



**TEMPORIS -**  
Lifecycle Management wissenschaftlicher Daten

# Management von Forschungsdaten – Der ETHZ-Lösungsansatz

Dr. Susanne Scheid Jakobi

## Gliederung

1. Einführung
2. IT-Architektur an der ETHZ
3. Vorgehen zur Konzipierung einer Workspace
4. Umsetzung
5. Zusammenfassung
6. Diskussion

# 1. Einführung

- ✓ ETHZ-weite Umfrage „Umgang mit Forschungsdaten“ (Sept. 2010 – Sept. 2011)
- ✓ 80 % Rücklaufquote
- ✓ Hauptteilnehmergruppen: Professor, Senior Research Assistent

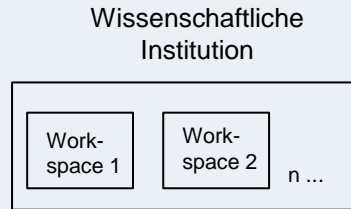
Welche Herausforderungen stellen sich?

- ✓ Hauptanteil der Forschungsdaten wird in Einzelprojekten erarbeitet (PhD-Thesis)
- ✓ Potentielle Vielfalt an Datenformaten
- ✓ wenige bekannte bzw. etablierte Standards für den Umgang mit Forschungsdaten in den Fachcommunities
- ✓ Speicherung von Forschungsdaten auf Stufe *peer-reviewtes Paper*
- ✓ Aufbewahrungszeitraum von 10 Jahren gemäss guter wissenschaftlicher Praxis (Nachprüfbarkeit)
- ✓ Bei Bedarf: Langzeitarchivierung, nicht erste Priorität
- ✓ Bei Bedarf: Bereitstellung von Forschungsdaten für die Sekundärnutzung (Open Data)
- ✓ Bereitschaft der Forschenden, sich mit der Thematik „*Datenmanagement und Archivierung*“ auseinanderzusetzen
- ✓ Kommunikation des Mehrwertes für Forschenden

## 2. IT-Architektur an der ETHZ

### Lokales, dezentrales Datenmanagement

pro kleinster akademischer Einheit **Workspace** mit Zugriffsbeschränkung auf deren Mitglieder

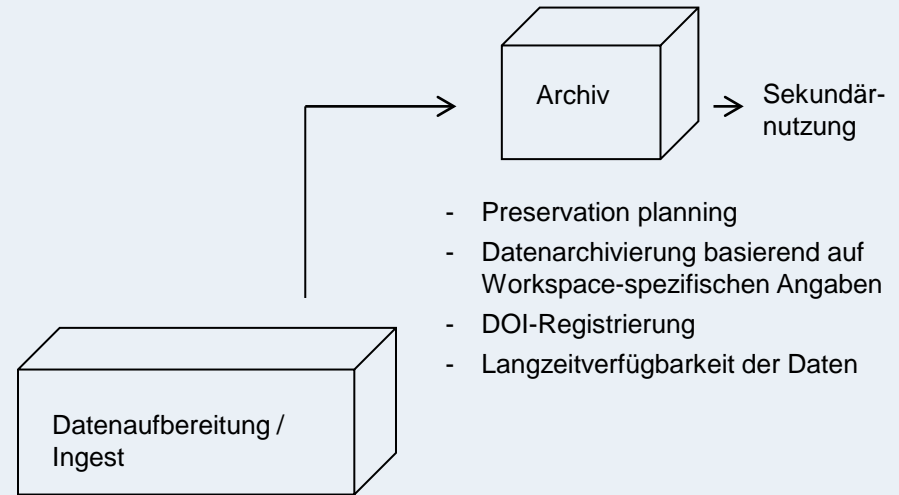


Wechsel der Datenhoheit

### Applikation **Docuteam Packer**

- Start via Weblink / Webstart
- Java-basiert
- Betriebssysteme Windows, Mac OS X, Linux
- kontextbezogene Erfassung von Forschungsdaten
- Erfassung technischer u. administrativer Metadaten für Archivierung (gemäss ISAD(G))

### Zentrale Archivierung



- Preservation planning
- Datenarchivierung basierend auf Workspace-spezifischen Angaben
- DOI-Registrierung
- Langzeitverfügbarkeit der Daten

- Vorbereitung Ingest-Prozess Archiv
- Formatvalidierung
- Viruscheck
- Erstellung METS für Archivierung

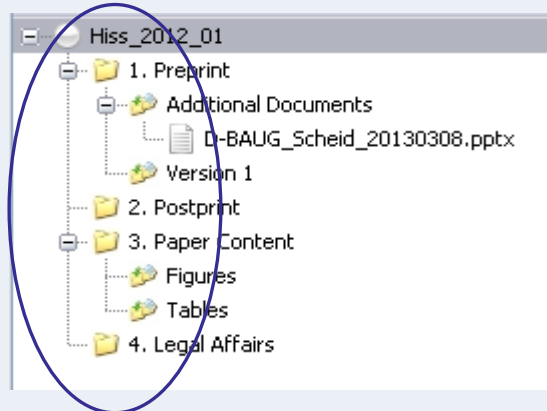
### 3. Vorgehen zur Konzipierung einer Workspace

- Klärung der [Arbeitsweise](#) innerhalb der Professur/Forschungsgruppe im Hinblick auf
  - ✓ Hauptdatenproduzenten (Doktoranden, Postdoc, Wissenschaftliche Mitarbeiter etc.)
  - ✓ Hauptoutput (Paper, Forschungsberichte etc.)
  - ✓ Art der Forschungsdaten (Messdaten, Umfragedaten, Bilder etc.)
  - ✓ Vorhandensein einer Daten-Policy / eines Archivs
  - ✓ Datenzugriff via DOI für Dritte / Sekundärnutzung
  
- Erarbeitung einer [Baumstruktur](#) für die kontextbezogene Datenerfassung
  - ✓ Erstellung eines/mehrerer Templates als Vorlage
  
- Definition der zu erfassenden [Metadaten](#) (unter Berücksichtigung existierender Daten-Policies)
  - ✓ Vorbereitung der Forschungsdaten für die Archivierung und DOI-Registrierung
  
- Erstellung von fix hinterlegten [Picklisten](#) für Metadaten (Vermeidung Schreibfehler integrierte Suchfunktion)
  
- Erarbeitung von [Kurztexten](#), die als Orientierung für den User fix hinterlegt werden können
  
- Zusammenstellung aller möglichen [Datenformate](#) der zu erfassenden Forschungsdaten und Bestimmung der PRONOM-ID (Format Library National Archives)
  - ✓ [Verifizierung](#) der Ergebnisse mittels *DROID* (Digital Record Object Identification) anhand von Originaldatenfiles

## 4. Umsetzung

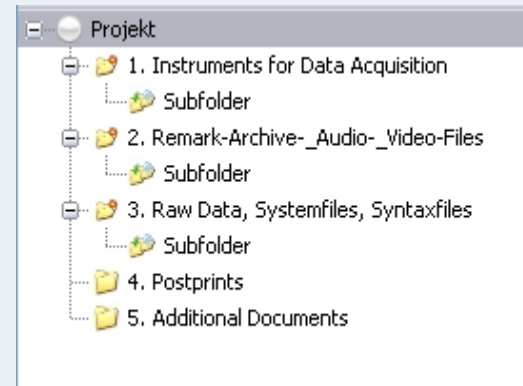
### ✓ Baumstrukturen

#### Paper-bezogene Datenerfassung

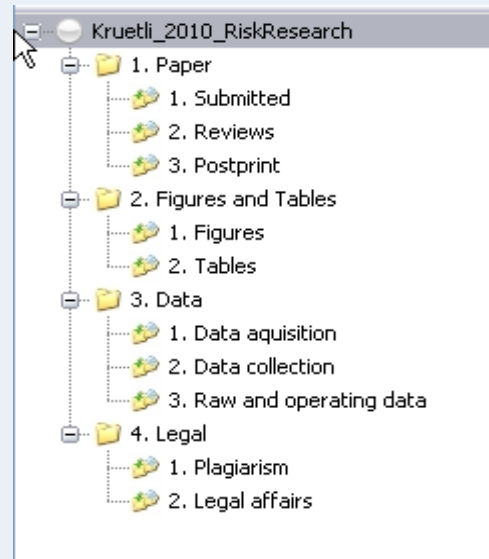


Paket

#### Projekt-bezogene Datenerfassung



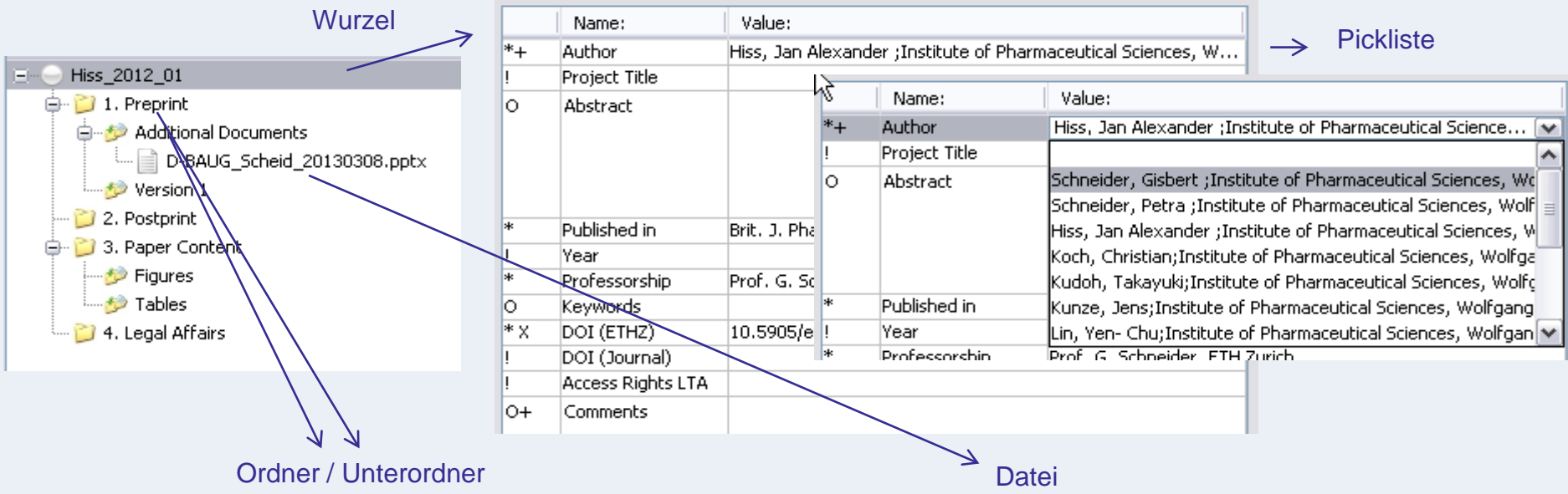
#### Kombination



Professur-/ FG-spezifische Templates  
zur Standardisierung der  
Datenerfassung

# 4. Umsetzung

## ✓ Metadaten und Picklisten

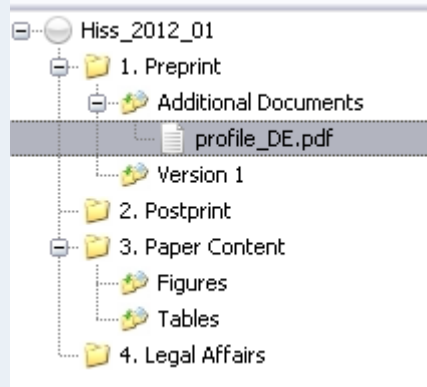


|     | Name:             | Value:  |
|-----|-------------------|---|
| O   | Title             |   |
| I+  | Project Title     |   |
| *+  | Author            | Hiss, Jan Alexander ;Institute of Pharmaceutical Sciences, W... |
| *   | Professorship     | Prof. G. Schneider, ETH Zurich                                  |
| * X | DOI (ETHZ)        | 10.5905/ethz-1004-252   |
| !   | Access Rights LTA |   |
| O+  | Comments          |   |

|     | Name:             | Value:  |
|-----|-------------------|---|
| I+  | Project Title     |   |
| *+  | Author            | Hiss, Jan Alexander ;Institute of Pharmaceutical Sciences, W... |
| * X | Filename          | D-BAUG_Scheid_20130308.pptx                                     |
| *   | Professorship     | Prof. G. Schneider, ETH Zurich                                  |
| *   | Published in      | Brit. J. Pharmacol.   |
| !   | Year              |   |
| * X | DOI (ETHZ)        | 10.5905/ethz-1004-3   |
| !   | Access Rights LTA |   |
| O   | Comments          |   |

## 4. Umsetzung

### ✓ Formatüberprüfung



### Technische Metadaten zur Datei

|              |  |
|--------------|--|
| Level:       | Datei  |
| Name:        | profile_DE.pdf   |
| Path:        | Hiss_2012_01/1. Preprint/Additional Documents/profile_DE.pdf |
| Type:        | file   |
| MIME type:   | application/pdf  |
| Format:      | Acrobat PDF 1.4 - Portable Document Format                   |
| Format ID:   | fmt/18   |
| Size [kB]:   | 3879   |
| Size [%]:    | 100  |
| Children:    | -  |
| Descendants: | -  |
| Events:      | 1  |



Automatische Formatidentifikation  
mittels *DROID*  
(Digital Record Object Identification)

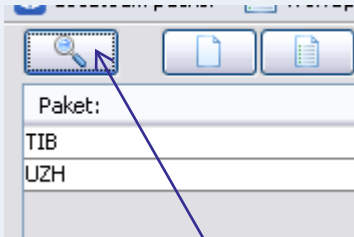
| Summary            |  |
|--------------------|--|
| <b>Name</b>        | Acrobat PDF 1.4 - Portable Document Format   |
| <b>Version</b>     | 1.4  |
| <b>Other names</b> | PDF (1.4)  |
| <b>Identifiers</b> | MIME: application/pdf<br>Apple Uniform Type Identifier: com.adobe.pdf<br>PUID: <b>fmt/18</b> |

Preservation planning

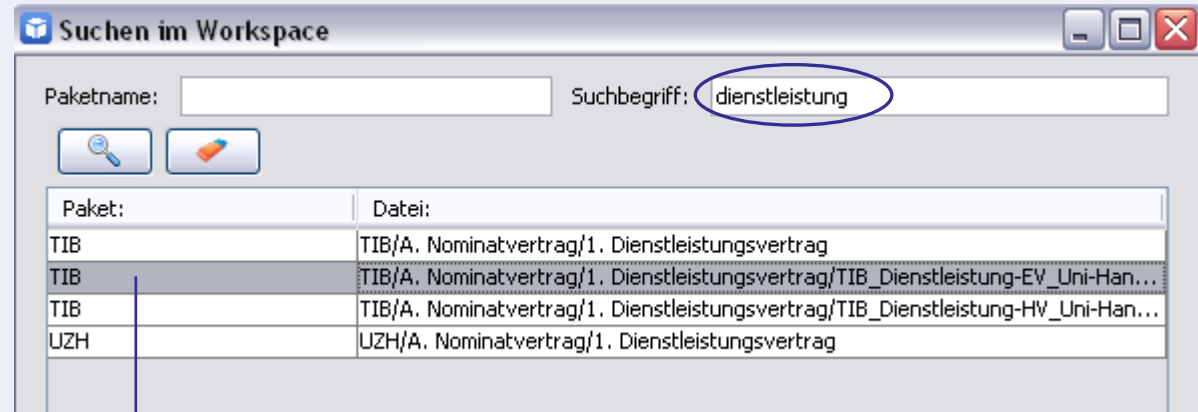


## 4. Umsetzung

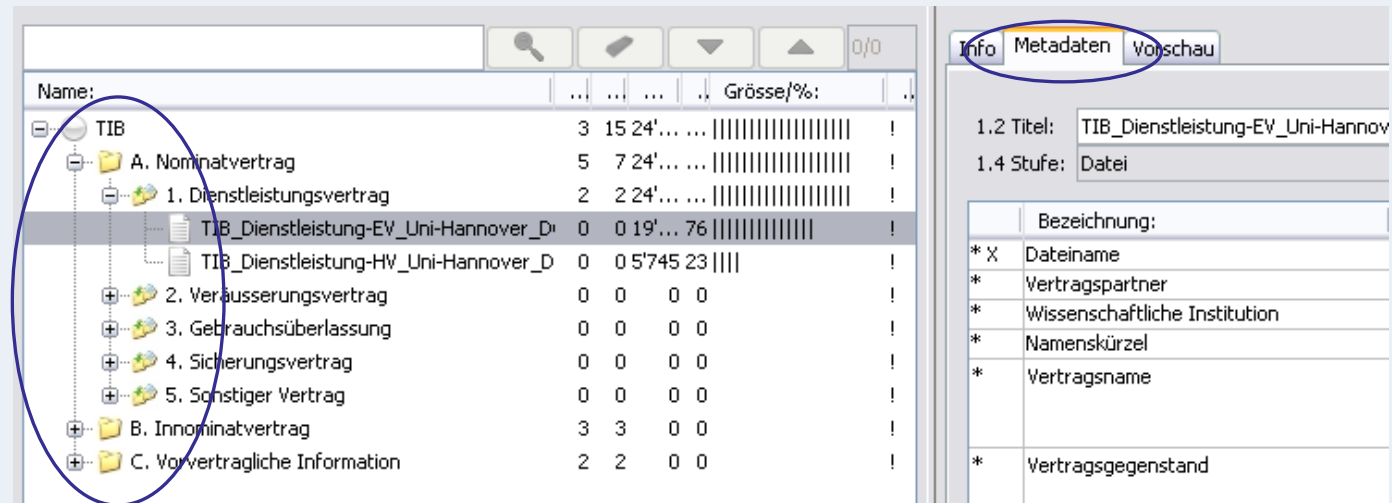
### ✓ Suchfunktionalität



Suchbutton



Anzeige aller Suchtreffer für die Workspace



## 5. Zusammenfassung

Management von Forschungsdaten mit *Docuteam Packer* bedeutet:

- ✓ kontextbezogene Erfassung von Forschungsdaten und deren Management über einen selbstbestimmten Zeitraum mit eigener Datenhoheit bis zur Übergabe ans Archiv
- ✓ kontextbezogene Erfassung von Zusatzdokumenten
- ✓ Paper-bezogene als auch Forschungsprojekt-bezogene Baumstrukturen plus deren Kombination
- ✓ flexible Vergabe von Metadaten zur Beschreibung von Forschungsdaten auf verschiedenen Stufen innerhalb der Baumstruktur (Wurzel, Ordner, Unterordner, Datei)
- ✓ Vererbungsmechanismus für Metadaten von der Wurzel über Ordner/Unterordner bis auf Dateiebene (keine Doppeleingaben, Konsistenz der Angaben für DOI-Registrierung)
- ✓ DOI-Erstellung für langfristige Zitierbarkeit von Forschungsdaten
- ✓ Template-Vorlagen ermöglichen standardisierte Erfassung von Forschungsdaten
- ✓ Aufbau eines internen Repository aufgrund der integrierten Suchfunktionalität

Vielen Dank für Ihre Aufmerksamkeit.