



# Migration Digitaler Objekte

## Ein Rahmenwerk zur automatisierten Unterstützung "Kontrollierter Migrationen"

Treffen

Arbeitskreis Digitale Langzeitarchivierung München  
an der Universität der Bundeswehr München

25.10.2007

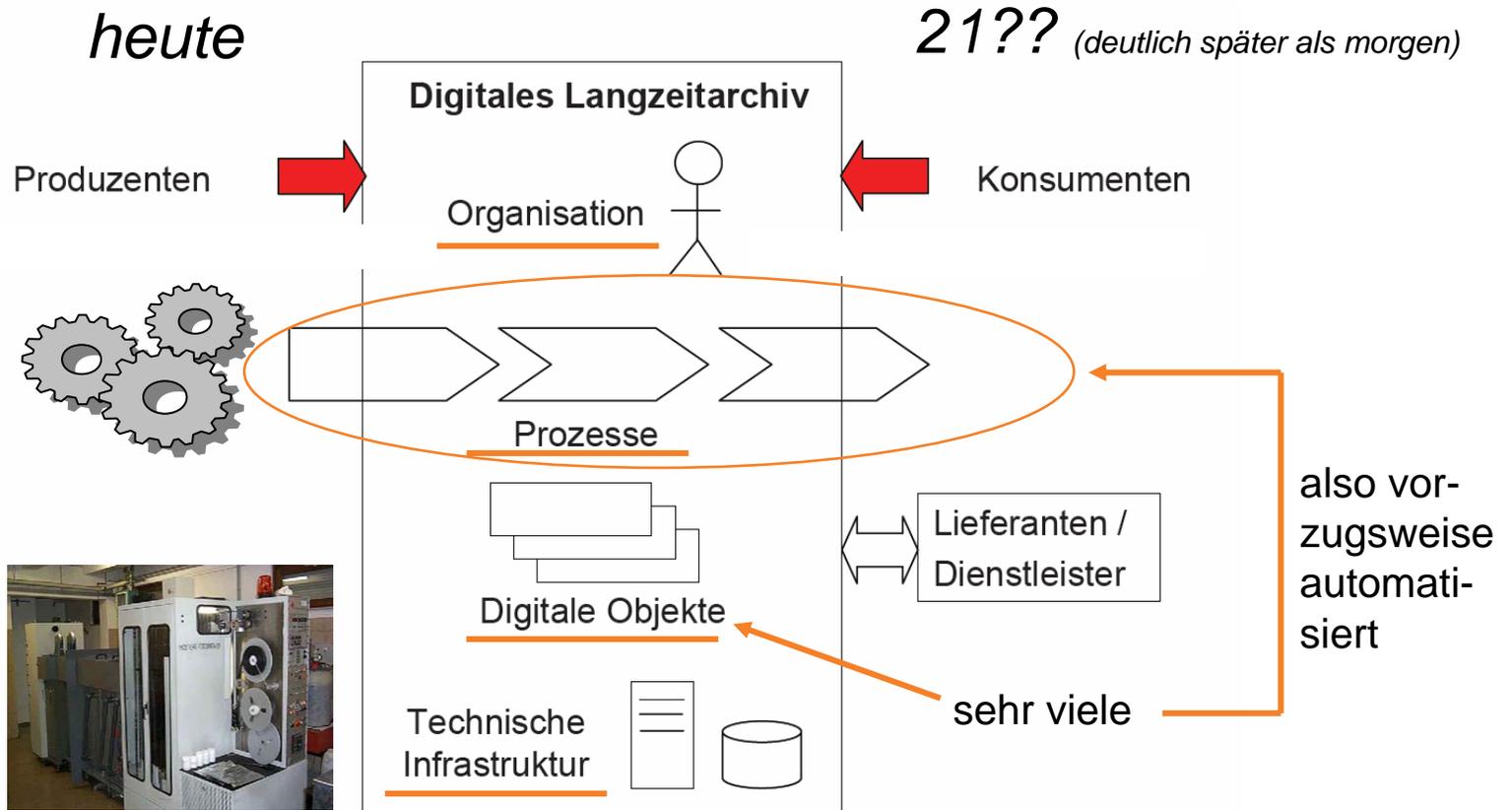
Thomas Triebsees

[Thomas.Triebsees@unibw.de](mailto:Thomas.Triebsees@unibw.de)

Neubiberg, 25. Oktober 2007

- I. Zieldefinition & Einordnung in Gesamtkontext**
- II. Ein Rahmenwerk zur automatisierten Unterstützung "Kontrollierter Migrationen"**
- III. Wesentliche Konzepte**
- IV. Fallstudie – Formattransformation**
- V. Zusammenfassung**

# Migrationsprozesse in digitalen Archiven



relevante Eigenschaften  
Objekt-Versionen

- aufbauend auf einem geeigneten Erhaltungsbegriff einen Beitrag

- zur verbesserten Automatisierung von Migrationsprozessen in digitalen Archiven

Software-Unterstützung

- bei gleichzeitiger Berücksichtigung eines hohen Grades an Vertrauenswürdigkeit

leisten.

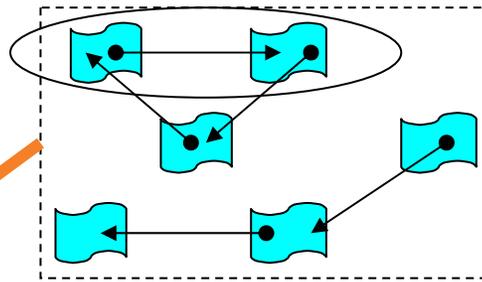
Formalisierung

**Organisation** → Unterstützung bei Formulierung der Anforderungen

**Prozesse** → Integration und Verifikation der Anforderungen über Objekthistorien hinweg

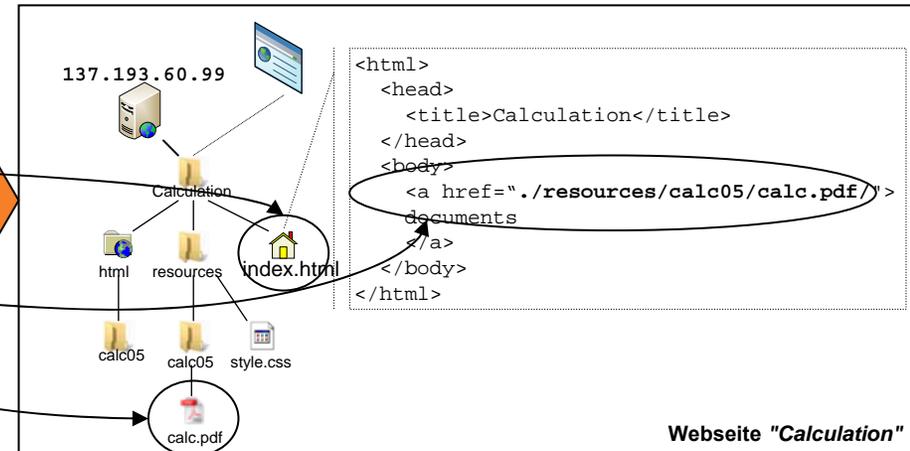
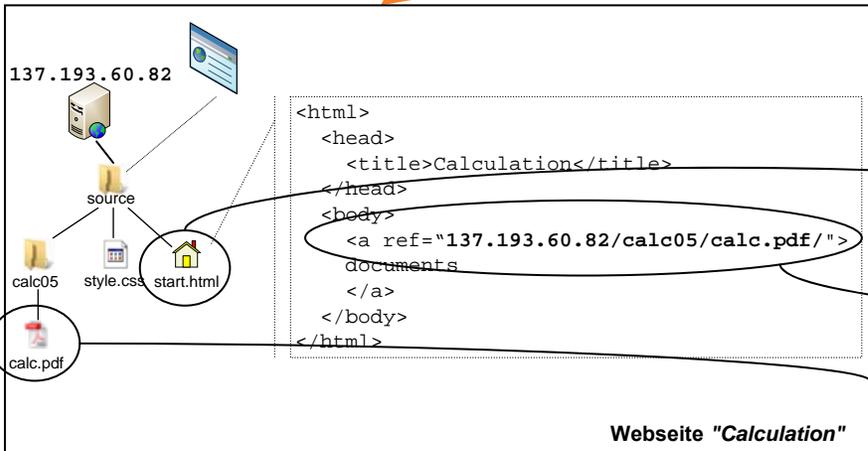
## Web-Dokument

Quellrepräsentation  
→ Quellkontext

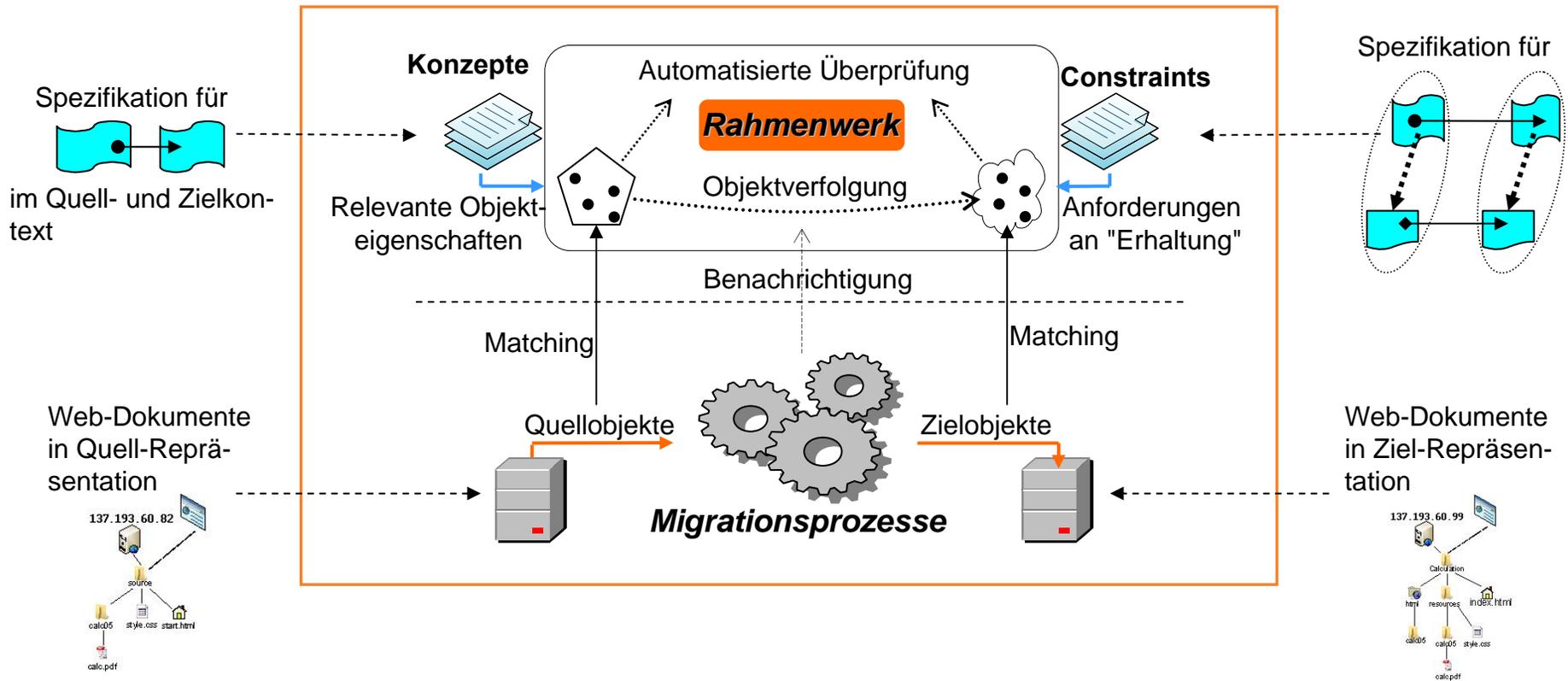


Zielrepräsentation  
→ Zielkontext

Ziel: höhere Portierbarkeit

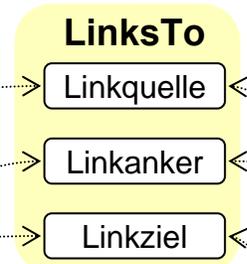
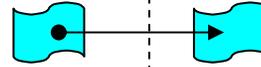


## Rahmenwerk für den Ablauf einer "kontrollierten" Migration



# (1) Spezifikation von Konzepten

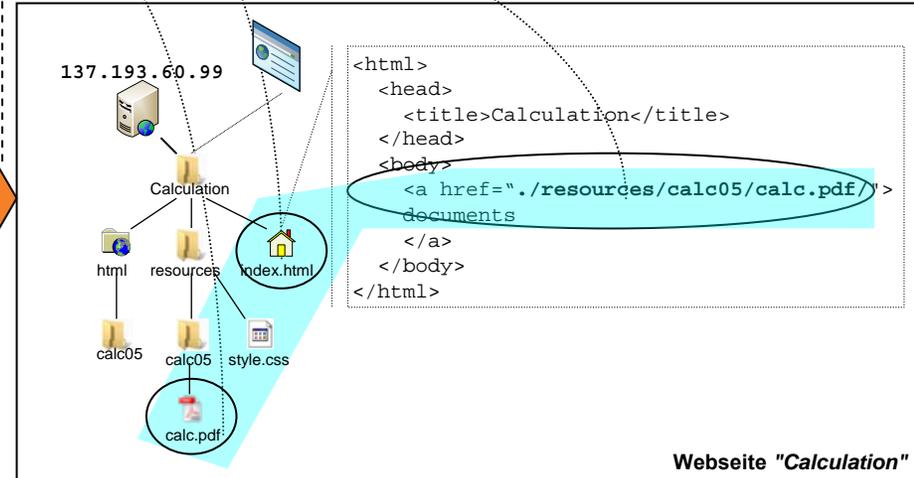
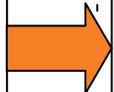
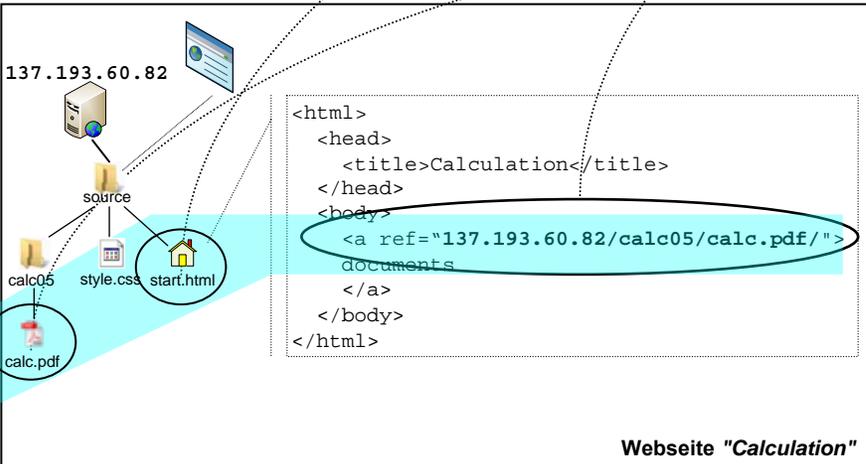
## Konzept + Interface



- Rollennamen definieren
- Objekten Rollennamen zuweisen

### Kontext *LinkAbs*

### Kontext *LinkRel*



## (2) Spezifikation zu erhaltender Eigenschaften - Constraints

### Anforderung:

Bei der Transformation von Web-Dokumenten soll die Linkkonsistenz erhalten werden.

### Semiformale Sprechweise mit Konzepten und Kontexten:

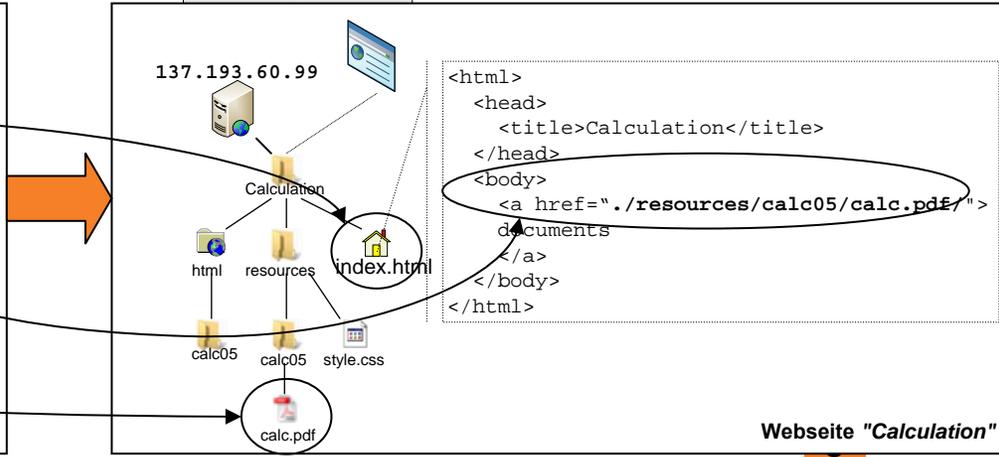
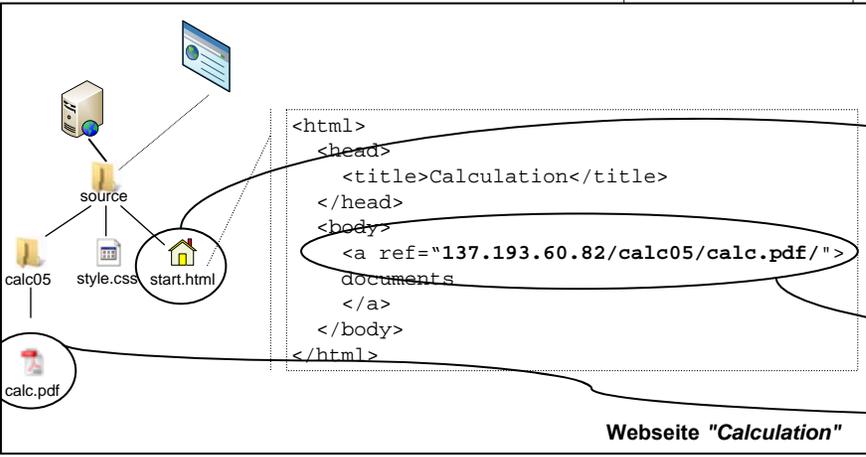
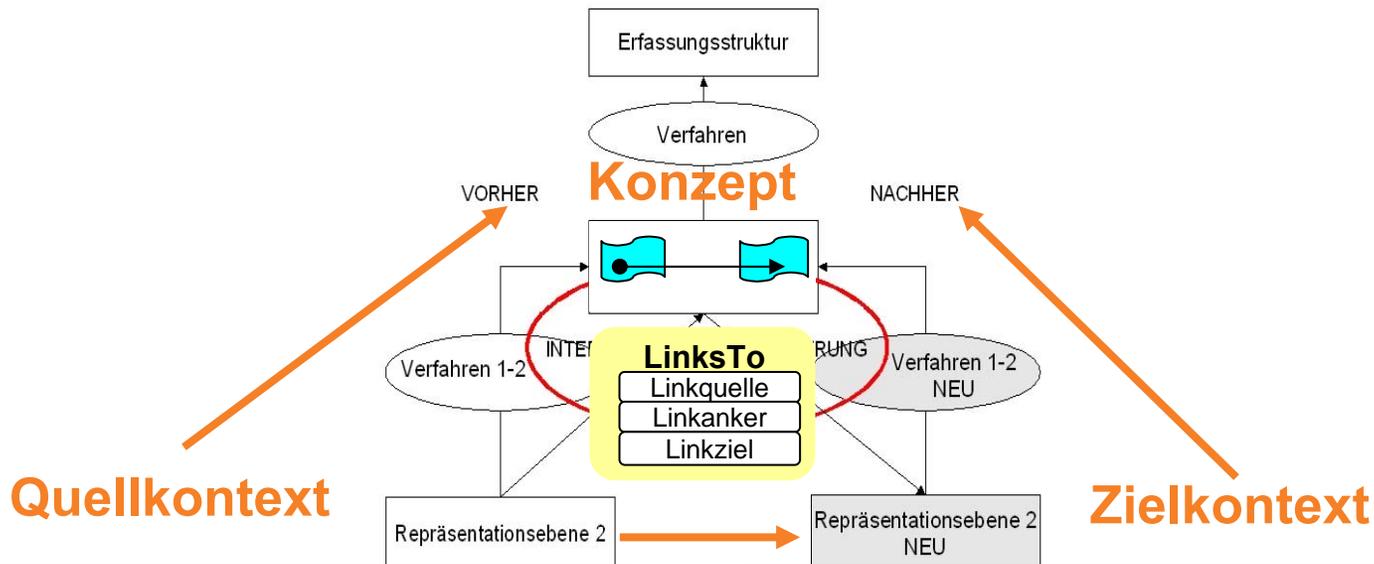
Erhalte bei der Transformation von Linkquelle, Linkanker und Linkziel zu einer neuen Repräsentation das Konzept **LinksTo** für diese Objekte im Zielkontext **LinkRel**, sofern die Objekte im Kontext **LinkAbs** vorlagen.

### Formal ausgedrückt als Constraint:

```
presK( {q → Linkquelle, a → Linkanker, z → Linkziel},  
  LinksTo (q, a, z),  
  {LinkAbs},  
  {LinkRel})
```

Maschinenverarbeitbar &  
abstrakt (ohne Bezug zu  
konkreter Implementierung)

# Erhaltung im Modell der Repräsentationsebenen



Im Rahmen einer Diplomarbeit durchgeführt

## Aufgabe

Transformation XHTML nach ODF und zurück mit CSS-Formatierung

## Ziel

- (1) Evaluierung der konzeptionellen Vorgehensweise
- (2) Bewertung der Anwendbarkeit der theor. Konzepte
- (3) Implementierung der Formattransformation
- (4) Laufzeitmessungen bzgl. Overhead durch formale Verifikation der Erhaltungs-Constraints

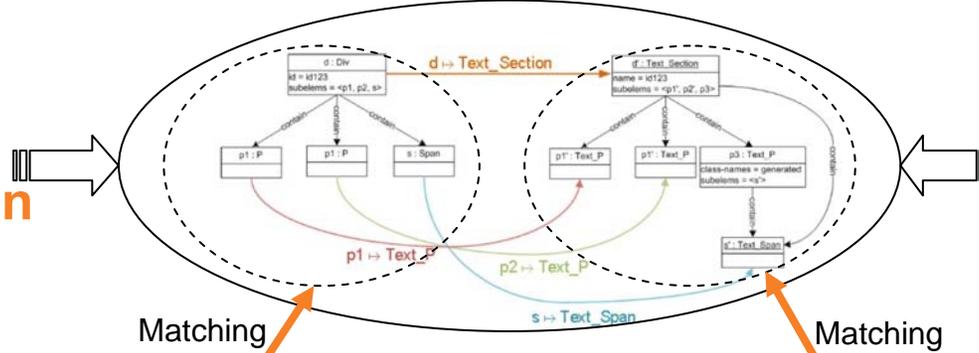
## Besondere Herausforderungen

- (1) Layout / Rendering Zieldokument im Browser (Erhaltung des Layouts)
- (2) Linkkonsistenz bei mehreren Dokumenten
- (3) Zieldokument valide gemäß Formatspezifikation

## Constraints

XHTML-Element	XHTML-Attribut	ODF-Attribut	ODF-Element
html			office:document
body			office:body
head			office:meta
link		meta:name=link	meta:userdefined
		meta:value-type="false"	
meta	content	content	meta:userdefined
	meta_name	meta name	
		meta:value-type="true"	
title	content	content	dc:title
div	class	style-name	text:section
	id	text name	
br			text:line-break
p	class	text: class-names	text:p
	id	text: id	

Formale Verifikation



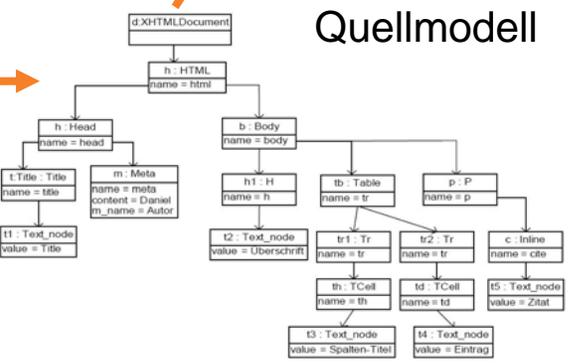
Quellformat

```

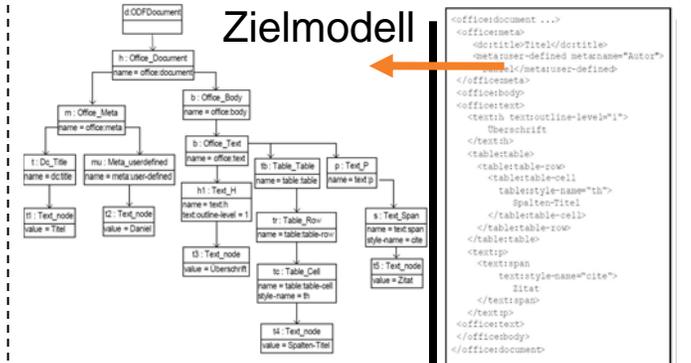
<html>
<head>
<title>Titel</title>
<meta name="Autor" content="Daniel"/>
</head>
<body>
<h1>Überschrift</h1>
<table>
<tr>
<td>Spalten-Titel</td>
</tr>
<tr>
<td>Eintrag</td>
</tr>
</table>
<div>
<cite>Zitat</cite>
</div>
</body>
</html>
    
```

Quellstruktur

Quellmodell



Zielmodell



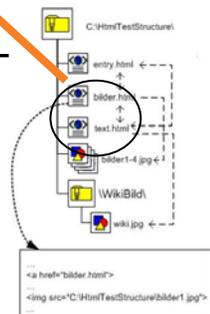
Zielstruktur

Zielformat

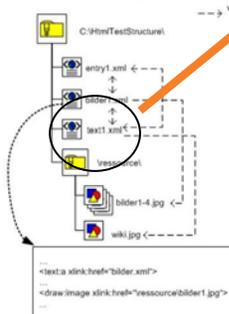
```

<office:document ...>
<office:meta ...>
<dc:title>Titel</dc:title>
<meta:user-defined meta:name="Autor" />
</office:meta ...>
</office:document ...>
<office:body ...>
<text:h text:outline-level="1">
Überschrift
</text:h>
<table:table-row ...>
<table:table-cell ...>
Spalten-Titel
</table:table-cell ...>
</table:table-row ...>
<table:table-cell ...>
text:span
text:style-name="cite"
Zitat
</table:table-cell ...>
</table:table-row ...>
</office:body ...>
</office:document ...>
    
```

Quell-Repräsentation (Referenzdok.)



Ziel-Repräsentation (Referenzdok.)



Transformation  
 (in JAVA implementiert)

## Güte der Transformationsresultate:

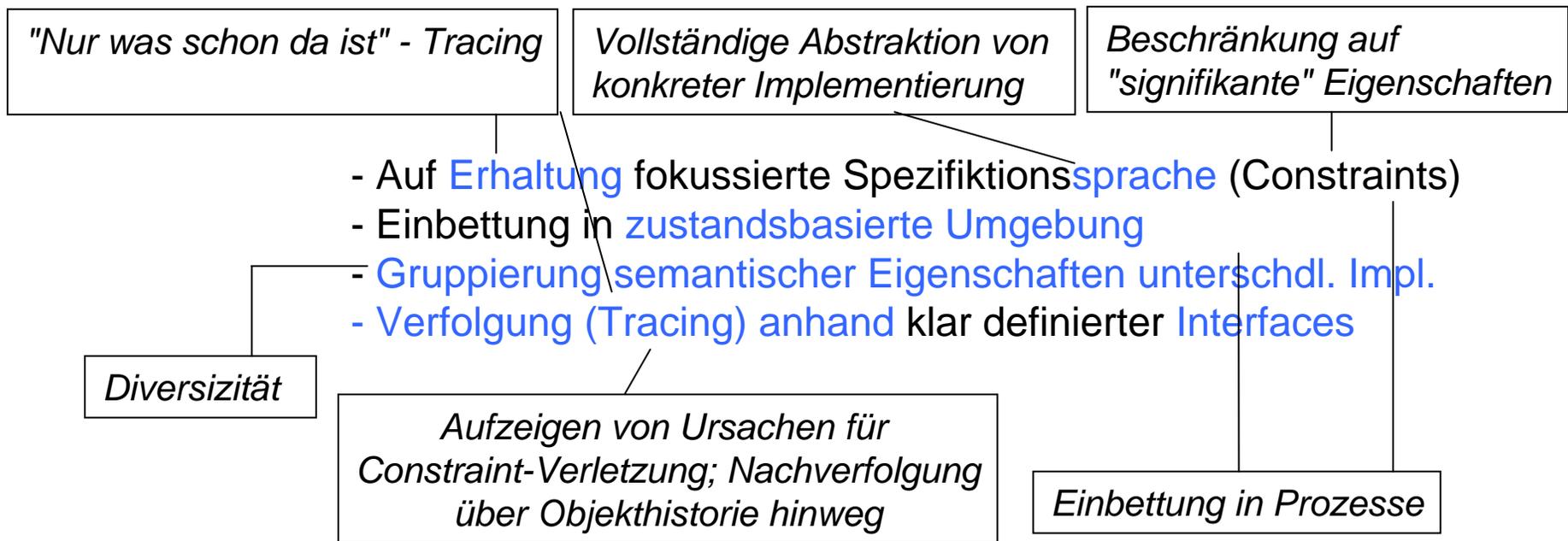
Hin- und Rückrichtung entsprechen den Anforderungen (Erhaltung Linkkonsistenz, Inhalte, optisches Erscheinungsbild) mit kleinen Abstrichen bei eingebetteten Bildern

## Allgemeine Erfahrungen:

- (1) - Ansatz über Konzept- und Rollenidentifikation eingängig und praktikabel
  - Ansatz hilft bei Implementierung (Fehler bei Layout-Transformation konnten behoben werden)
- (2) Dokumentinhalte und Anforderungen an Erhaltung konnten vollständig formal spezifiziert werden
  - Anzahl Konzepte: 10 (je Richtung)
  - Anzahl Constraints: 27 (je Richtung)
- (3) Implementierungstechnische Einbettung in das formale Framework mit entsprechendem Einarbeitungsaufwand gut möglich
- (4) - vertretbares Laufzeitverhalten der Implementierung
  - mit etwas formalem Aufwand sind Linkkonsistenz und Transformation der Style-Informationen sogar vollautomatisiert handhabbar (Laufzeitverhalten grenzwertig → (Optimierungspotential))

**Grundidee:** Erhaltung definiert über Vorher-Nachher Vergleich und Abstraktion (Kernkonzept der Informatik)

## Merkmale und Bestandteile des entwickelten Systems:



Vielen Dank für Ihre Aufmerksamkeit...



**Thomas Triebsees**  
Universität der Bundeswehr München  
Fakultät für Informatik

[Thomas.Triebsees@unibw.de](mailto:Thomas.Triebsees@unibw.de)