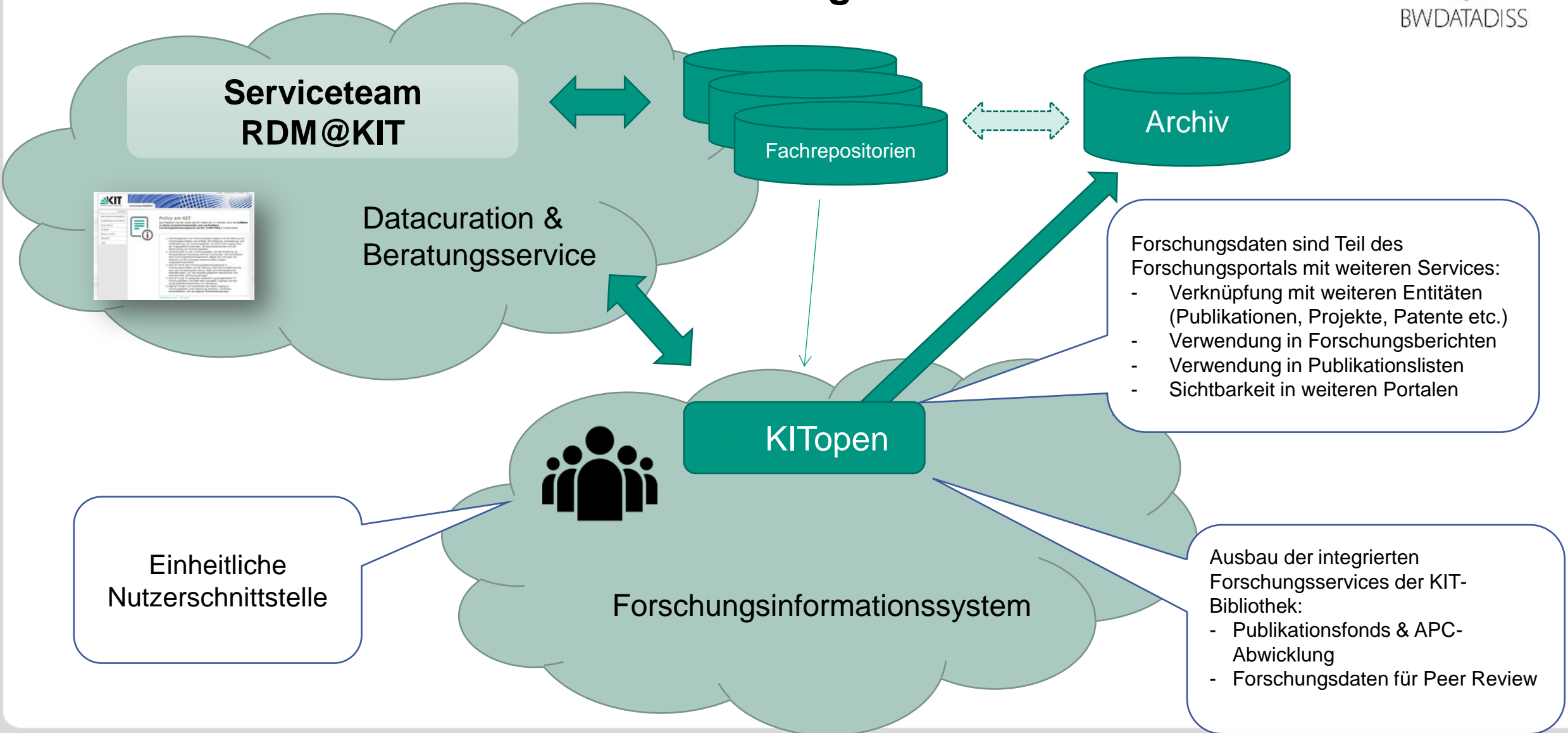
Persistente
Identifikatoren
(DOI)

Data Citation
Index

Langzeitarchivierung



Service ist erst am Entstehen: Forschungsdaten im KIT ...



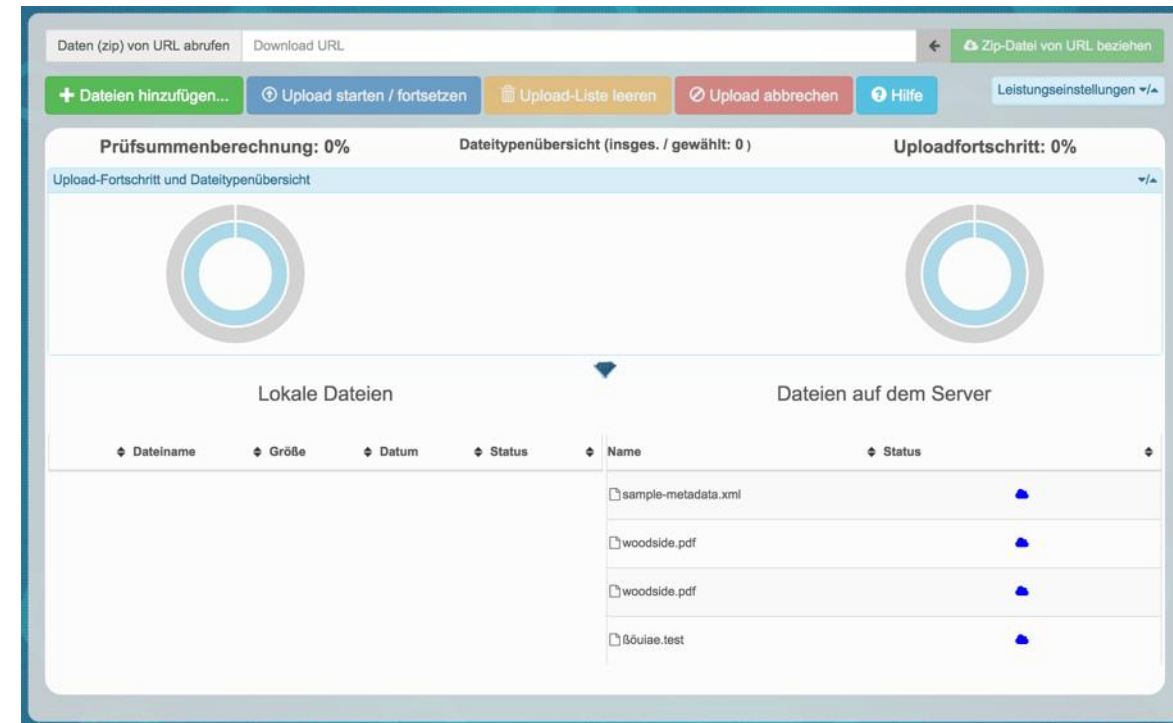


Detailaspekte

Upload, Charakterisierung, Authentifizierung und Metadaten

Upload

- Hochgeladen werden können:
 - Beliebig* viele einzelne Dateien die nicht hierarchisch organisiert sind (in Ordnern, etc.)
 - Eine ZIP-Datei, die die eigentlichen Forschungsdaten enthält und auch Ordner enthalten darf und
 - Entweder als ZIP-Datei archiviert wird
 - ODER auf dem Server entpackt und auch entpackt archiviert wird
- Alternativ zum Upload kann auch die URL einer ZIP-Datei angegeben werden, die dann:
 - vom Server heruntergeladen
 - entpackt
 - und archiviert wird

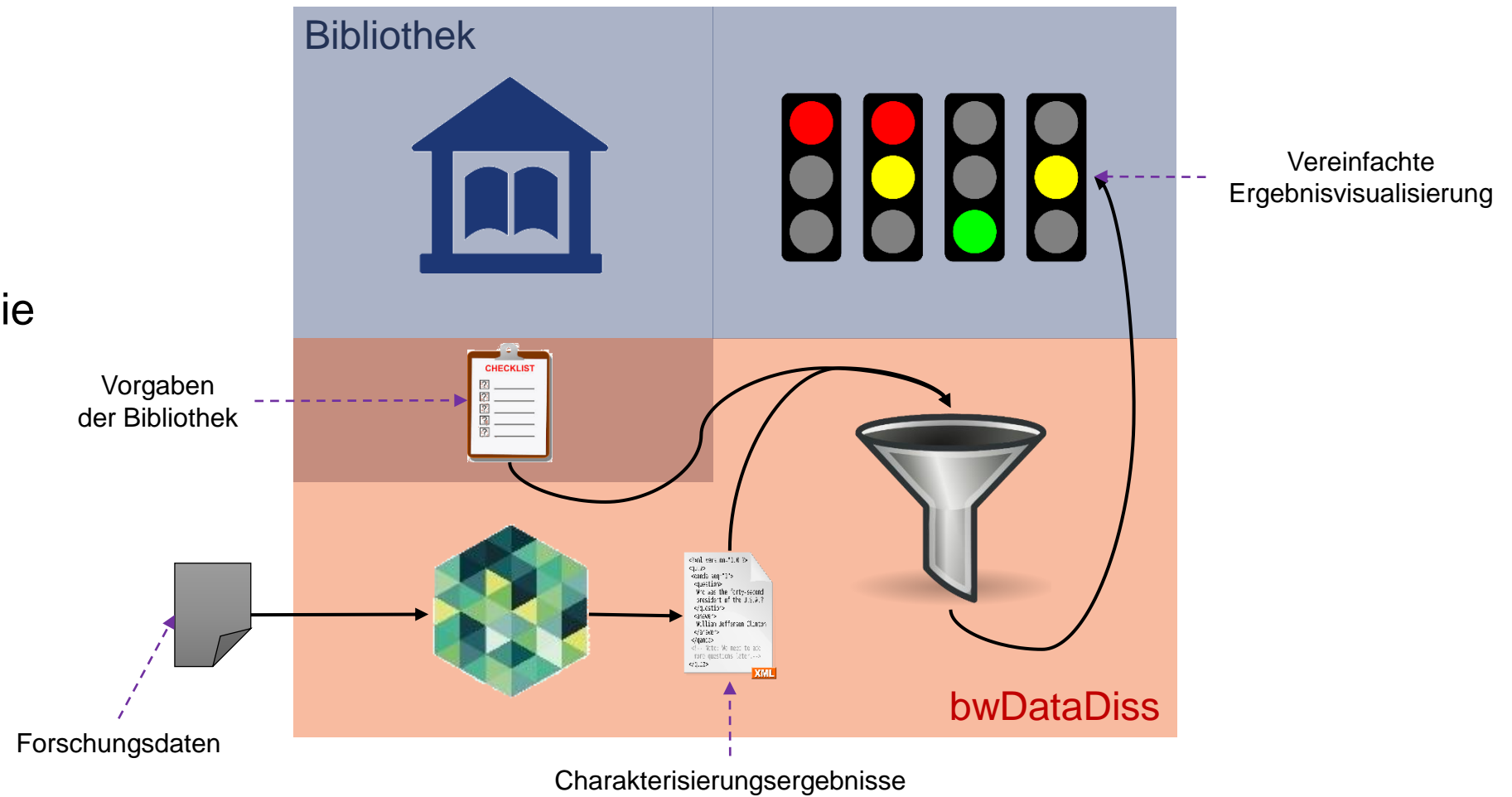


The screenshot shows the BW DATADISS upload interface. At the top, there are buttons for '+ Dateien hinzufügen...', 'Upload starten / fortsetzen', 'Upload-Liste leeren', 'Upload abbrechen', and 'Hilfe'. Below these are progress indicators: 'Prüfsummenberechnung: 0%', 'Dateitypenübersicht (insges. / gewählt: 0)', and 'Uploadfortschritt: 0%'. The main area is divided into 'Lokale Dateien' and 'Dateien auf dem Server'. The 'Dateien auf dem Server' section contains a table with the following data:

| Dateiname | Größe | Datum | Status | Name | Status |
|-----------|-------|-------|--------|---------------------|--------|
| | | | | sample-metadata.xml | 📄 |
| | | | | woodside.pdf | 📄 |
| | | | | woodside.pdf | 📄 |
| | | | | 85ouiae.test | 📄 |

Charakterisierung - Prinzip

- Unterstützung für Bibliotheken bei der Qualitätssicherung
- Grundlage bzw. Entscheidungshilfe für die Erstellung von Erhaltungsstrategien





Charakterisierung – Policies (1)

Charakterisierungspolicy bearbeiten

* Beschreibung

Erste bwDD Policy: - default / unbekannt ist rot - gelb/rot bzw. gelb/grün: die Notwendigkeit von zusätzlicher Software (zBsp weil es ein Binärformat ist, -> Rot wäre dann, wenn man die notwendige Software nicht langfristig bereitstellen kann/will -> gelb, wenn man davon ausgehen muss, dass Software nicht

* Wert bzw. Farbe für nicht konfigurierte PRONOM Typen

rot

Liste der PRONOM Typen

- Acrobat PUA/A - Portable Document Format (version: 3u) [pdf] <fmt/479>
- Acrobat PDF/A - Portable Document Format (Version: 3a) [pdf] <fmt/479>
- Acrobat PDF/A - Portable Document Format (Version: 3b) [pdf] <fmt/480>
- Acrobat PDF/A - Portable Document Format (Version: 3u) [pdf] <fmt/481>
- Acrobat PDF/E - Portable Document Format for Engineering PDF/E-1 [pdf] <fmt/493>
- Acrobat PDF/X - Portable Document Format - Exchange 1.1999 [pdf] <fmt/144>
- Acrobat PDF/X - Portable Document Format - Exchange 1.2001 [pdf] <fmt/145>
- Acrobat PDF/X - Portable Document Format - Exchange 1a.2001 [pdf] <fmt/157>
- Acrobat PDF/X - Portable Document Format - Exchange 1a.2003 [pdf] <fmt/146>
- Acrobat PDF/X - Portable Document Format - Exchange 2.2003 [pdf] <fmt/147>
- Acrobat PDF/X - Portable Document Format - Exchange 3.2002 [pdf] <fmt/138>
- Acrobat PDF/X - Portable Document Format - Exchange 3.2003 [pdf] <fmt/148>
- Acrobat PDF/X - Portable Document Format - Exchange PDF/X
- Acrobat PDF/X - Portable Document Format - Exchange PDF/X
- Acrobat PDF/X - Portable Document Format - Exchange PDF/X
- Acrobat PDF/X - Portable Document Format - Exchange PDF/X
- Active Server Page [asp] <x-fmt/138>
- ActiveX License Package file [pk] <fmt/498>
- Adaptive Multi-Rate Audio [amr] <fmt/356>
- Additive Manufacturing File Format [amf] <fmt/697>
- Adobe ACD [acd] <x-fmt/217>
- Adobe After Effects [aef] <fmt/796>
- Adobe Air (Version: 1.0) [air] <fmt/937>
- Adobe Air (Version: 1.5) [air] <fmt/942>
- Adobe Air (Version: 2.0) [air] <fmt/943>
- Adobe Content Server Message File [acsm] <fmt/871>
- Adobe Flash (Version: 10) [swf] <fmt/507>
- Adobe Flash (Version: 11) [swf] <fmt/757>
- Alpha Flash (Version: 1.1) [swf] <fmt/758>

* Wert bzw. Farbe für unbekannte PRONOM Typen

rot

* Text der Charakterisierungspolicy

fmt/969 GREEN
fmt/96 GREEN
fmt/97 GREEN
fmt/98 GREEN
fmt/99 GREEN
fmt/100 GREEN
fmt/471 GREEN
x-fmt/14 GREEN
x-fmt/15 GREEN
fmt/136 GREEN
fmt/290 GREEN
fmt/291 GREEN
fmt/45 GREEN
fmt/46 GREEN
fmt/47 GREEN

Aktualisieren

← zurück

- Policies gehören immer zu einer Einrichtung und können nur von dieser verändert werden
- Alle Policies stehen aber immer allen Einrichtung zur Nutzung offen und können geklont und dann bearbeitet werden
- Policy Editor bietet eine grafische Oberfläche die eine einfache Bearbeitung und Erstellung von Policies ermöglicht

Charakterisierung – Policies (2)


- "Basis-Policy" zur Bewertung von Dateiformaten
 - 3 (bzw. 4) Klassen (Ampel-Modell)
 - "grün", alle Formate, die mit sehr großer Wahrscheinlichkeit in Zukunft interpretiert werden können
 - ohne besondere Software, wie "plain-text" Formate (ASCII, UTF-8)
 - Formate, die explizit zur Archivierung entwickelt wurden (z. Bsp. PDF/A)
 - "hellgrün" ("grün"), alle anderen binären Formate, die mit sehr großer Wahrscheinlichkeit in Zukunft interpretiert werden können
 - Z.Bsp. Formate mit weiter Verbreitung, (offener) Dokumentation und quelloffener (Referenzsoftware (PNG, GIF etc.)
 - "gelb", alle Formate, die wahrscheinlich noch in Zukunft interpretiert werden können
 - Formate, für die es eine Migrationsstrategie gibt
 - Formate, deren Software ebenfalls archiviert wird und langfristig z.B. mittels Emulation nutzbar bleibt
 - "rot", alle anderen Formate

Charakterisierung – Ergebnisse

| | | | | |
|---|--|------------|-------------------|-------------------------|
| fmt/18 (Acrobat PDF 1.4 - Portable Document Format) | Ø Beschreibung des Problems und der Algorithmen. pdf | 143.1 KiB | 10.01.17 14:48:25 | Details |
| fmt/17 (Acrobat PDF 1.3 - Portable Document Format) | Ø FüßeSchreibtisch1.pdf | 418.89 KiB | 10.01.17 14:48:25 | Details |
| fmt/531 (AutoCAD Drawing) | Ø FüßeSchreibtisch1.bak | 401.58 KiB | 10.01.17 14:48:25 | Details |
| fmt/18 (Acrobat PDF 1.4 - Portable Document Format) | Ø html5boilerplate-pdf.pdf | 228.23 KiB | 10.01.17 14:48:27 | Details |
| x-fmt/263 (ZIP Format) | Ø classicthesis-pkrefined.zip | 143.91 KiB | 10.01.17 14:48:27 | Details |
| fmt/20 (Acrobat PDF 1.6 - Portable Document) | Ø haushaltsplan_bw_2015_16.pdf | 45.99 MiB | 16.01.17 09:24:28 | Details |

Charakterisierungsergebnisse

Der Datensatz wurde charakterisiert. Nachfolgend die Ergebnisse.



```

[
  {
    "value": "GREEN",
    "type": "fmt/279",
    "count": 1,
    "files": [
      ".\\BIS1447-002-flac_24.flac"
    ]
  },
  {
    "value": "GREEN",
    "type": "fmt/20",
    "count": 1,
    "files": [
      ".\\haushaltsplan_bw_2015_16.pdf"
    ]
  },
  {
    "value": "RED",
    "type": "x-fmt/263",
    "count": 2,
    "files": [

```

Ergebnisse der Charakterisierung

- werden für jede Datei farblich hervorgehoben
- und für den Gesamtdatensatz als Ampel dargestellt
 - Mittelung der Einzelwerte der Dateien
- können auch über die bwDataDiss Schnittstelle abgerufen werden

Benutzeranmeldung / Authentifizierung

- bwIDM (Föderiertes Identitätsmanagement der bw-Hochschulen)
 - Bei bwDataDiss können sich Nutzer per bwIDM anmelden
 - Bei Integration z.B. des Uploaders auf der Bibliothekswebseite, kümmert sich bwDataDiss um die bwIDM Anmeldung
- Einladungskomponente von bwDataDiss
 - Nutzer können von der Bibliothek zu bwDataDiss eingeladen werden
 - Eingeladene Nutzer können sich dann – wie oben beschrieben – authentifizieren
 - Dies ist insbesondere dann nützlich, wenn keine ‘normale’ Anmeldung mehr mit einem Hochschulnutzerkonto möglich ist
- bwDataDiss Kontennutzung
 - Die Bibliothek kann Nutzernamen + Passwort Kombinationen ihrer zugeordneten Nutzer überprüfen und dies als Authentifizierung nutzen

Benutzereinladung



BWDATADISS

Neue Policy > Policy bearbeiten > Charakterisierungspolicies > **Bib: Nutzereinladung** > Benachrichtigungen

Bibliothek: Benutzer zu bwDataDiss einladen

Mit diesem Formular können Sie Benutzer zu bwDataDiss einladen. Die Funktion ist für Benutzer vorgesehen, die sich anderfalls (über bwIDM) nicht anmelden können, weil z.B. der benötigte Universitätsaccount deaktiviert wurde. Sollte der Benutzer die Einladung annehmen, wird er ihrer Bibliothek zugeordnet.

* Vorname
Vorname des

* Nachname
Nachname des

* E-Mail Adresse des Empfängers
E-Mail Adresse

Gültigkeit der Einladung
Bis wann soll die Einladung gültig sein?

* Betreff
Der Betreff der E-Mail Nachricht an den einzuladenden Benutzer

* Nachricht an den Einzeladenden

```
<!DOCTYPE HTML>
<html>
<head>
  <meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
</head>
<body>
<h3>Accounteinladung für bwDataDiss</h3>

Hallo <span id="firstName">{{ firstName }}</span> <span id="surName">{{ surName }}</span><br>
<br>
Sie wurden von Admin User (SYSTEM) zu bwDataDiss eingeladen.
<p>
Um ihren Account zu aktivieren, klicken Sie auf nachfolgenden Link <br>
<a href="http://localhost:8000/invite/;inviteToken;/accept">Einladung annehmen</a>
</p>
</body>
```


Text der E-Mail Nachricht an den einzuladenden Benutzer (kann auch HTML enthalten)

* Absenderadresse
Ihre Absenderadresse

[Zur Liste der Benutzereinladungen](#)

Hinweise zur Gültigkeit der Einladung
Sie können eine Gültigkeitsbegrenzung für die Einladung hinterlegen. Ist diese überschritten wird die Einladung ungültig und kann nicht mehr akzeptiert werden.

- Können Nutzer bwIDM nicht mehr nutzen, können neue Accounts durch Einladung zu bwDataDiss durch die Bibliothek angelegt werden
 - Das Abschicken einer Benutzereinladung ist direkt über das bwDataDiss Portal möglich (als Bibliotheksmitarbeiter)

Metadaten

- Klassifizierung:
 - Neben DFG (nach Fachsystematik) und DDC ist es nun auch möglich, beliebige andere Klassifizierungen anzugeben
 - Für DDC und DFG bietet das bwDD Portal eine Auswahlhilfe
- MARC21 Format
 - In Zusammenarbeit mit UBFR wird bwDataDiss Metadatenschema in MARC21 abgebildet
 - bwDataDiss unterstützt MARC21 per OAI-PMH (Protokoll für Metadaten-Harvesting)
- Metadaten können auch über die bwDataDiss Schnittstelle eingespielt werden
- DublinCore kann auch (per OAI-PMH) angeboten werden
 - Entsprechende Erweiterung nötig
 - In Abstimmung mit zukünftigem Erstnutzer kann Schema festgelegt werden

Sonstiges

- Datensätze können einem Embargo unterworfen werden
 - Währenddessen kann einzelnen Nutzern trotzdem Zugriff auf den Datensatz gewährt werden
 - Ein Embargo hat immer ein Ablaufdatum, nach welchem der Datensatz frei zugänglich ist
- Datensätze können auch durch die Bibliothek gesperrt werden und sind dann – bis zur Aufhebung der Sperre – nicht mehr zugänglich
- bwDataDiss vergibt DOIs für Datensätze, wenn dies die jeweilige Bibliothek wünscht

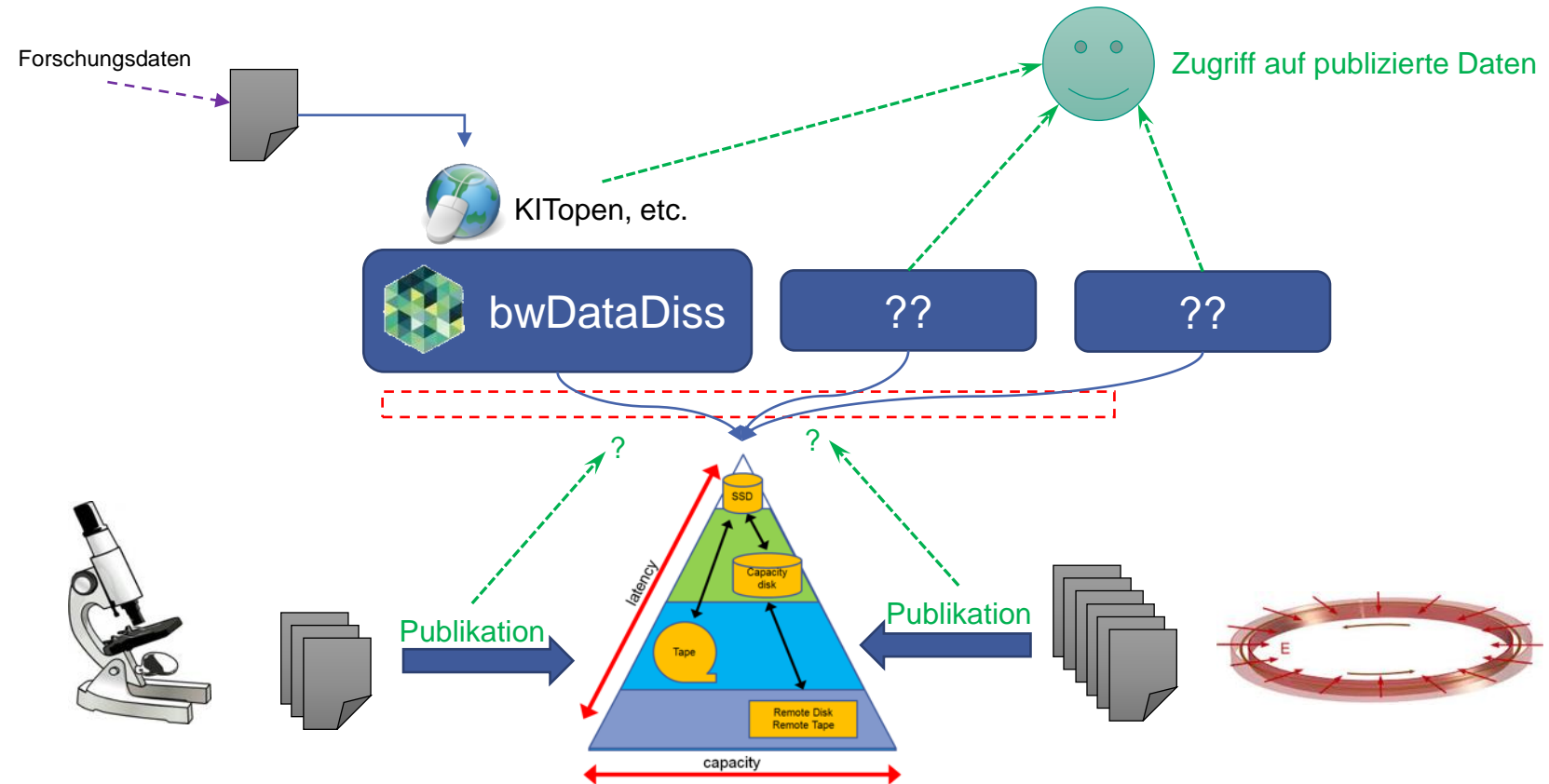


Zukünftige Erweiterungsmöglichkeiten



Erweiterungen bzgl. Archivsystem bzw. Archivschnittstelle

- Archiv basiert auf High Performance Storage System (HPSS), <http://www.hpss-collaboration.org/> und ist ein Hierarchisches Speichersystem (SSD-Caches ↔ Festplatten ↔ Speicherbänder)
- Daten können auch direkt über Geräte bzw. Forschungsprojekte ins Archiv (bzw. „in dessen Nähe“) kommen
 - Diese sollen dann auch (ggf. nur teilweise) publiziert und mit Metadaten versehen werden können



Erweiterungsmöglichkeiten: wie geht's weiter?

- Öffnung gegenüber Spezialumgebungen von Forschungsdaten
 - Statt Upload kann Datenübertragungen zwischen den Systemen erfolgen
- Import von Daten die bereits im Archivsystem bzw. "in dessen Nähe" gespeichert sind
 - Auszeichnung mit Metadaten
 - Vergabe von PIDs



Besten Dank

Bei Interesse und Fragen wenden Sie sich gerne an uns!

regine.tobias@kit.edu & tobias.kurze@kit.edu