

# The PARSE.Insight Gap Analysis

PARSE.Insight Final Symposium

*Carrying the Flame Forward*

*Paris, 25<sup>th</sup> June 2010*

Matthias Hemmje, Moritz Gomm,  
Björn Werkmann, **Holger Brocks**  
(FernUniversität in Hagen)

# Overview

---

- Executive Summary
- Objectives
- Working Practices
- Methodology & Gap Analysis Framework
- IT-Tool
- Identified Gaps
  - Libraries Community
  - Publisher Community (Summary)
- Discussion & Summary

# Executive Summary

---

- The **Gap Analysis Tool** supports the analysis of survey data to identify gaps in awareness, knowledge, implementation and commitment of digital preservation within different communities.
- **Gap analyses** were performed within different communities, using a specially developed tool to explore the gaps between vision and reality.

Contribution *"Proposing a Framework and a Visual Tool for Analysing Gaps in Digital Preservation Practice - A Case Study among Scientific Libraries in Europe"* accepted at iPRES2010.

# Overview

---

- Executive Summary
- Objectives
- Working Practices
- Methodology & Gap Analysis Framework
- IT-Tool
- Identified Gaps
  - Libraries Community
  - Publisher Community (Summary)
- Discussion & Summary

# Objectives

---

*... identification of gaps within stakeholder communities across various dimensions*

- Validate results from WP3 and identify gaps not covered by survey
- Develop a generic gap analysis framework for DP
- Implement appropriate IT-Tool
- Analyze gaps in DP of representative domains (Publishers, Libraries)
- Validate results by consultations with PARSE.Insight communities

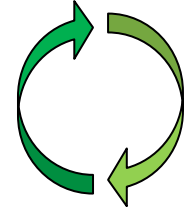
# Overview

---

- Executive Summary
- Objectives
- Working Practices
- Methodology & Gap Analysis Framework
- IT-Tool
- Identified Gaps
  - Libraries Community
  - Publisher Community (Summary)
- Discussion & Summary

# Working Practices

- Development of framework and gap analysis methodology
- Validation within expert workshops
- Implement IT-Tool based on established methods
- Use framework and IT-Tool on PARSE survey data
  - Domain workshops with PARSE community experts
  - Application of framework and IT-Tool on survey data
  - Analyse and iteratively feed back results to expert
- Enhance IT-Tool and framework
- Publish Results (Gap Analysis Report)



# Overview

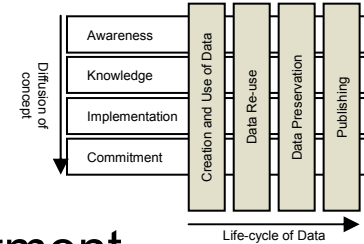
---

- Executive Summary
- Objectives
- Working Practices
- Methodology & Gap Analysis Framework
- IT-Tool
- Identified Gaps
  - Libraries Community
  - Publisher Community (Summary)
- Discussion & Summary

# Methodology

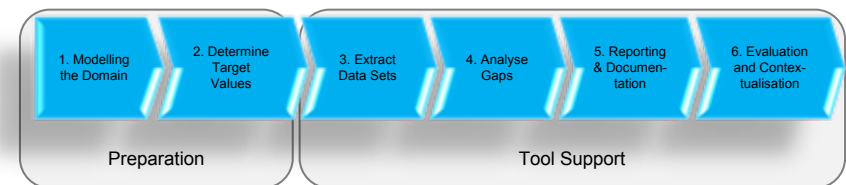
## Gap Analysis Framework

- Diffusion of Concepts and Lifecycle of Data
- DP Awareness, Knowledge, Implementation and Commitment

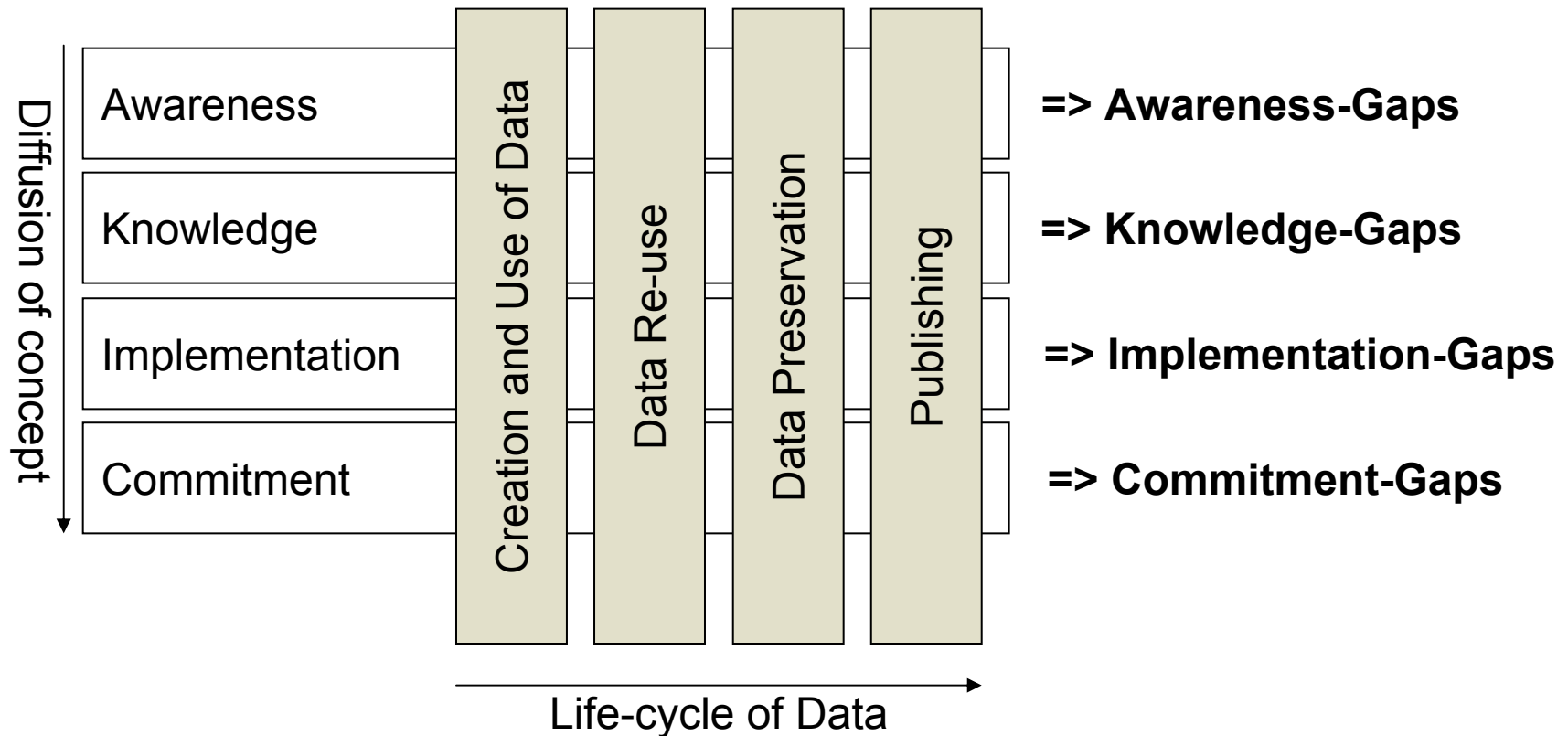


## Methodology

- 1) Modelling the domain
- 2) Define targets values for gaps
- 3) Load IT-tool with data sets from survey
- 4) Analyse the gaps
- 5) Reporting/documentation
- 6) Evaluation/contextualisation

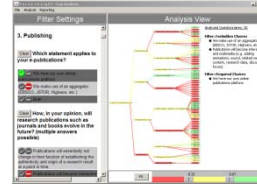
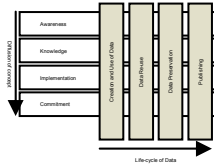


# Gap Analysis Framework



# Gap Analysis Process

LIBER-experts  
using framework  
and Excel-Tool



Interpretation, feedback  
and validation  
by LIBER-experts

1. Modelling  
the Domain

2. Determine  
Target  
Values

3. Extract  
Data Sets

4. Analyse  
Gaps

5. Reporting  
& Documen-  
tation

6. Evaluation  
and Context-  
tualisation

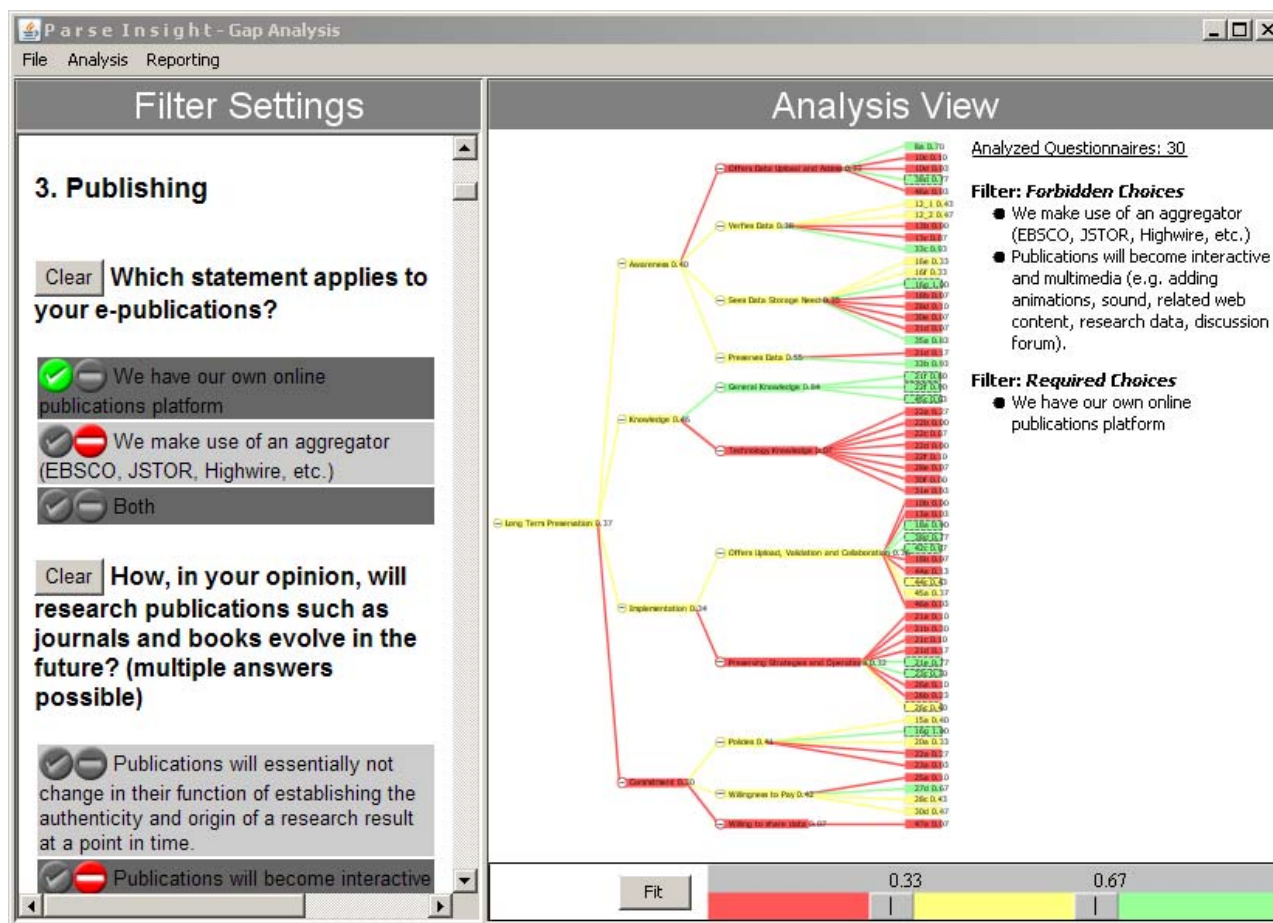
Support by FUH / IT-Development-Team

# Overview

---

- Executive Summary
- Objectives
- Working Practices
- Methodology & Gap Analysis Framework
- IT-Tool
- Identified Gaps
  - Libraries Community
  - Publisher Community (Summary)
- Discussion & Summary

# Gap Analysis IT-Tool

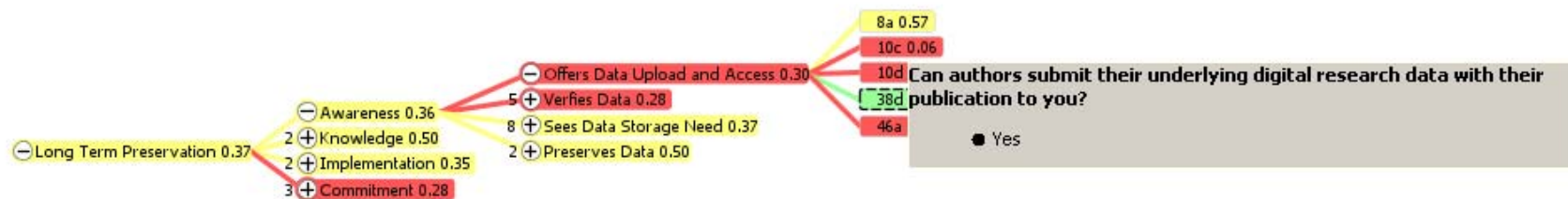


Filled-out  
Surveys  
(Data)

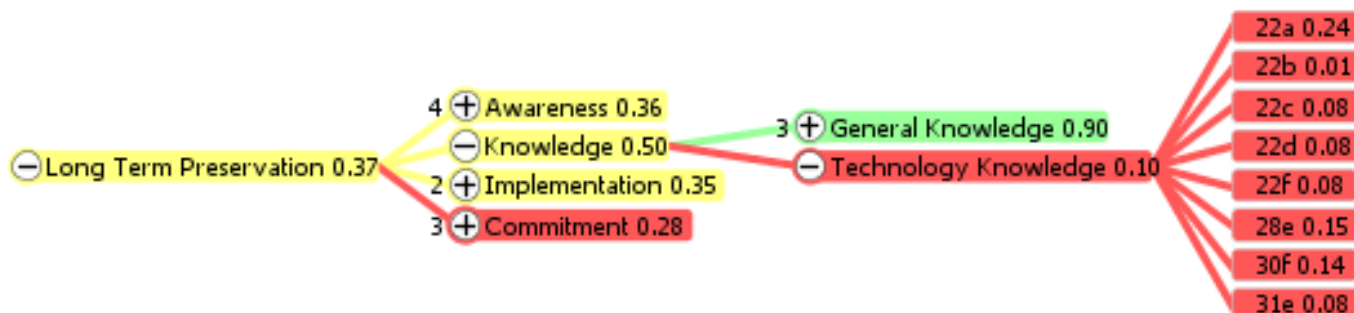
Gap Analysis Tool

# Gap Analysis IT-Tool

## Visual Drill-Down



*Drill-down with question details*



*Degree-of-Interest-based pruning of context around focused "Technology Knowledge" branch*

# Overview

---

- Executive Summary
- Objectives
- Working Practices
- Methodology & Gap Analysis Framework
- IT-Tool
- Identified Gaps
  - Libraries Community
  - Publisher Community (Summary)
- Discussion & Summary

# Top-Level Overview: Libraries (LIBER)



- Only **awareness** is visualized as positive as was assumed by the experts.
- **Commitment** is on a modest level, due to some preservation policies that libraries have implemented.
- A larger gap exists in the area of **knowledge** than pure looking at the survey results had indicated.
- The **implementation** gap that was assumed can be proven.

# Community of Libraries

## Summary of Findings

- Libraries that have **policies and infrastructures** are more *committed* to digital preservation and are better prepared in terms of *implementation*.
- The **kind of material** that libraries collect has no impact on libraries' preparedness for digital preservation.
- The larger the **amount of data** that a library has to deal with, the smaller the gaps in the area of *implementation* and *commitment* are.
- Those institutions that do not have **preservation strategies** in place have higher gaps in knowledge, implementation and commitment.
- Libraries that think that more **training** is needed shows the smallest gap in the *knowledge* area. They obviously are aware of the complexity of digital preservation.
- Libraries that are less confident that their infrastructure will **scale with future requirement** have bigger gaps in *knowledge* and *commitment*.
- The gap analysis with the tool proved the assumption right that there is **mainly an implementation gap**, especially in the implementation of preservation strategies.
- **Overall, the results indicate a gap between well prepared and less prepared libraries. The less prepared libraries must be attentive that they do not fall behind.**

# Libraries (1/5)

## Policies

Selected Findings

1	"Do you have policies and an infrastructure to guarantee that data are properly managed and maintained to ensure continued access and usability?"
Yes	<pre> graph LR     Root[Root 0,72] --- Awareness[Awareness 0,94]     Root --- Knowledge[Knowledge 0,66]     Root --- Implementation[Implementation 0,57]     Root --- Commitment[Commitment 0,72]     Awareness --- A1[14 + Recognises reasons for digital preservation 0,94]     Awareness --- A2[21 + Recognizes importance of digital preservation 0,94]     Knowledge --- K1[8 + Knows about infrastructure requirements 0,66]     Implementation --- I1[7 + Has data and access management strategies implemented 0,76]     Implementation --- I2[5 + Has preservation strategies implemented 0,38]     Commitment --- C1[8 + Recognises the responsibility of libraries 0,57]     Commitment --- C2[7 + Has preservation policies in place 0,87]         </pre>
No	<pre> graph LR     Root[Root 0,58] --- Awareness[Awareness 0,93]     Root --- Knowledge[Knowledge 0,52]     Root --- Implementation[Implementation 0,34]     Root --- Commitment[Commitment 0,54]     Awareness --- A1[14 + Recognises reasons for digital preservation 0,93]     Awareness --- A2[21 + Recognizes importance of digital preservation 0,92]     Knowledge --- K1[8 + Knows about infrastructure requirements 0,52]     Implementation --- I1[7 + Has data and access management strategies implemented 0,45]     Implementation --- I2[5 + Has preservation strategies implemented 0,23]     Commitment --- C1[8 + Recognises the responsibility of libraries 0,53]     Commitment --- C2[7 + Has preservation policies in place 0,55]         </pre>

# Libraries (2/5)

## Policies cont.

Selected Findings

2	"Do you have policies and procedures in place which determine what kind of data is accepted for storage/preservation by your organisation, and how and when it needs to be submitted?"
Yes	<p>             - Root 0,63             <ul style="list-style-type: none"> <li>- Awareness 0,94                     <ul style="list-style-type: none"> <li>14 + Recognises reasons for digital preservation 0,94</li> <li>21 + Recognizes importance of digital preservation 0,93</li> </ul> </li> <li>- Knowledge 0,54                     <ul style="list-style-type: none"> <li>8 + Knows about infrastructure requirements 0,54</li> </ul> </li> <li>- Implementation 0,43                     <ul style="list-style-type: none"> <li>7 + Has data and access management strategies implemented 0,57</li> <li>5 + Has preservation strategies implemented 0,29</li> </ul> </li> <li>- Commitment 0,62                     <ul style="list-style-type: none"> <li>8 + Recognises the responsibility of libraries 0,53</li> <li>7 + Has preservation policies in place 0,71</li> </ul> </li> </ul> </p>
No	<p>             - Root 0,45             <ul style="list-style-type: none"> <li>- Awareness 0,95                     <ul style="list-style-type: none"> <li>14 + Recognises reasons for digital preservation 0,95</li> <li>21 + Recognizes importance of digital preservation 0,95</li> </ul> </li> <li>- Knowledge 0,35                     <ul style="list-style-type: none"> <li>8 + Knows about infrastructure requirements 0,35</li> </ul> </li> <li>- Implementation 0,21                     <ul style="list-style-type: none"> <li>7 + Has data and access management strategies implemented 0,22</li> <li>5 + Has preservation strategies implemented 0,20</li> </ul> </li> <li>- Commitment 0,31                     <ul style="list-style-type: none"> <li>8 + Recognises the responsibility of libraries 0,33</li> <li>7 + Has preservation policies in place 0,29</li> </ul> </li> </ul> </p>

# Libraries (3/5)

## Volume of data

Selected Findings

4	"Please provide us with an estimate of the volume of stored digital data (excluding backups)"	
100MB-1GB	small	
1GB-1TB		
1TB-1PB	big	

# Libraries (4/5)

## Preservation Strategies

Selected Findings

5a	"Does your organisation have any of the following preservation strategies in place?"
Yes either: Emulation Migration Outsourced	 <ul style="list-style-type: none"> <li>Root 0,73 <ul style="list-style-type: none"> <li>Awareness 0,94 <ul style="list-style-type: none"> <li>14 Recognises reasons for digital preservation 0,93</li> <li>21 Recognizes importance of digital preservation 0,94</li> </ul> </li> <li>Knowledge 0,70 <ul style="list-style-type: none"> <li>8 Knows about infrastructure requirements 0,70</li> </ul> </li> <li>Implementation 0,61 <ul style="list-style-type: none"> <li>7 Has data and access management strategies implemented 0,68</li> <li>5 Has preservation strategies implemented 0,53</li> </ul> </li> <li>Commitment 0,66 <ul style="list-style-type: none"> <li>8 Recognises the responsibility of libraries 0,62</li> <li>7 Has preservation policies in place 0,71</li> </ul> </li> </ul> </li> </ul>
Emulation - No	 <ul style="list-style-type: none"> <li>Root 0,47 <ul style="list-style-type: none"> <li>Awareness 0,96 <ul style="list-style-type: none"> <li>14 Recognises reasons for digital preservation 0,96</li> <li>21 Recognizes importance of digital preservation 0,95</li> </ul> </li> <li>Knowledge 0,32 <ul style="list-style-type: none"> <li>8 Knows about infrastructure requirements 0,32</li> </ul> </li> <li>Implementation 0,26 <ul style="list-style-type: none"> <li>7 Has data and access management strategies implemented 0,30</li> <li>5 Has preservation strategies implemented 0,23</li> </ul> </li> <li>Commitment 0,34 <ul style="list-style-type: none"> <li>8 Recognises the responsibility of libraries 0,31</li> <li>7 Has preservation policies in place 0,37</li> </ul> </li> </ul> </li> </ul>
Migration - nein	 <ul style="list-style-type: none"> <li>Root 0,42 <ul style="list-style-type: none"> <li>Awareness 0,96 <ul style="list-style-type: none"> <li>14 Recognises reasons for digital preservation 0,96</li> <li>21 Recognizes importance of digital preservation 0,96</li> </ul> </li> <li>Knowledge 0,24 <ul style="list-style-type: none"> <li>8 Knows about infrastructure requirements 0,24</li> </ul> </li> <li>Implementation 0,20 <ul style="list-style-type: none"> <li>7 Has data and access management strategies implemented 0,22</li> <li>5 Has preservation strategies implemented 0,18</li> </ul> </li> <li>Commitment 0,28 <ul style="list-style-type: none"> <li>8 Recognises the responsibility of libraries 0,25</li> <li>7 Has preservation policies in place 0,31</li> </ul> </li> </ul> </li> </ul>

# Libraries (5/5)

## Infrastructure

*Selected Findings*

7	"Do you think your current infrastructure will scale with future requirements?"
Yes	<p>Diagram illustrating the factors influencing the 'Yes' response to the question 'Do you think your current infrastructure will scale with future requirements?'. The diagram shows a hierarchy of factors and their associated scores.</p> <ul style="list-style-type: none"> <li>Root (0,71) <ul style="list-style-type: none"> <li>Awareness (0,94) <ul style="list-style-type: none"> <li>Recognises reasons for digital preservation (0,95) [14]</li> <li>Recognizes importance of digital preservation (0,93) [21]</li> </ul> </li> <li>Knowledge (0,72) <ul style="list-style-type: none"> <li>Knows about infrastructure requirements (0,72) [8]</li> </ul> </li> <li>Implementation (0,48) <ul style="list-style-type: none"> <li>Has data and access management strategies implemented (0,62) [7]</li> <li>Has preservation strategies implemented (0,34) [5]</li> </ul> </li> <li>Commitment (0,68) <ul style="list-style-type: none"> <li>Recognises the responsibility of libraries (0,64) [8]</li> <li>Has preservation policies in place (0,72) [7]</li> </ul> </li> </ul> </li> </ul>
No	<p>Diagram illustrating the factors influencing the 'No' response to the question 'Do you think your current infrastructure will scale with future requirements?'. The diagram shows a hierarchy of factors and their associated scores.</p> <ul style="list-style-type: none"> <li>Root (0,59) <ul style="list-style-type: none"> <li>Awareness (0,91) <ul style="list-style-type: none"> <li>Recognises reasons for digital preservation (0,90) [14]</li> <li>Recognizes importance of digital preservation (0,92) [21]</li> </ul> </li> <li>Knowledge (0,50) <ul style="list-style-type: none"> <li>Knows about infrastructure requirements (0,50) [8]</li> </ul> </li> <li>Implementation (0,39) <ul style="list-style-type: none"> <li>Has data and access management strategies implemented (0,52) [7]</li> <li>Has preservation strategies implemented (0,25) [5]</li> </ul> </li> <li>Commitment (0,56) <ul style="list-style-type: none"> <li>Recognises the responsibility of libraries (0,53) [8]</li> <li>Has preservation policies in place (0,59) [7]</li> </ul> </li> </ul> </li> </ul>

# Top-Level Overview Publisher



- The **knowledge** about long term preservation in the entire sample seems to be high
- **Awareness** and **implementation** is on a moderate level
- The biggest gap lies in the **commitment** to actually “live” and promote preservation. Particularly the willingness to share data seems to be low among the publishers.

# Community of Publishers

## Summary of Findings

- Publishers, that allow authors to **submit underlying digital research data** with their publication have higher awareness and implementation levels.
- Publishers without a **validation process** have bigger awareness, implementation and commitment gaps and also their knowledge on DP is still on a moderate level.
- The publishers without a **policy for preservation** have very little awareness and commitment.
- Offering “**online collaboratories**” leads to higher DP knowledge and implementation even though commitment stays on a moderate level.
- Publishers with a “**disaster recovery policy**” have smaller gaps in awareness and also in commitment and implementation.
- Those companies who are not **planning** DP initiatives in the near future have big gaps in awareness, implementation and commitment, whilst those with short term plans have high values in all dimensions except in “implementation”.
- Those publishers who believe, that digital data has **economic value** have a high degree of awareness and knowledge and smaller gaps in implementation and commitment than others.
- **Discrepancies** due to size of publishers, “outsource effect” and emphasis on preservation of data-sets

# Overview

---

- Executive Summary
- Objectives
- Working Practices
- Methodology & Gap Analysis Framework
- IT-Tool
- Identified Gaps
  - Libraries Community
  - Publisher Community (Summary)
- Discussion & Summary

# Discussion

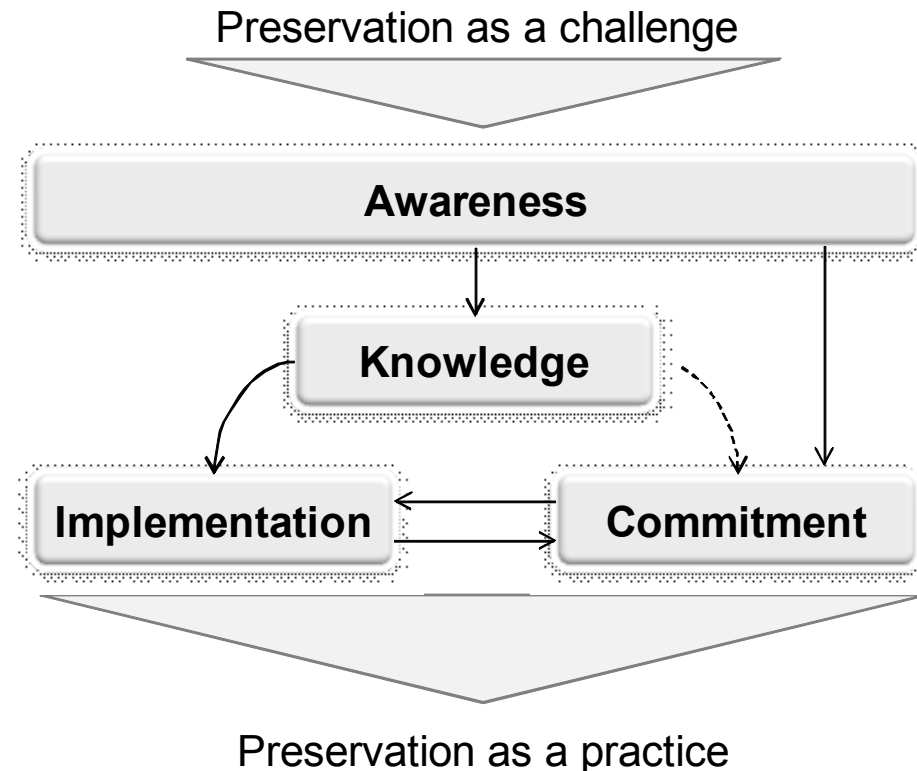
---

- Gap Analysis operates on a different level than the survey
  - Aggregation of survey results along predefined dimensions
  - “Scoring model” allows to “calculate” gaps
  - Visualization of gaps with interactive Drill-Down
- Majority of results is consistent with Insight-Report
- Variances occur due to different perspectives
  - E.g. Gap: “Technical Knowledge amongst publishers is low”
  - ➔ Most publishers outsource DP, thus knowledge less necessary
- Some robustness due to aggregation of multiple items

Overall the Gap Analysis gives deeper insight into correlations and stimulates discussion amongst stakeholders

# Summary

- Gap analysis methodology and IT-Tool finalized
  - Supporting all phases: preparation, visualization, analysis and documentation of gaps
- Methodology and tool validated with “Publishers” and “Libraries”
  - Gaps & Challenges identified
  - Validation and contextualization through workshops and consultation
- Full details
  - D4.3 Gap Analysis Final Report
  - D4.4 Gap Analysis Tool plus supporting database



Thank you!

Any Questions?

[Holger.Brocks@FernUni-Hagen.de](mailto:Holger.Brocks@FernUni-Hagen.de)