

Do we need metadata? An on-line survey in german archives

Marcel Ruhl

University of Applied Science Potsdam, Faculty of Information Sciences,
Friedrich-Ebert-Str. 4, 14467 Potsdam, Germany

ruhl@fh-potsdam.de

<http://iw.fh-potsdam.de>

Abstract. The paper summarizes the results of an on-line survey which was executed 2010 in german archives of all branches. The survey focused on metadata and used metadata standards for the annotation of audio-visual media like pictures, audio and video files (analog and digital). The findings motivate the question whether archives are able to collaborate in projects like europeana if they do not use accepted standards for their orientation. Archives need more resources and archival staff need more training to execute more complex tasks in an digital and semantic surrounding.

Keywords: on-line survey, metadata, archives, audiovisual, semantic web

1 Introduction

Stefan Gradmann said in his inaugural lecture at the Humboldt-University Berlin that digital information objects should be understandable without reading them all [1]. This statement is gaining in weight. Through the power of the Internet more and more information are available, which are crawled with search algorithms. These generate a large number of hits. The resulting amount of information are often unmanageable for the seekers. The location and time independent web resources are accompanied by localized and time-based accessible information resources. This archival material is only accessible to users via analog finding aids. In the context of the semantic web both described kinds of information are information silos. The reason for this is the missing link between the individual stocks. To bring these silos together one requirement is well-formed, standard based metadata [2], [4]. This is also against the background of the development of portals like europeana relevant. For though german archives provide many million on-line metadata objects, it seems that the majority of archives is not deep in this topic. Reason enough to ask precisely what the situation is in the archives. Because it is mandatory for most portals like europeana that the provided metadata describes objects with digital representations like scanned documents [3], this paper focuses on multimedia objects.

The objective was to gain information about metadata for audio-visual objects in archives, how relevant is metadata and how is it used at the moment. The needs and deficiencies of german archives should be determined. To reach these goals an on-line survey was designed.

The paper summarizes the results of an on-line survey from the year 2010. German archives of all branches were asked about their use of metadata. The text is structured as follows. **Section 2 On-line Survey** describes design and execution of the survey. **Section 3 Interpretation and Discussion** summarizes the answers to selected questions. Tables and diagrams are used for visualization. The last **Section 4 Conclusion** lists the main findings and gives a forecast to possible conclusions to change the founded situation.

2 On-line Survey

The survey was conducted from 28.10.2010 until 12.11.2010. It was designed as an open on-line survey. The high number of potential participants spoke for this decision. A paper based survey couldn't be analyzed in appropriate time. Institutions were invited via email to complete the on-line questionnaire. The survey was created with LimeSurvey [5], an open source survey application.

The survey was not personalized, no login or password was needed. The implied problem of multiple attendance was solved by using the possibility of using cookies. So the questionnaire couldn't be completed from one workstation several times.

The participants had to answer 28 questions (one-choice, multiple-choice and free text), from which some were based on the answers of previous ones. Topically the questions covered the archived media (analog and digital pictures, audio and video), used metadata standards and the participation in projects for metadata exchange.

The institutions were chosen from a database provided by [6]. It contains 2733 datasets with addresses from german archives. Not all of them had an email-address, so 2056 institutions from all kinds of archives could be contacted. 191 email-addresses were invalid, so that 1865 Institution could have participate. Within the survey period an reminder email was send on 09.11.2010.

3 Interpretation and Discussion

From the above mentioned 1865 institutions 873 institutions attended the survey, but 485 participants stopped before the end and didn't finish the survey. Altogether 388 complete data sets were created and could be analyzed. The return rate was 46.81% and the drop-out rate was 55.56%, so 20.8% of the potential participants finished the survey.

After answering the first optional question about the name of the institution, the second question was about the archival branch. 18 (4.64%) state archives, 244 (62.89%) city archives, 31 (7.99%) church archives, 4 (1.03%) nobility archives,

25 (6.44%) economy archives, 14 (3.61%) parliament archives, 16 (4.12%) media archives, 29 (7.47%) academy archives and 7 (1.80%) free archives took part.

The following questions were about the media objects the institutions archive. The survey asked separate for analog and digital pictures, audio- and video-objects. See table 1 for details. The participated archives have a huge amount of analog pictures. 64.95% store between 1001 and 100000 objects. The quantity of analog audio- and video-objects is rather low. 41.75% (audio) respectively 44.33% (video) of the participants store between 11 and 100 analog objects. Most of the archives have a little amount of digital media. Especially the number of digital audio- and video-objects is mostly below 100 items. It seems that this kind of archival objects hasn't arrived in the archives yet or that existing analog objects are not broad digitized.

	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%

Table 1. How many objects are stored in the archive?

Up to this point, all participants saw the similar questions. The following question number five (Is metadata captured?) was designed to exclude institutions which do not capture metadata from the following block. Answering with NO the participants could not see the following questions. This design was chosen, because it was supposed that this institutions could not answer the questions concerning the metadata elements and used standards. This decision was possibly the reason for the high drop-out rate (see above), because the following sites of the survey were shown empty. The participants had to click several times to get to the last block with the comment field. 227 (58.51%) answered this question with YES and 161 (41.49%) answered with NO.

92.51% of the institutions which capture metadata for audiovisual objects do this manually, 17 participants (7.49%) capture the metadata automatically. After that the survey asked about the used metadata fields for descriptive, administrative, technical and structural metadata for all kinds of objects (analog and digital). The tables 2 and 3 show the fields mostly used. Technical metadata is captured depending on the kind of the object. For pictures the color (148) and the file-format (104), for audio the medium (99) and the material (71) and for video the color (111) and the film-format (95) are the most often used metadata fields. In all three categories many institutions answered, that they don't collect metadata. The exact numbers are for pictures (32), audio (58) and video (49).

The captured fields for structural metadata are the same in all three categories, only the number of mentions differs. Table 4 gives further information.

Pictures (mentions)	Audio (mentions)	Video (mentions)
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 contents (110)	duration (123)
place (183)	duration (105)	description of contents (120)
description of contents (148)	place (85)	original title (95)
year of publication (143)	keyword (66)	place (94)
title (138)	original title (66)	producer (77)
keyword (117)	speaker (53)	keyword (71)
country (41)	producer (51)	director (52)
genre (35)	original language (30)	actors (41)
language (10)	language (25)	film location (38)
remarks (3)	director (25)	language (31)
event (2)	genre (23)	genre (30)
others (3)	others (23)	others (75)
none (4)	none (35)	none (31)

Table 2. Which descriptive metadata do you collect?

Pictures (mentions)	Audio (mentions)	Video (mentions)
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)
retention period (53)	references (32)	retention period (29)
license (45)	availability (30)	availability (27)
others (80)	others (32)	others (49)
none (15)	none (50)	none (43)

Table 3. Which administrative metadata do you collect?

When asked about the use of metadata standards for the annotation of objects, 79.30% (180) answered not to use standards for their guidance. Just 47 institutions (20.70%) used standards like IPTC (19), ISAD-G (16) or EAD (12)¹(see table 5). Furthermore, in this context, the question was asked, what

¹ Multiple-choice was possible.

	Pictures	Audio	Video
Holding	164	118	118
Series	88	61	59
Sequence	37	39	37
Classification	5	2	2
None	36	66	69

Table 4. Which structural metadata do you collect?

reasons were there for non-use of standards. Mentioned reasons were lack of resources, higher priority for classical records and lack of information on standards. In the comment field participants mentioned additionally the annotation before the introduction of appropriate standards, the use of its own regulations and the low holding size.

standard	mentions	standard	mentions
IPTC	19	FIAF cataloguing rules	2
ISAD-G	16	RNA	2
EAD	12	EAC-CPF	1
EXIF	9	METS	1
Dublin Core	8	MIDAS	1
RAK (NBM)	7	PND	1
Regelwerk ARD-ZDF	3	GKD	1
in-house guidelines	3	SWD	1
MAB	2	XMP	1
MAB2	2	others	5
MPEG7	2		

Table 5. Which standards are used for annotation?

An major intention of the survey was to find out if archives are participating in projects which focus on the exchange of metadata. The survey shows that 196 institutions are not attending in such projects. 11 institutions provide metadata in the BAM-portal [8] and 6 in europeana [7]. For details see figure 1. The participants which do not participate in projects yet, were asked if it is planned in the future. Here 161 (82.14%) participants answered with NO. Just 35 (17.86%) said YES (see figure 2). Against the background of the intentions of metadata exchange this is an very bad outcome. At this point the answers differ depending on the archival branch. Most of the smaller archives like city archives are not planning the participation in portals. Reasons are the same mentioned above. Institutions with better resources like state archives can afford the commitment of human and financial costs easier.

At the end of the survey respondents had the option to enter their notes and comments into a comment field. Here a large number of participants pointed out

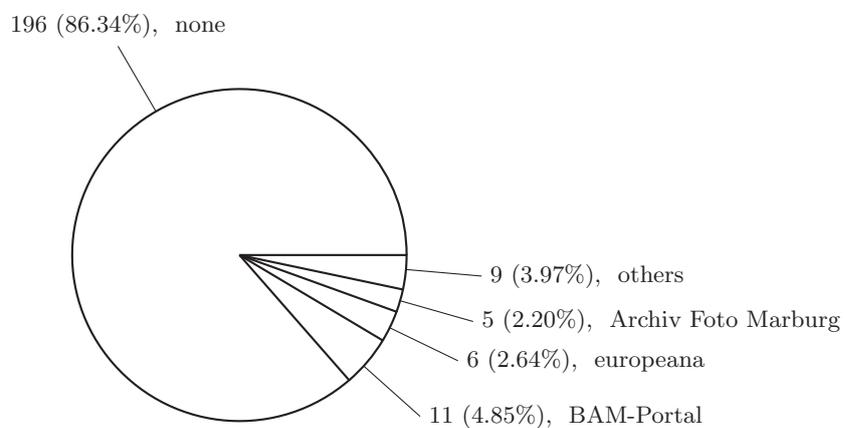


Fig. 1. In which project (metadata exchange) are you participating?

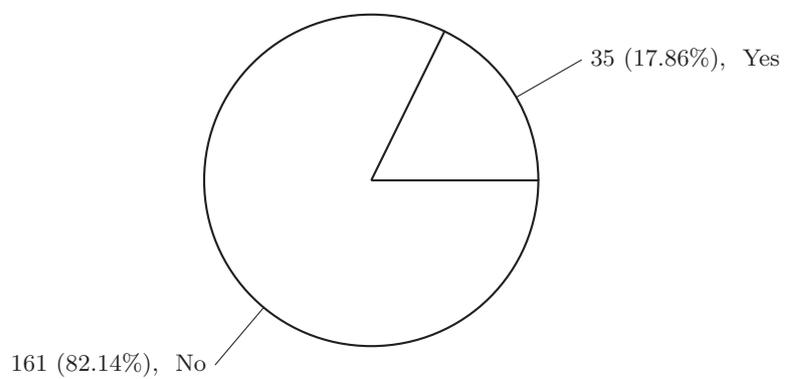


Fig. 2. Is a participation in projects (metadata exchange) planned?

that the archives have to deal with a lack of human and financial resources and therefore no opportunities for the annotation of audiovisual media are present. Archives focus on analog records at the moment. Another frequently mentioned issue is the ambiguity of terms such as *metadata* and *audiovisual*. Here training and understandable guidelines on this topic are demanded.

4 Conclusion

The paper summarized the results of an on-line survey which was executed 2010 in german archives of all branches. The survey found that the issue of metadata for audiovisual objects, metadata standards and their exchange plays a minor role for most of the german archives. Though archives are professionals in making classical records accessible, some archival branches like nobility archives or small city archives somehow can not use this professionalism for archival material with a technical smell. The question has to be asked if, under this results, german archives could be interested in semantic web technologies if they have not the ability to annotate their objects close to accepted international standards and are not planning to share their metadata. This can only be attributed to a lack of knowledge of the subject and its benefits to the archival landscape or to an extreme lack of personnel, temporal and financial resources in the archives. This shortage could be relieved by increased training and advertising of this issue. An other possible solution could be the wider use of pool resources for archival issues like indexing.

References

1. Gradmann, St.: Signal. Information. Zeichen. Zu den Bedingungen des Verstehens in semantischen Netzen. (2008), <http://edoc.hu-berlin.de/humboldt-v1/157/gradmann-stefan-3/PDF/gradmann.pdf>
2. Hitzler, P., Krötzsch, M., Rudolph, S., Sure, Y.: Semantic Web - Grundlagen. Springer, Berlin (2008)
3. europeana Aggregators Handbook (2011), <http://pro.europeana.eu/documents/858566/858665/Aggregators+Handbook>
4. Heath, T., Bizer, Ch.: Linked Data: Evolving the Web into a Global Data Space. Morgan & Claypool, San Rafael (2011), <http://linkeddatabook.com/editions/1.0/>
5. LimeSurvey - the open source survey application, <http://www.limesurvey.org/>
6. Verband deutscher Archivarinnen u. Archivare: Archive in Deutschland, Österreich und der Schweiz. Ein Adressverzeichnis. 20. aktual. Aufl., Ardey, München (2009)
7. europeana.eu, <http://europeana.eu/portal/>
8. Gemeinsames Portal für Bibliotheken, Archive und Museen - ein Online-Informationssystem, <http://www.bam-portal.de/>