

---

# Planung des Forschungsdaten-Managements:

## Einführung

---

Jens Ludwig

SUB Göttingen

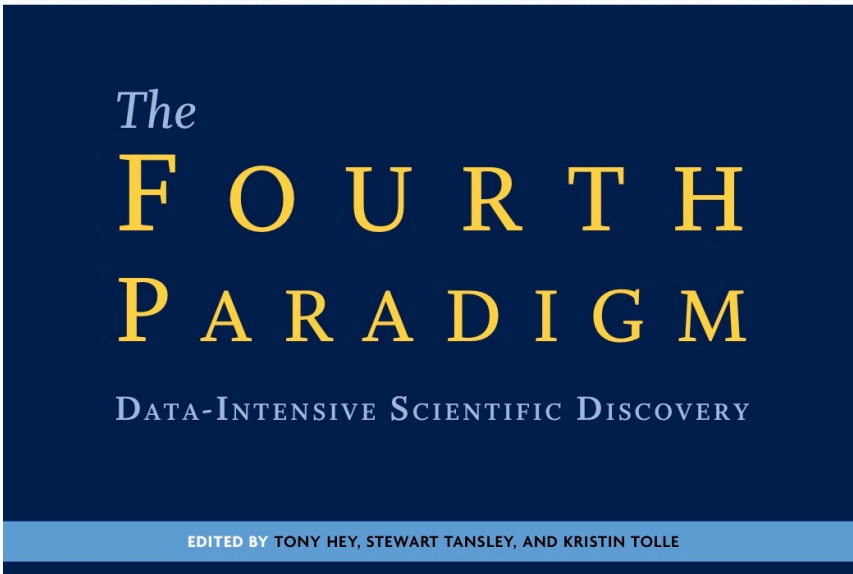
[ludwig@sub.uni-goettingen.de](mailto:ludwig@sub.uni-goettingen.de)

6. Dezember 2011, Göttingen

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung



# Forschungsdaten als Thema der Informationsinfrastruktur

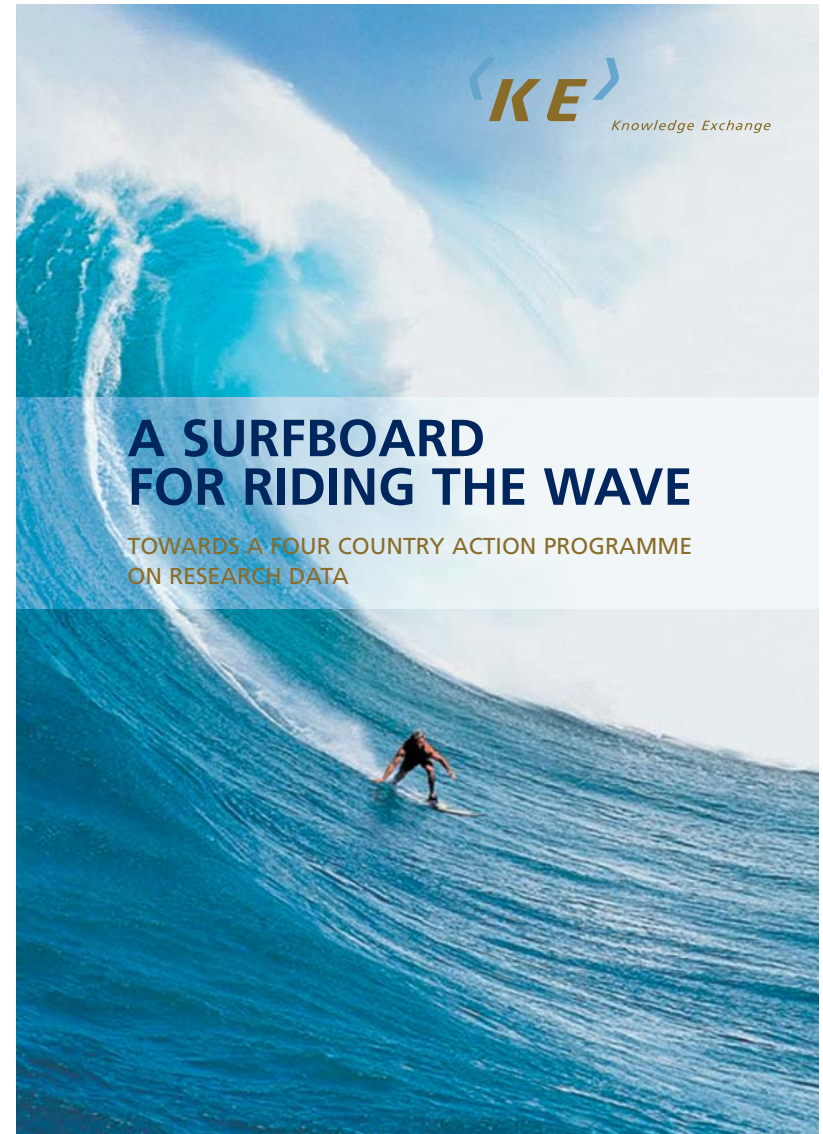


Kommission Zukunft der  
Informationsinfrastruktur

## Gesamtkonzept für die Informationsinfrastruktur in Deutschland

Empfehlungen der  
Kommission Zukunft der Informationsinfrastruktur im Auftrag der  
Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder

April 2011





# Forschungsdaten als Thema der Informationswissenschaft



„[...] Wenn aus Projektmitteln systematisch (Mess-)Daten erhoben werden, die für die Nachnutzung geeignet sind, legen Sie bitte dar, welche Maßnahmen ergriffen wurden bzw. während der Laufzeit des Projektes getroffen werden, um die Daten nachhaltig zu sichern und ggf. für eine erneute Nutzung bereit zu stellen [...]“

*DFG, Antragsformular, Kapitel Umgang mit den im Projekt erzielten Forschungsdaten, seit 2010*

„Proposals submitted or due on or after January 18, 2011, must include a supplementary document of no more than two pages labeled ‚Data Management Plan‘. This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results.“

*NSF, Data Management Plan Requirements, seit 2010*

„[...] Wenn aus Projektmitteln systematisch (Mess-)Daten erhoben werden, die für die Nachnutzung geeignet sind, legen Sie bitte dar, welche Maßnahmen ergriffen wurden bzw. während der Laufzeit des Projektes getroffen werden, um die Daten nachhaltig zu sichern und ggf. für eine erneute Nutzung bereit zu stellen [...]“

*DFG, Antragsformular, Kapitel Umgang mit den im Projekt erzielten Forschungsdaten, seit 2010*

„Proposals submitted or due on or after January 18, 2011, must include a supplementary document of no more than two pages labeled ‚Data Management Plan‘. This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results.“

*NSF, Data Management Plan Requirements, seit 2010*



# Datenmanagementpläne

COMPONENT	MIT	DCC	ANDS	CESSDA	JPL	QUT	DAAC	Cornell
<b>OBJECT CHARACTERISTICS</b>								
Data type (raw, computational, observational)	✓	✓	X	X	X	✓	✓	✓
Naming conventions	✓	X	✓	✓	X	✓	✓	X
Format types	✓	✓	✓	X	✓	✓	✓	✓
<b>CONTEXTUAL INFORMATION</b>								
Software (resilience, obsolescence)	✓	✓	X	X	X	✓	✓	X
Required and stored metadata	✓	✓	✓	X	X	✓	✓	✓
Documentation (may relate to metadata)	✓	✓	✓	✓	✓	✓	✓	✓
Intellectual control (copyright, ownership)	✓	✓	✓	✓	X	✓	✓	X
Transfer of Responsibility	?	?	?	?	?	?	?	✓
<b>ENVIRONMENTAL / STORAGE FACTORS</b>								
Data retention	✓	✓	✓	X	X	✓	✓	✓
Anticipated volume	✓	✓	X	X	✓	X	✓	X
Backup	✓	✓	✓	✓	X	✓	✓	✓
<b>USAGE</b>								
Access restrictions	✓	✓	✓	X	X	✓	X	✓
Privacy / anonymization	✓	✓	✓	✓	X	✓	✓	✓
Security	✓	✓	✓	✓	X	✓	✓	✓
Audience	✓	✓	✓	X	X	X	✓	X



Projekt von Vertretern virtueller  
Forschungsumgebungen im D-Grid zu  
organisatorischer und technischer Nachhaltigkeit:

- Astronomie
- Geisteswissenschaften
- Hochenergiephysik
- Klimawissenschaften
- Medizin



# Forschungsdaten als Thema von WissGrid



AP 2:  
Blaupausen  
für neue  
Community-  
Grids

Checkliste und  
Leitfaden zum  
Forschungs-  
daten-  
Management

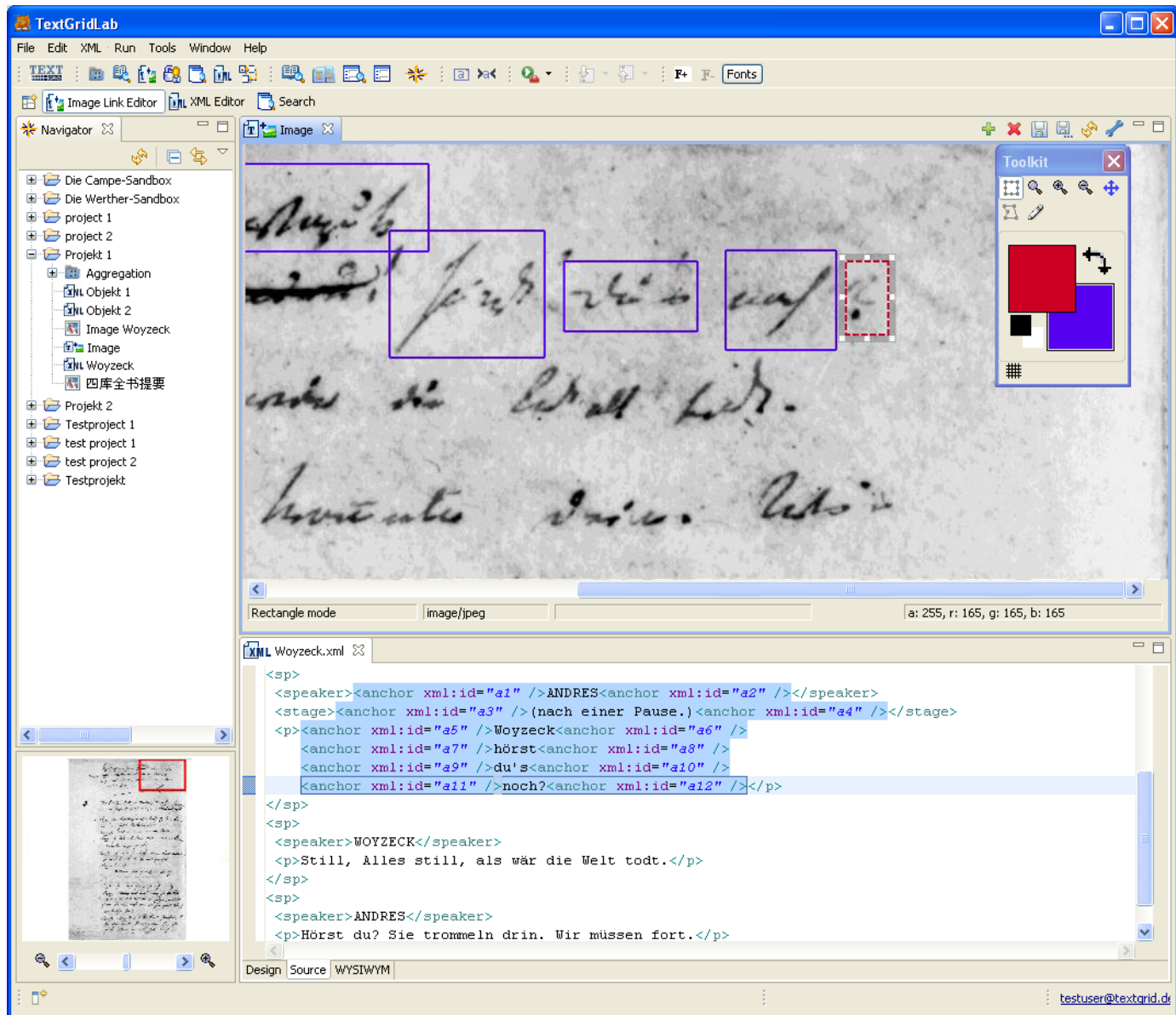
AP 3: Langzeit-  
archivierung von  
Forschungsdaten



# Was sind Forschungsdaten?

Depth [m]	Depth c [mcd]	Depth top [m]	Depth bot [m]	Label
d180 [per mil PDB]	Uvigerina sp. d180 [per mil PDB]			
8.68	10.845	10.83	10.86	177-1090D-2H-1,128-130.5 3.64
8.78	10.945	10.93	10.96	177-1090D-2H-1,138-140.5 3.56
8.88	11.030	11.03		177-1090D-2H-1,148-150.5 3.57
8.98	11.145	11.13	11.16	177-1090D-2H-2,8-10.5 3.55
9.08	11.245	11.23	11.26	177-1090D-2H-2,18-20.5 3.45
9.18	11.345	11.33	11.36	177-1090D-2H-2,28-30.5 3.36
9.28	11.445	11.43	11.46	177-1090D-2H-2,38-40.5 3.21
9.38	11.545	11.53	11.56	177-1090D-2H-2,48-50.5 3.25
9.48	11.645	11.63	11.66	177-1090D-2H-2,58-60.5 3.37
9.58	11.745	11.73	11.76	177-1090D-2H-2,68-70.5 3.38
9.68	11.845	11.83	11.86	177-1090D-2H-2,78-80.5 3.07
9.78	11.945	11.93	11.96	177-1090D-2H-2,88-90.5 3.06
9.88	12.045	12.03	12.06	177-1090D-2H-2,98-100.5 2.75
9.98	12.145	12.13	12.16	177-1090D-2H-2,108-110.5 2.72
10.08	12.245	12.23	12.26	177-1090D-2H-2,118-120.5 2.63
10.18	12.345	12.33	12.36	177-1090D-2H-2,128-130.5 2.65
10.28	12.445	12.43	12.46	177-1090D-2H-2,138-140.5 2.62
10.38	12.530	12.53		177-1090D-2H-2,148-150 2.65
10.48	12.645	12.63	12.66	177-1090D-2H-3,8-10.5 2.77
10.58	12.745	12.73	12.76	177-1090D-2H-3,18-20.5 3.12
10.68	12.845	12.83	12.86	177-1090D-2H-3,28-30.5 3.24

# Was sind Forschungsdaten?



The screenshot displays the TextGridLab application window. The main area shows a handwritten document with several rectangular annotations in purple and red. A 'Toolkit' panel on the right provides editing tools. Below the document, an XML editor shows the following code:

```

XML Woyzeck.xml
<sp>
<speaker><anchor xml:id="a1" />ANDRES<anchor xml:id="a2" /></speaker>
<stage><anchor xml:id="a3" />[nach einer Pause.]<anchor xml:id="a4" /></stage>
<p><anchor xml:id="a5" />Woyzeck<anchor xml:id="a6" />
<anchor xml:id="a7" />hörst<anchor xml:id="a8" />
<anchor xml:id="a9" />du's<anchor xml:id="a10" />
<anchor xml:id="a11" />noch?<anchor xml:id="a12" /></p>
</sp>
<sp>
<speaker>WOYZECK</speaker>
<p>Still, Alles still, als wär die Welt todt.</p>
</sp>
<sp>
<speaker>ANDRES</speaker>
<p>Hörst du? Sie trommeln drin. Wir müssen fort.</p>

```

The status bar at the bottom indicates 'Rectangle mode', 'image/jpeg', and coordinates 'a: 255, r: 165, g: 165, b: 165'. The bottom right corner shows the email address 'testuser@textgrid.de'.



# Was sind Forschungsdaten?

ELAN 3.9.1 File Edit Annotation Tier Type Search View Options Window Help

Elan - pear story.eaf

Grid Text Subtitles **Audio Recognizer** Video Recognizer Metadata Controls

Recognizer: Tag vowels (volume peaks of voiced timespans) File(s): pear.wav

Selections: Add Remove Add Tier...

Parameters: Pitch ceiling [Hz], eg. 300 male / 550 female (150.0 - 2500.0) 604.8; Intensity change [dB] to start/end a peak (0.5 - 5.0) 2.0; Minimum amplitude (0..1) for pitch analysis (0.01 - 0.3)

Segmentations: Create Tier(s)... Progress: Ready Start Report...

00:00:16.840 Selection: 00:00:00.000 - 00:00:00.000 0

#intensity/dB: 76.005, 55.3015, 32.598; #pitch/Hz: 604.475, 340.3425, 76.21

00 00:00:09.000 00:00:10.000 00:00:11.000 00:00:12.000 00:00:13.000 00:00:14.000 00:00:15.000 00:00:16.000 00:00:17.000 00:00:18.000 00:00:19.000 00:00:20.000

Event (1)  
 Clause Transcri (16)  
 Motion (16)  
 Gesture # (23)  
 Gs Hand (23)

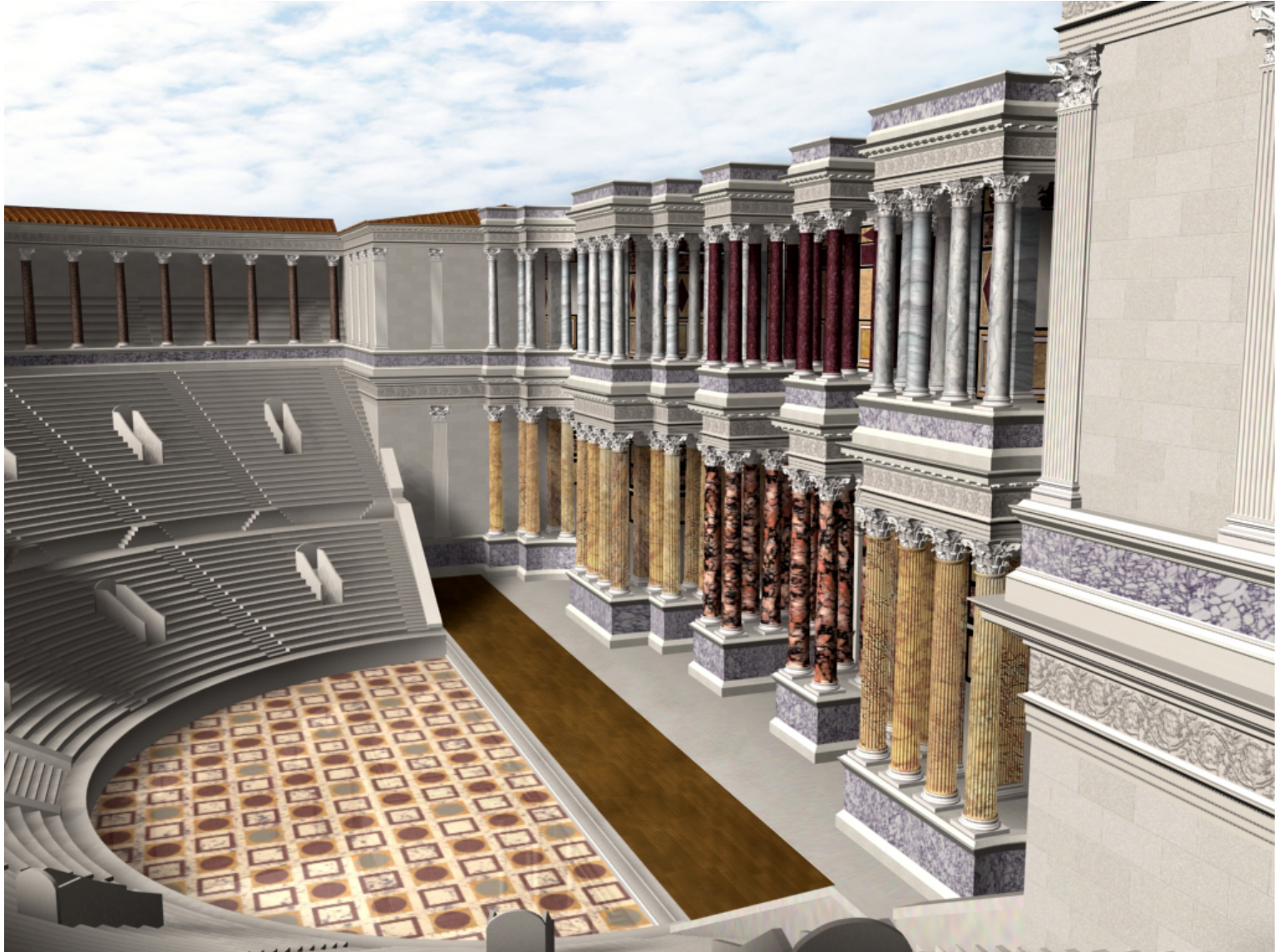
and so he climbs up a tree and he starts with the ladder and he starts picking pears off the tree and he puts the pears into an apron

motion non-motion non-motion

gestu gestu gesture 4 gestur gesture 6 gesture 7 gesture gesture 9 gesture 1

R R R R R B B B

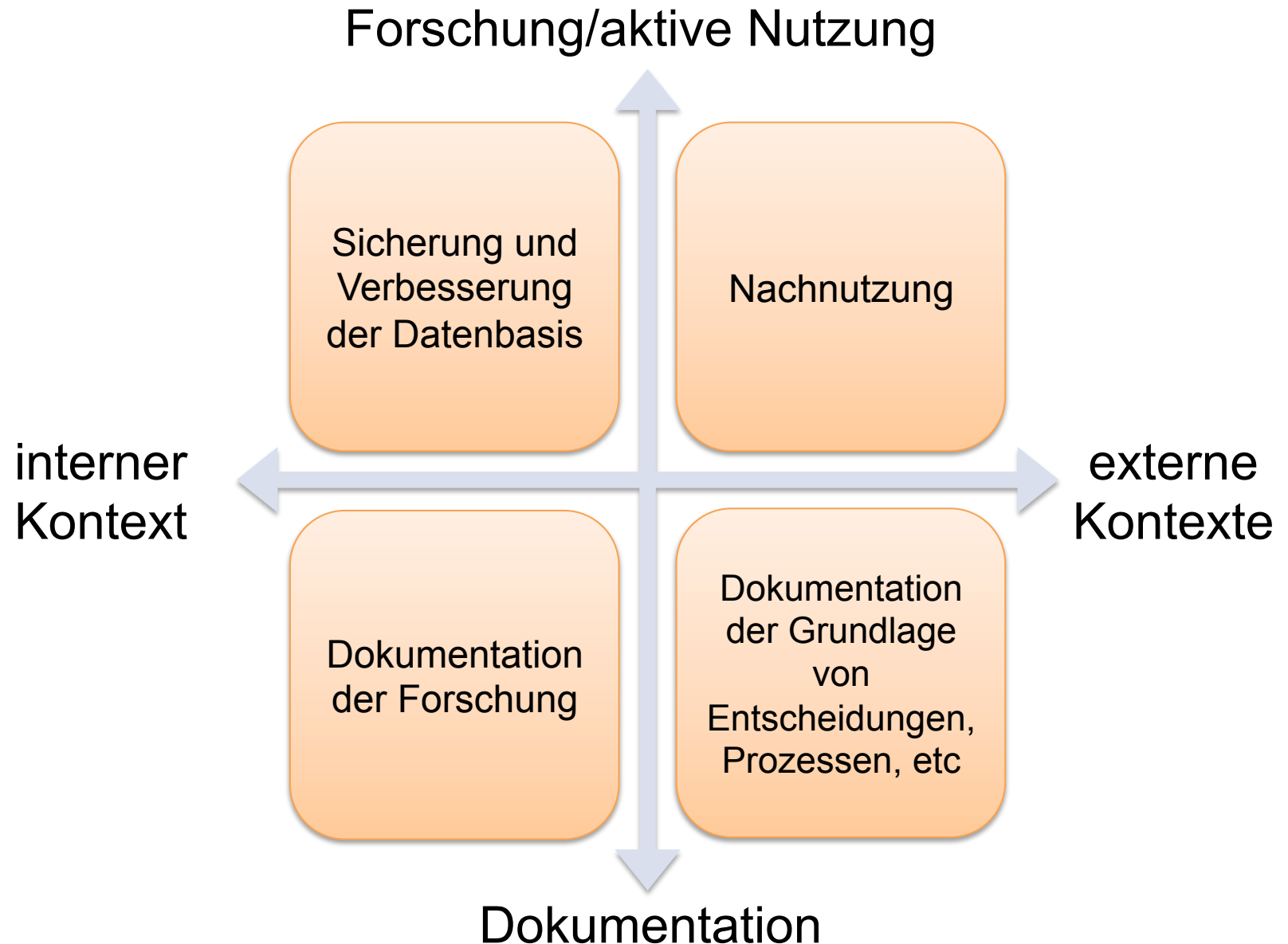
# Was sind Forschungsdaten?







# Gründe für Forschungsdaten-Management





# Aspekte des Forschungsdaten-Managements

Verantwortlichkeiten,  
Infrastrukturpartner, finanzielle  
Nachhaltigkeit, Rechte, ...

**intellektuelle Nachnutzbarkeit**  
Kontextinformationen, Objektmodelle, inhaltliche  
Versionierung, ...

**technische Nachnutzbarkeit**  
technische Qualitätskontrollen, Konvertierungen, ...

**Bitstream Preservation: technische Stabilität**  
genug unabhängige Kopien, Integritätsprüfung, ...

# Forschungsdaten-Management in unterschiedlichen Kontexten



**Private  
Domäne**

**Gruppendomäne**

**Dauerhafte  
Domäne**

**Zugang und  
Nachnutzung**

Daten-  
management

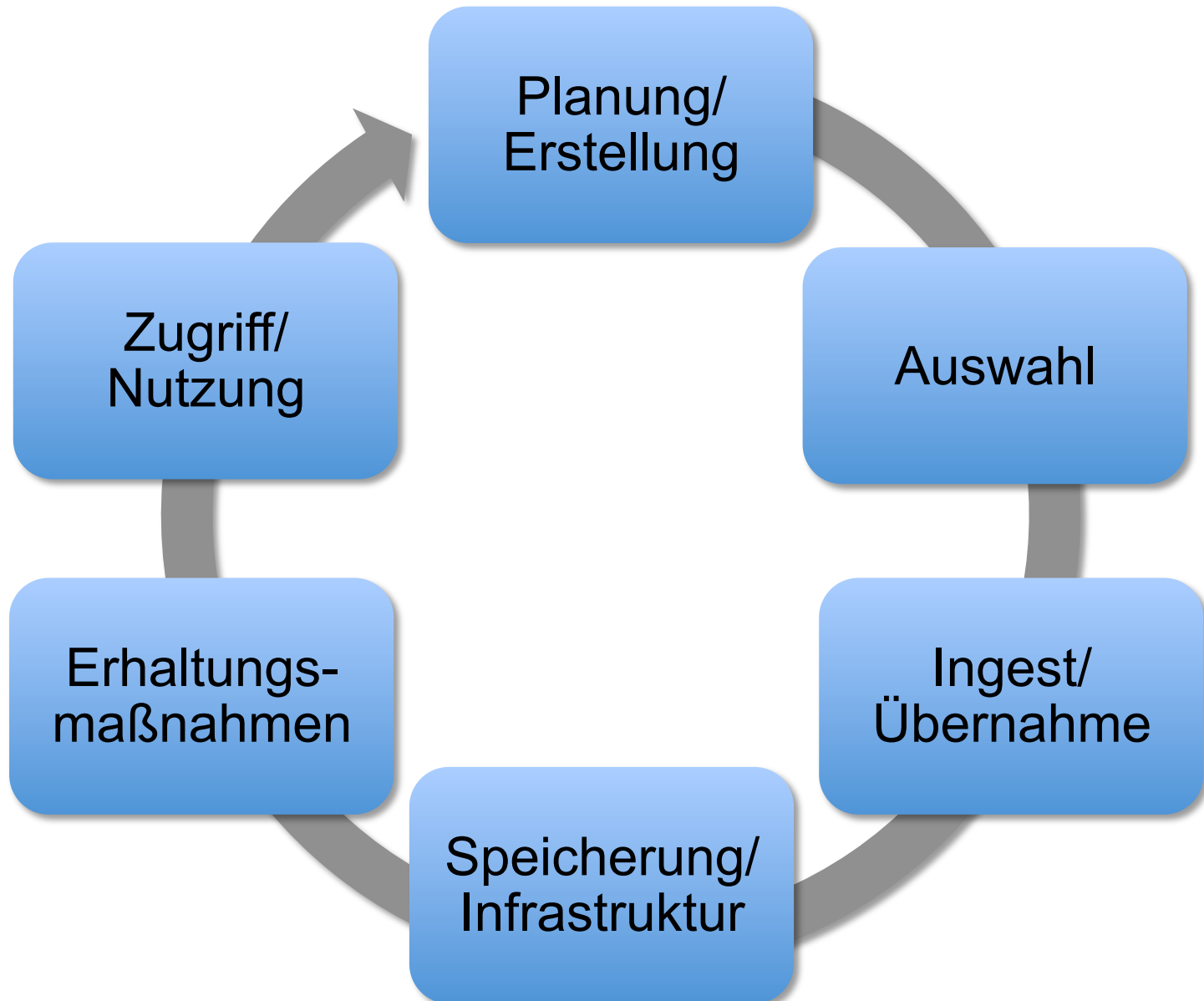
erweitertes  
Daten-  
management

Zugriffs-  
bedingungen

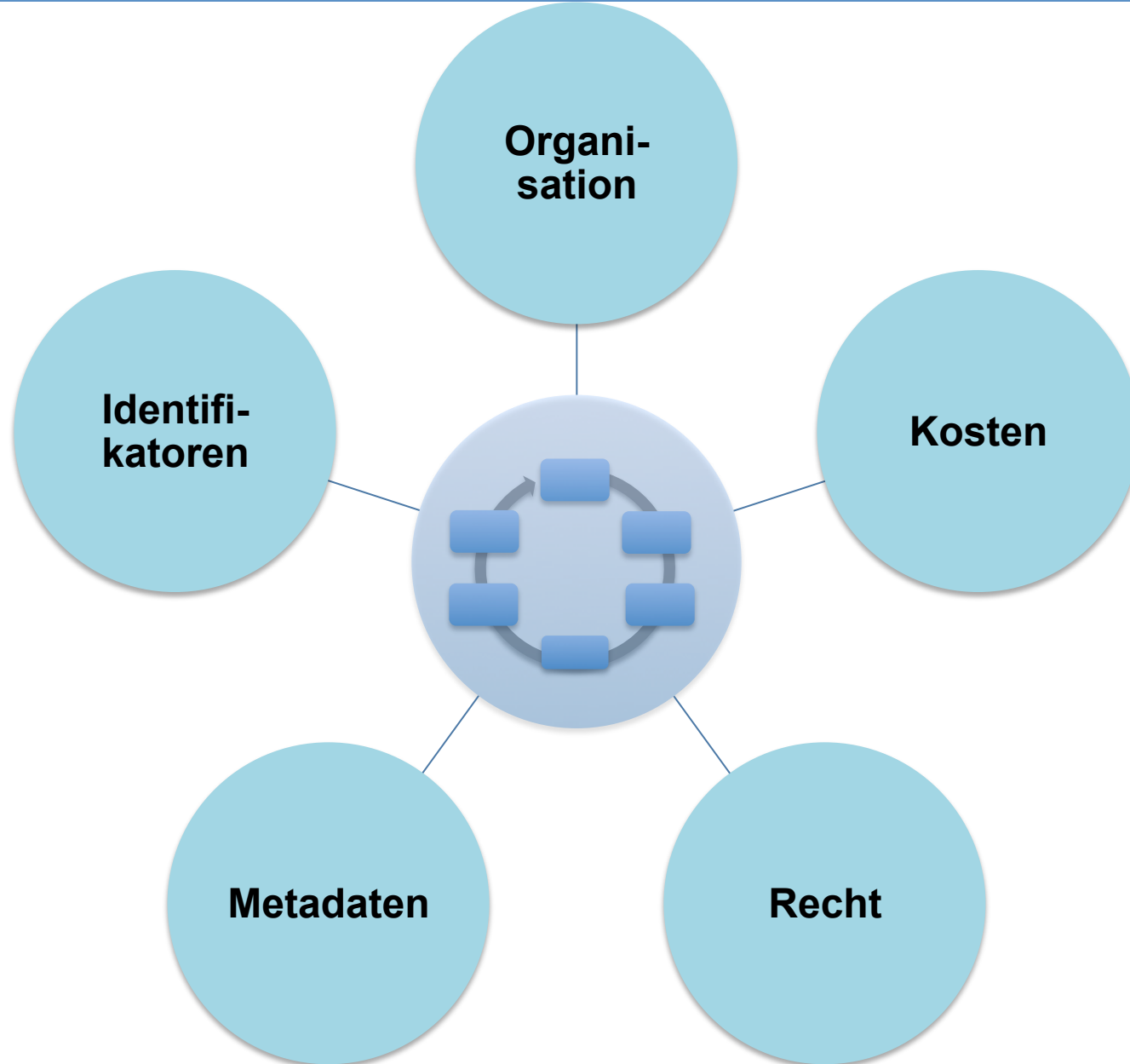
## WissGrid Checkliste

- anzupassen an Kontext der Disziplin und Domäne
- Zusammenarbeit von Wissenschaft und Infrastruktur

# Checkliste und Leitfaden: Aufgabenbereiche im Lebenszyklus



# Checkliste und Leitfaden: Übergreifende Aufgabenbereiche







Workshop-Feedback

Publikation als Entwurf zur öffentlichen  
Kommentierung auf Webseite

Englische Übersetzung

Erfahrungen

Überarbeitete Version am Ende des Projekts



## Fragen für alle Aufgabenbereiche

---

Wie wird der Planungsprozess organisiert?

Wie granular sollte die Planung jeweils sein? (z.B. unterschiedlich granulare Checklisten sinnvoll?)

Mit welchen Kontexten/Domänen haben Sie zu tun?

Welche Erfahrungen haben Sie mit der Planung? Nutzen Sie bestimmte Instrumente? Was würden Sie sich wünschen?



# Agenda

11:00 – 11:30	Begrüßung und Einführung	Jens Ludwig (SUB)
11:30 – 13:00	Vorstellung und Diskussion: Aufgaben im Lebenszyklus von Forschungsdaten	
	Datenproduktion, Auswahl und Ingest	Gabriel Stöckle (ZAH)
	Speicherung, Erhaltung und Nachnutzung	Torsten Rathmann (DKRZ)
13:00 – 13:45	Mittagsimbiss	
13:45 – 15:00	Vorstellung und Diskussion: übergreifende Aufgabenbereiche	
	Management, Recht und Finanzierung	Harry Enke (AIP)
	Metadaten und Identifier	Timo Gnadl (SUB)
15:00 – 15:30	Pause	
15:30 – 16:30	Zusammenfassung und Abschlussdiskussion: Wie kann Forschungsdaten-Management sinnvoll geplant werden?	Moderation: Jens Ludwig (SUB)



# Bild- und Quellennachweise

---

Folien 2 – 6: Die Cover entstammen den jeweils gezeigten Quellen und unterliegen den Lizenzen der jeweiligen Rechteinhaber.

Folie 7: DataOne, Data Management Plans, <https://www.dataone.org/plans>

Folie 10: Becquey, S; Gersonde, R (2002): Distribution of planktonic foraminifera and paleotemperature reconstruction for the Subantarctic Zone of the South Atlantic. doi:10.1594/PANGAEA.706226, <http://doi.pangaea.de/10.1594/PANGAEA.706226>

Folie 11: TextGrid, TextGrid-Bild-Link-Editor, <http://www.textgrid.de>

Folie 12: Max Planck Institute for Psycholinguistics, ELAN, complex annotations on video and audio resources <http://www.lat-mpi.eu/tools/elan/>

Folie 13: King's College London, Visualisation Lab, Restoring the Theater of Pompey in the 21st Century: The Blazeby Reconstructions, <http://www.pompey.cch.kcl.ac.uk/Blazeby.htm>

Alle anderen Grafiken sind im WissGrid-Projekt erstellt worden.