

# Do we need metadata?

## An On-line Survey in German Archives

Marcel Ruhl, University of Applied Sciences Potsdam

- 2<sup>nd</sup> International Workshop on Semantic Digital Archives (SDA 2012) -

# Outline

- Introduction
- Design of the survey
- Findings
- Conclusions

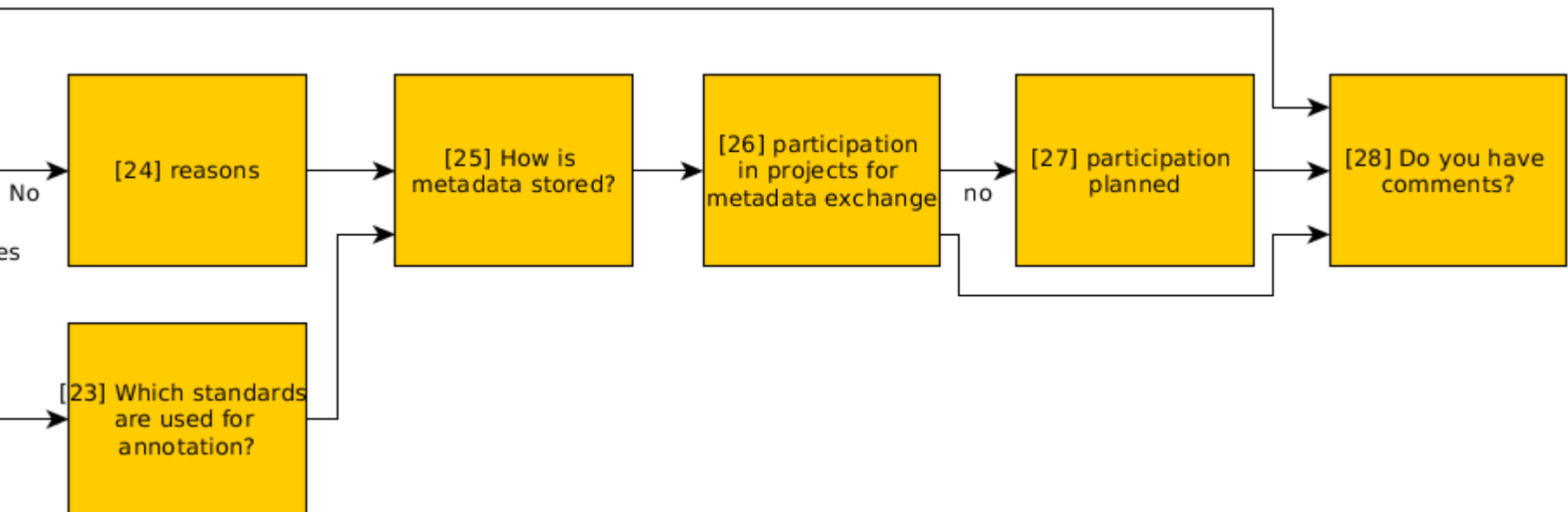
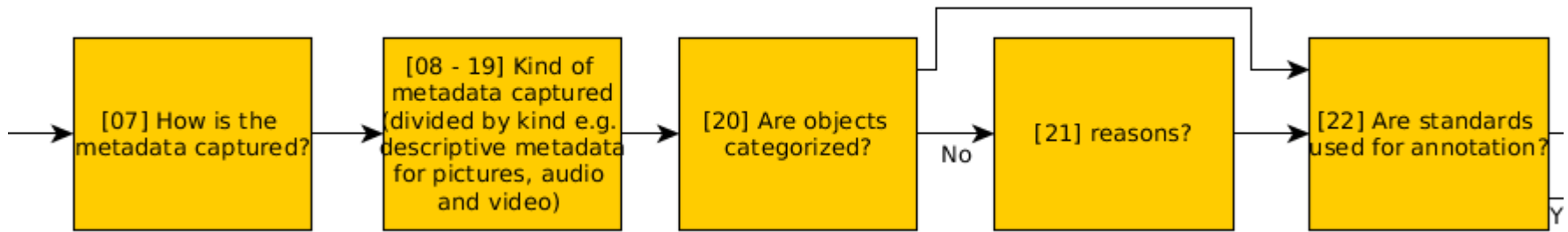
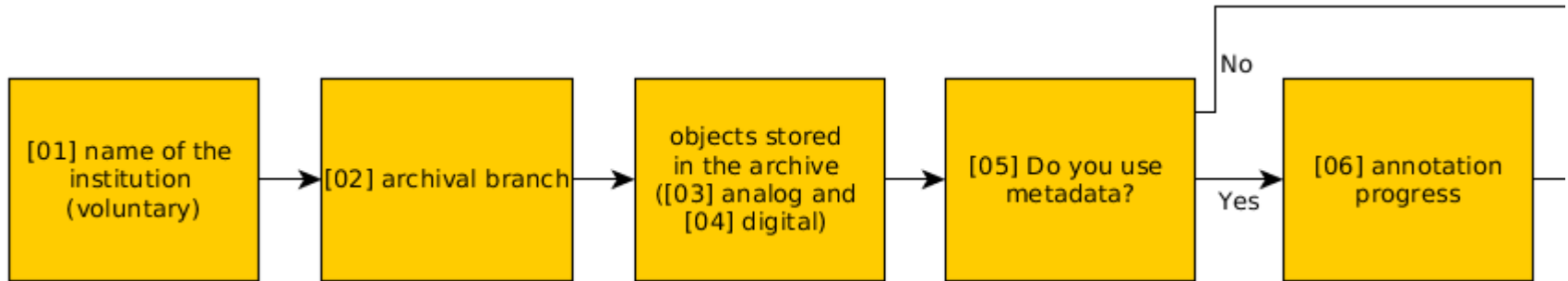
# Introduction

- Preliminary study
- Current state of metadata for audio-visual objects
- Information silo
- Well formed, standardized metadata needed
- Big archives provide many million on-line metadata objects, but smaller ones?

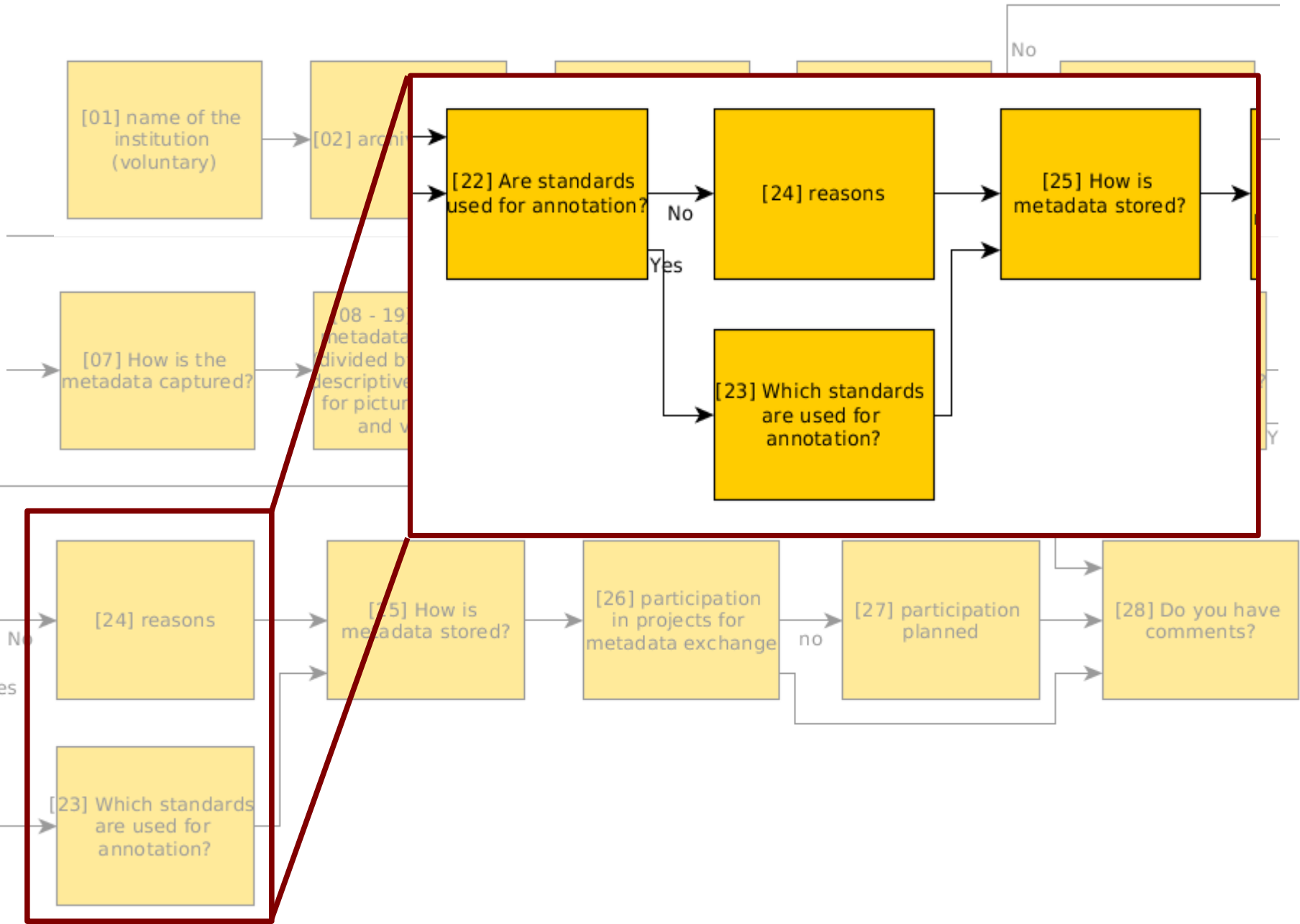
# Design of the survey

- Open not personalized on-line survey
- October – November 2010
- 1865 german institution were contacted via eMail
- 388 institutions take part (20.8 %)
- Focus on audio-visual objects

# Design of the survey

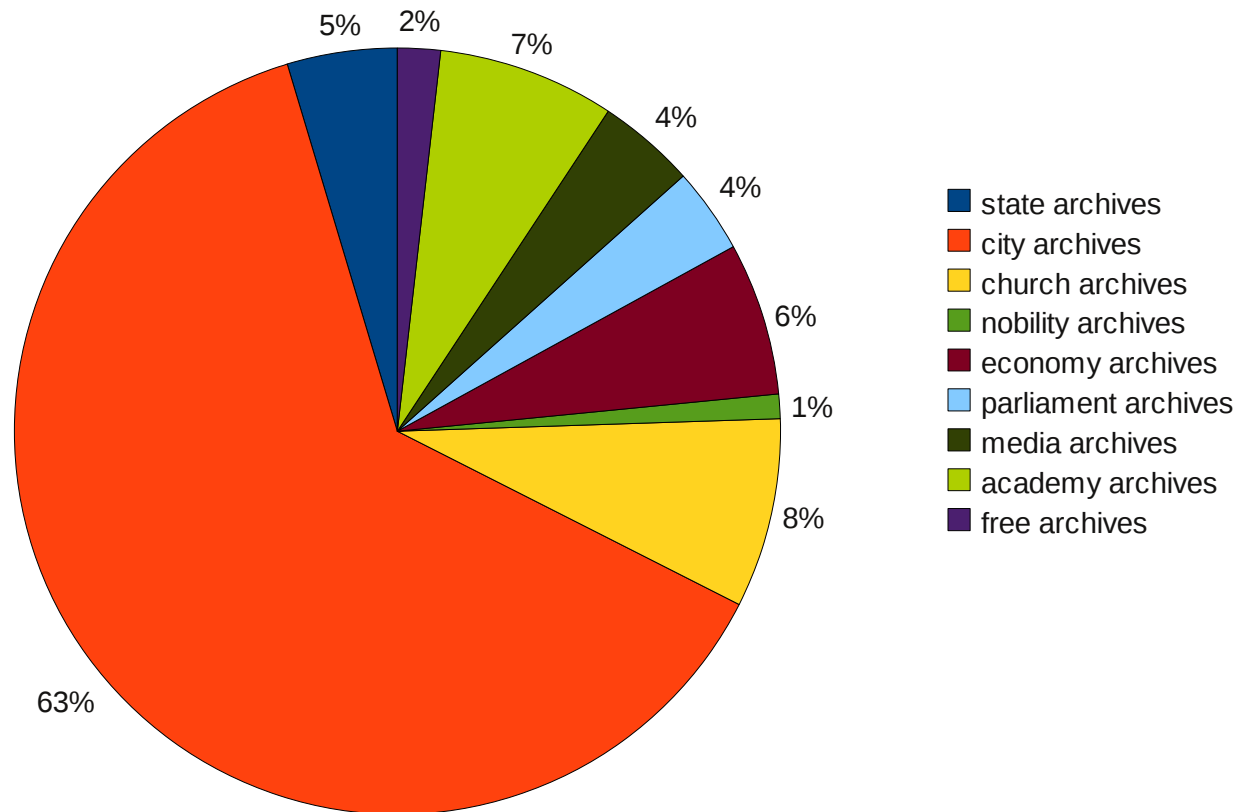


# Design of the survey



# Findings

Q2: Wich archival branch do your institution belong to? (% rounded)



# Findings

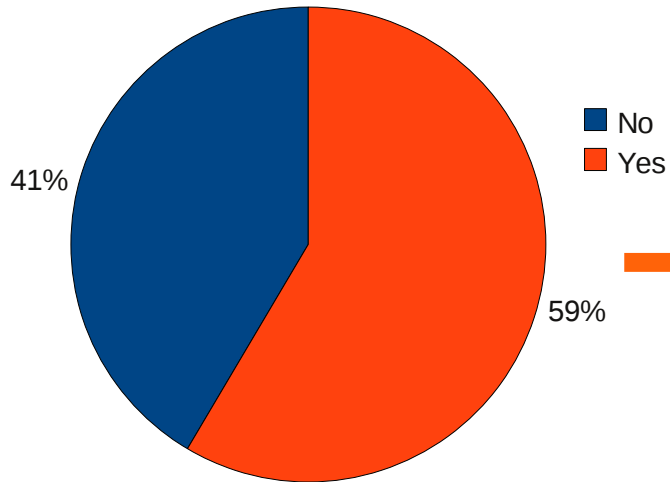
	0	1-10	11-100	101-1000	1001-10000	10001-100000	>100000
Analog pictures	4.12%	1.29%	3.35%	13.40%	37.89%	27.06%	12.88%
Analog audio	12.63%	18.04%	41.75%	21.91%	3.61%	1.55%	0.52%
Analog video	11.34%	18.04%	44.33%	19.85%	4.64%	1.29%	0.52%
Digital pictures	14.69%	3.61%	9.28%	22.16%	26.80%	16.75%	6.70%
Digital audio	22.94%	19.33%	35.57%	18.56%	1.80%	1.03%	0.78%
Digital video	21.13%	24.23%	35.57%	15.98%	2.58%	0.52%	0%

Q3/4: How many objects are stored in the archive?

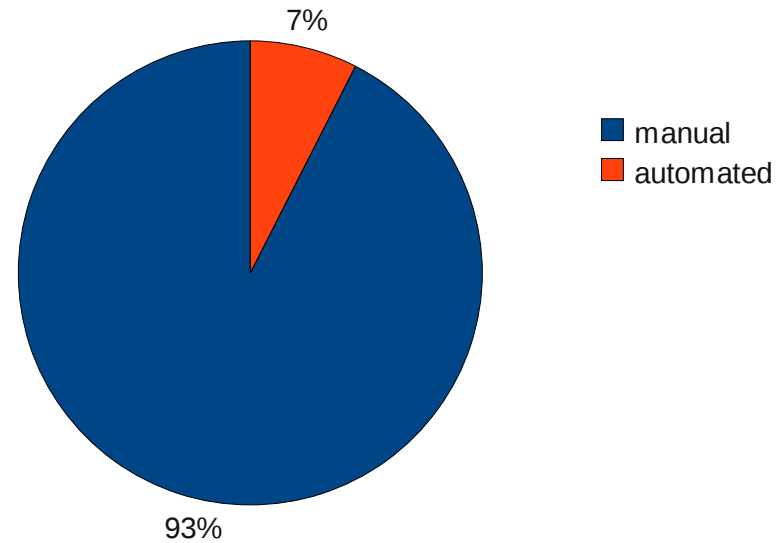


# Findings

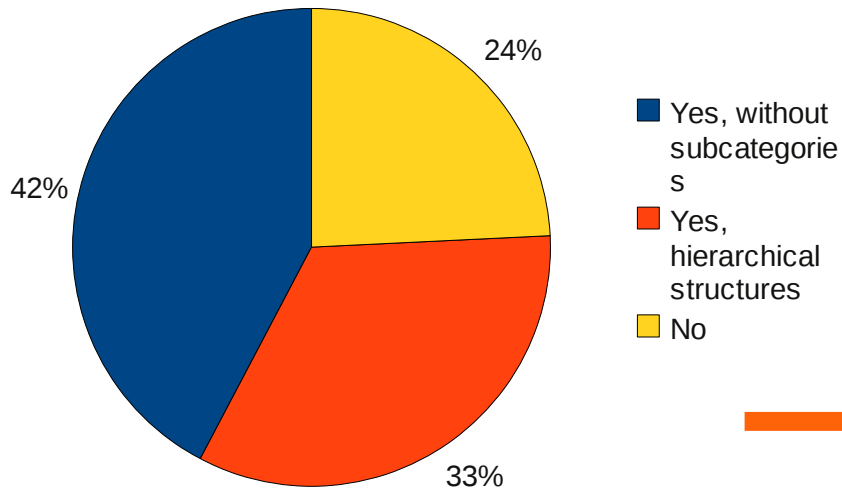
Q5: Is metadata captured?



Q7: How is metadata captured?



Q20: Are objects categorized?



If No, possible reasons:

- holding too small
- no staff
- no time
- no money

# Findings

<b>pictures</b>	<b>audio</b>	<b>video</b>
pictured person (200)	date (134)	date (144)
pictured object (200)	year of publication (127)	year of publication (137)
date (194)	title (119)	title (125)
photographer (186)	description of content (110)	duration (123)
place (183)	duration (105)	description of content (120)
none (4)	none (35)	none (31)

Q8: Which descriptive metadata do you collect?

(mentions)

# Findings

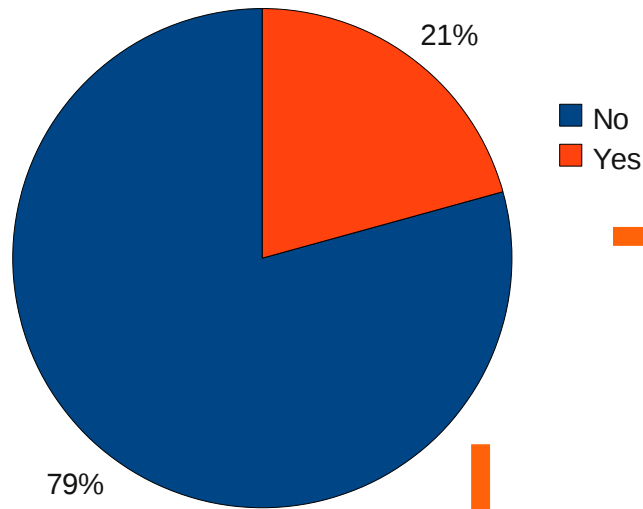
<b>pictures</b>	<b>audio</b>	<b>video</b>
signature (174)	signature (130)	signature (135)
creator (162)	provenience (105)	creator (119)
provenience (156)	creator (98)	provenience (112)
terms of use (82)	terms of use (58)	terms of use (71)
references (71)	retention period (37)	references (40)
none (15)	none (50)	none (43)

Q9: Which administrative metadata do you collect?

(mentions)

# Findings

Q22: Are standards used for annotation?



Which standards are used?

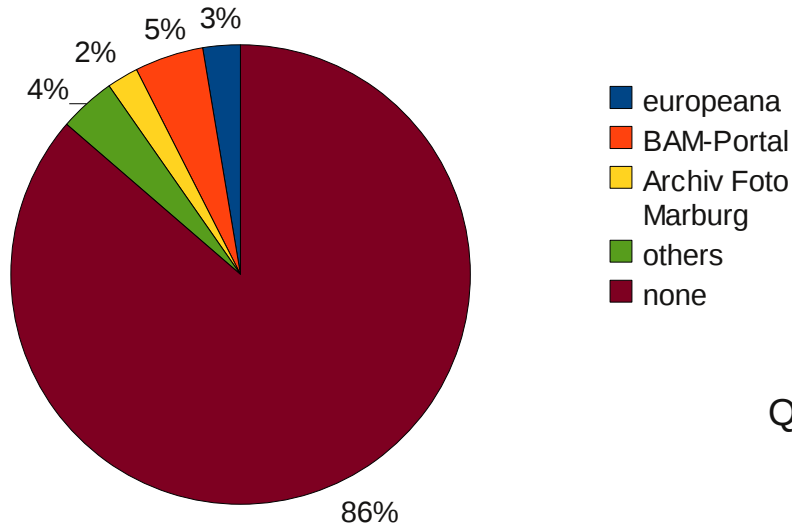
- IPTC (18, 4.64%)
- ISAD-G (16, 4.12%)
- EAD (12, 3.09%)
- EXIF (9, 2.32%)
- Dublin Core (8, 2.06%)
- MPEG7, MAB, MAB2, MARC21

If No, possible reasons:

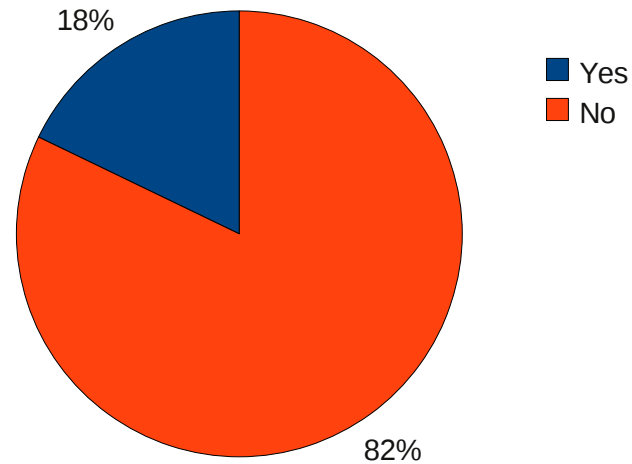
- holding too small
- no staff
- number of standards too big
- use of own regulations

# Findings

Q26: In which project (metadata exchange) do you participate?



Q27: Is a participation in projects (metadata exchange) planned?



# Conclusions

- Metadata issue plays minor role in most german archives (there are exceptions)
- Objects are not annotated close to international standards
- Increased training and advertising
- Broader use of pool resources e.g. for indexing
- But we need metadata to find and use the information stored within documents

Thank you for your attention.

@marcelruhl  
ruhl@fh-potsdam.de