



www.axes-project.eu

Project acronym	AXES
Project full title	Access to Audiovisual Archives
Project No	269980
	Large-scale integrating project (IP)

Deliverable D1.8

Final Use Case Evaluation Report

PROJECT DELIVERABLE REPORT

Project

Grant Agreement number	269980
Project acronym:	AXES
Project title:	<i>Access to Audiovisual Archives</i>
Funding Scheme:	<i>Large-Scale Integrating project (IP)</i>
Date of latest version of Annex I against which the assessment will be made:	24 September 2010

Document

Deliverable number:	D1.8
Deliverable title	Final Use Case Evaluation Report
Contractual Date of Delivery:	31/12/2014
Actual Date of Delivery:	19/12/2014
Author (s):	Martijn Kleppe, Jill Briggeman
Contributors	Andy O'Dwyer, Max Kemman, Henri Beunders, Peggy van der Kreeft, Roeland Ordelmans, Denise Bland
Reviewer (s):	Kevin McGuinness
Work package no.:	1
Work package title:	Data distribution and end user perspective
Work package leader:	Henri Beunders (EUR)
Version/Revision:	7
Draft/Final:	Final
Total number of pages (including cover):	70

CHANGE LOG

Reason for change	Issue	Revision	Date
Initial version by EUR	1	0.1	17-11-2014
Chapter 2 added by Jill Briggeman	2	1.0	25-11-2014
Appendix B, C & D added by Jill Briggeman	3	1.0	27-11-2014
Text revised by Martijn Kleppe	4	1.0	28-11-2014
Chapter 3 added and revised by Martijn Kleppe, Roeland Ordelmans, Andy O'Dwyer, Denise Bland, Henri Beunders and Peggy van der Kreeft	5	2.0	05-12-2014
Final editing and lay-out improvements	6	3.0	11-12-2014
Review suggestions by Kevin McGuinness implemented	7	4.0	12-12-2014

DISCLAIMER

This document contains description of the AXES project work and findings.

The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any responsibility for actions that might occur as a result of using its content.

This publication has been produced with the assistance of the European Union. The content of this publication is the sole responsibility of the AXES project and can in no way be taken to reflect the views of the European Union.

The European Union is established in accordance with the Treaty on European Union (Maastricht). There are currently 27 Member States of the Union. It is based on the European Communities and the member states cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice and the Court of Auditors. (<http://europa.eu.int/>)



AXES is a project partly funded by the European Union.

TABLE OF CONTENTS

Change log.....	3
Disclaimer.....	4
Table of contents.....	5
Summary	7
1 Introduction.....	8
2 User study AXES RESEARCH	10
2.1 Introduction.....	10
2.2 Experimental set-up and method.....	11
2.2.1 Pilot.....	11
2.2.2 User study	11
2.2.3 Surveys	15
2.2.4 Comments and ecological validity	15
2.3 Participants	16
2.4 Results	17
2.4.1 General comparison of the information retrieval systems employed.....	17
2.4.2 Q1: How is browsing behaviour of SERPs affected by a lack of annotations?	17
2.4.3 Q2: How is relevance of AV search results assessed?.....	18
2.4.4 Q3: To what extent do humanities scholars consider a lack of annotations an issue?	20
2.4.5 Context and background	23
2.5 Conclusions	24
3 Evaluation AXES HOME	27
3.1 Defining the home user.....	27
3.2 AXES Home Digital Library System	28
3.3 Goal of the evaluation.....	29
3.4 Set-up of the evaluation	29
3.5 Results: Acceptance of the AXES HOME system.....	32
3.5.1 Performance.....	32
3.5.2 Navigation and functionality	33
3.5.3 Discovery of tools and understanding of what they offer.....	35
3.5.4 Impediments	39
3.5.5 Requested features	40
3.5.6 AXES HOME as entertainment tool.....	40

3.6	Conclusion: Home user access models to audio-visual archives.....	41
4	References	43
Appendix A	Surveys user study AXES RESEARCH.....	44
Appendix B	Graphs user study AXES RESEARCH	52
Appendix C	Improvements AXES HOME.....	56
Appendix D	Workflow AXES HOME evaluation	57
Appendix E	Survey AXES HOME evaluation	58
Appendix F	Results AXES HOME evaluation.....	63
Appendix G	Promotional Material AXES HOME Evaluation.....	70

SUMMARY

This report describes two evaluations that were undertaken by the BBC, Deutsche Welle, Netherlands Institute for Sound and Vision and the Erasmus University Rotterdam. First, a study was performed with the AXES RESEARCH system, focusing on the influence of a lack of associated archivists' annotations on the relevance judgments of humanities researchers. The second study evaluated the AXES HOME system and described the lessons learned when developing a system for home users that aimed to give them pathways into an archive containing (historical) audio-visual content.

1 INTRODUCTION

As the AXES project comes to an end, the focus of Work package 1 in the fourth year has been on finalizing the user study with AXES RESEARCH and the evaluation of the AXES HOME system. This deliverable contains the final results of the user evaluations as performed by the BBC, Deutsche Welle, Netherlands Institute for Sound and Vision and the Erasmus University. This deliverable reports on two main evaluation sections:

The Erasmus University Rotterdam performed an additional user study on the AXES RESEARCH system with the objective to help archivists, librarians and ICT-researchers and developers gain more insight into the various ways ***humanities scholars navigate and seek in audio-visual archives, and their ability to make relevance judgments when presented with poor archivist annotations***. One of the hypotheses formulated by the researchers of AXES states that video archives can be unlocked using image search and speech recognition. If true, this would imply that it is useful to digitize thousands of hours of undescribed and unused audio-visual material currently being stored in archives, notwithstanding the absence of manually entered metadata. In this user study, a crucial part of the aforementioned assumption was put to the test amongst 27 humanities researchers. The results are reported in chapter 2.

The BBC led the user evaluations of AXES HOME at their main office in London and were assisted by partners of Deutsche Welle, the Netherlands Institute for Sound and Vision and the Erasmus University Rotterdam. This evaluation had two aims: ***1) the evaluation of the overall acceptance of the AXES HOME system, and 2) an assessment of potentially interesting models for users to access audio-visual archives***. A total of 25 users participated in these evaluations sessions. The results are reported in chapter 3.

This deliverable is the final report of Work package 1. In our previous deliverables we reported on the data delivery (D1.1, D1.5 & D1.9), requirements studies (D1.2, D1.3, D1.6) and the evaluations of AXES PRO (D1.4) and AXES RESEARCH (D1.7). The AXES project aimed to have as many users involved as possible in the setting up the systems as well as evaluating them. With a total of 1.851 users participating in one or more of the user studies, we consider this goal to be achieved.

System		Study	Number of participants	Reported in deliverable
AXES PRO	Requirements	TRECVID benchmark session	14	D1.2
		Observations	9	D1.2
		Group elicitation	21	D1.2
		Concept mapping study	47	D1.6
	Evaluation	Evaluation study	5	D1.4
AXES RESEARCH	Requirements	Group interview	13	D1.2
		Interviews	5	D1.2
		Survey academic researchers	342	D1.6
		Eyetracking study	24	D1.6
		Survey journalists	231	D1.6
	Evaluation	Evaluation study	48	D1.7
		VideoHyperlinking Evaluation	30	D1.7
AXES HOME	Requirements	Interviews	7	D1.2
		Survey	970	D1.2
		MediaEval workshops	60	D1.6
	Evaluation	Evaluation study	25	D1.8
TOTAL			1.851	

Table 1: Participants per session

2 USER STUDY AXES RESEARCH

In D1.7, we described the results of the first user evaluations of the AXES RESEARCH system. In this chapter we present an additional user study that aimed to help archivists, librarians and ICT-researchers and developers to gain more insight into the various ways humanities scholars navigate and seek in audio-visual archives, and their ability to make relevance judgments when presented with poor archivist annotations.

2.1 Introduction

One of the hypotheses formulated by the researchers of AXES states that video archives can be opened up using image search and speech recognition (AXES project proposal 13-04-2010). If true, this would imply that it is useful to digitize thousands of hours of undescribed and unused audio-visual material currently being stored in archives, notwithstanding the absence of manually entered metadata. The AXES system enables users to search specific faces, events, logos and locations, without them being mentioned in the associated archivists' annotations (Van der Kreeft, Macquarrie, Kemman, Kleppe and McGuinness, 2014).

In this additional user study on the AXES RESEARCH system, we have put a crucial part of the aforementioned assumption to the test among 27 humanities researchers. Using image and speech recognition (Content-Based Image Retrieval, hereinafter referred to as CBIR), unannotated video archives are searchable by means of automatic linking to 'known entities'. However, although CBIR might offer the technical potential to generate search results without annotations, are scholars able to work with these? This issue led to our main research question:

- ***How does a lack of accompanying archivists' annotations influence the ability of humanities researchers to make relevance judgments about audio-visual documents?***

The objective of this study was to help archivists, librarians and ICT-researchers and developers to gain more insight into the various ways humanities scholars navigate and seek in audio-visual archives, and their ability to make relevance judgments when presented with poor archivists' annotations. Our operationalized sub-questions were:

- ***Q1: How is browsing behavior of SERPs (Search Engine Results Page) affected by a lack of annotations?***
- ***Q2: How is relevance of AV search results assessed?***

➤ **Q3: To what extent do scholars consider a lack of annotations an issue?**

To address these questions, mostly qualitative research methods were employed. In making their various relevance judgments, we expected that humanities researchers – ‘expert users of text’ (Duff and Johnson, 2002) and ‘intellectually able seekers but not technical in orientation’ (Buchanan, Cunningham, Blandford and Rimmer, 2005) – are still very dependent on the archivists’ annotations accompanying retrieved video results. We conducted our experiment using the AXES RESEARCH prototype, with functionalities and an interface designed specifically for a user group consisting of journalists and academic researchers (D7.2 and D7.5). The technical aspect of the premise, that it is possible to unlock audio-visual archives using CBIR, is beyond the scope of this experiment.

2.2 Experimental set-up and method

In our experimental set-up, we followed similar approaches of relevant studies by amongst others Choi and Rasmussen (2002), Yang and Marchionini (2004), Van den Heuvel (2010), Schamber (2000) and Chung and Yoon (2012).

2.2.1 Pilot

Our research plan consisted of two components: a pilot experiment (two participants) and a user study (27 participants). In the pilot phase, initial feedback was recorded, and consequently issues concerning the duration of the search tasks and the clarity of the questions were solved before the start of the actual user study.

2.2.2 User study

After a short introduction, the scholars were given two different search tasks and queries, performing **both a text-based/spoken words search** (meaning that the query matched particular words found in the archivists’ annotations) and **a content-based search** (meaning that the query matched visual aspects of the video’s content, facial features in particular) of ten minutes each.

Following methodical guidelines as set forth in i.a. Van Someren, Barnard and Sandberg (1994), the scholars were asked to verbalize their thought processes (‘think-aloud protocol’, respectively during and after each task) and answer multiple choice and open questions (e.g. about the relevance of the retrieved search results). The comments by the participants as well as the observations made by the researcher present during the experiment were transcribed from screen and audio recordings. At a later stage, these transcripts were carefully reviewed, where we identified which

homogeneous content categories could be distinguished in the responses and behavior of the participants. In accordance with methodical guidelines proposed by Gibbs (2008), all similar comments and interactions per task were assigned a unique code, tallied and analyzed.

The participants in the final experiment were divided into groups of approximately equal size, group A and group B. Group A started the experiment with a text-based search, group B with a CBIR search. For the second task, their search methods were reversed:

Query / Results page	Detailed view (with metadata)	Thumbnail view (without metadata)
<i>Text-based search</i>	Group A, Task 1	Group B, Task 2
<i>CBIR search</i>	Group B, Task 1	Group A, Task 2

Table 2: Schematic overview of the experimental set-up.

In the first task, both user groups were shown their search results in **'detailed view'**: a detailed display option showing a thumbnail along with an overview of the archivists' annotations (see fig. 1):

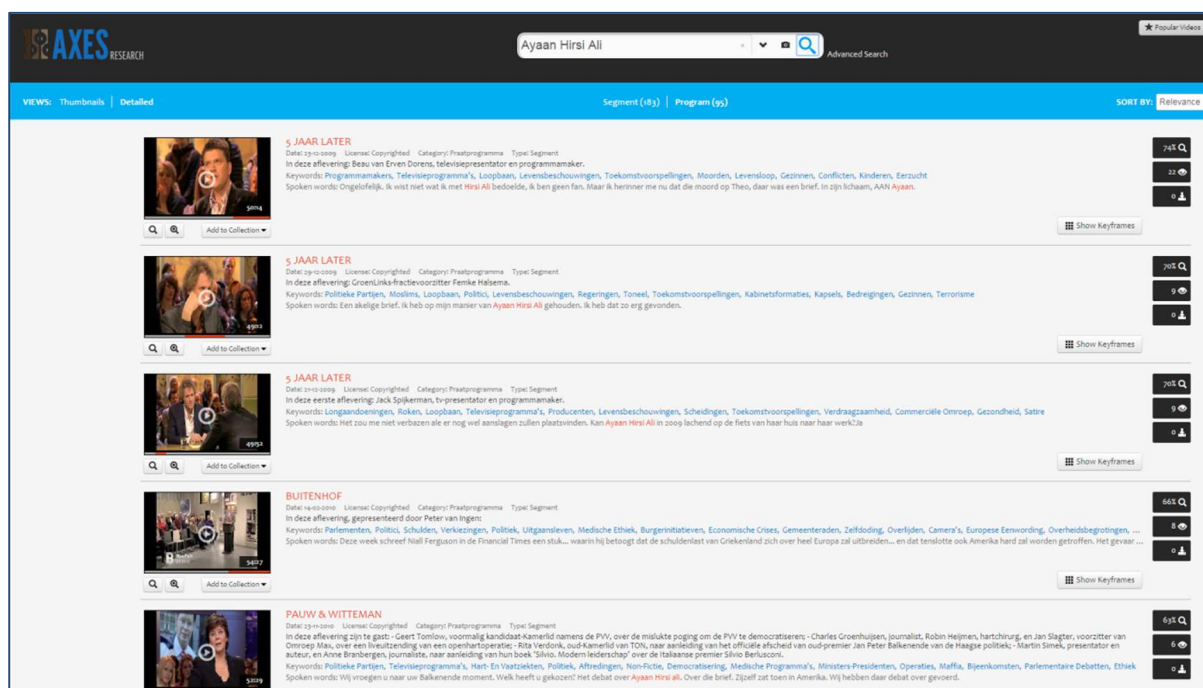
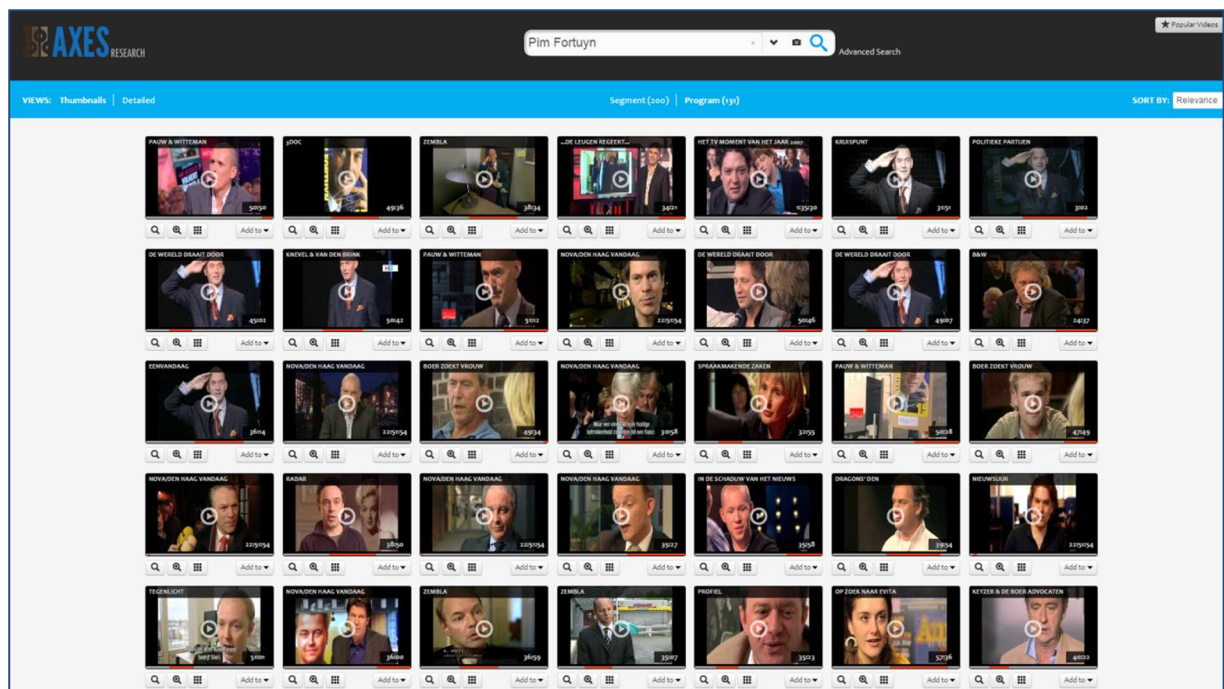


Figure 1: AXES RESEARCH - system setting 'detailed view'.

In the second task, the search results were displayed in **'thumbnail view'**, showing only a still frame per video accompanied by a program title (fig. 2):



The purpose of the categorization into two different user groups was to determine whether any discernible effects could be detected in relevance judgments ***when the queries were made to match the presentation of search results***. To examine this, the queries group A used for their tasks were optimally compatible with the display of the search results, whilst those of group B were not. This meant specifically that group A first performed a text search after which the result list was textually displayed as well (with parts of the associated metadata, system setting 'detailed view'), and secondly a visual search after which the result list was also visually displayed (with large still frames from the videos, setting 'thumbnail view'). The feedback of the two user groups was later ***compared within the group*** (text search task vs. visual search task per group) and ***between groups*** (group A vs. group B per task), coded, and analyzed.

In order to optimize the comparison within-groups as well as between-groups we chose two similar queries about well-known Dutch politicians. Consequently, they were asked twice to find as many relevant videos of opinions about these different controversial Dutch politicians – Pim Fortuyn and Ayaan Hirsi Ali – as possible within a time frame of ten minutes.

Research question:	How have attitudes in the media and in politics changed towards Ayaan Hirsi Ali's political thoughts	How have attitudes in the media and in politics changed towards Pim Fortuyn's political thoughts since his death in 2002?
---------------------------	--	---

	since her departure from the Dutch House of Representatives in 2006?	
Search task:	Evaluate and 'favorite' as many relevant clips as possible in 10 minutes, in which opinions are expressed about Ayaan Hirsi Ali's political thoughts.	Evaluate and 'favorite' as many relevant clips as possible in 10 minutes, in which opinions are expressed about Pim Fortuyn's political thoughts.
Query:	"Ayaan Hirsi Ali"	"Pim Fortuyn"
Search settings:	Metadata search	Visual search
First 50 search results shown:	±8 relevant, ±7 partially relevant	±6 relevant, ±7 partially relevant

Table 3: overview of the search tasks.

Queries were deliberately formulated to be ambiguous, meaning that the search results – both in thumbnail and detailed view – did not immediately give away any hints on whether the video was relevant to the given task. The intention was to force users to carefully examine and consider the search results and to exclude any possible 'gambling'. When the scholar deemed a particular search result useful for the query, the clip was to be added to their list of 'favorites' by clicking on the heart symbol. Thus, in our study, ***a favorite equaled a clip judged as relevant***. In order to prevent the task from being needlessly complicated, we made sure every query resulted in at least 13 (partly) relevant clips.

In the ***thumbnail view task***, users naturally had no other option than to click and watch clips to determine their relevance (apart from reading the program title). In the ***detailed view task*** however, scholars were given substantially more choices: participants could decide to view the clip, quick scan the parts of the associated metadata as displayed in the result list, or to click the expand option and read the extensive archivists' annotations accompanying each clip. Since the tasks had to be performed within a certain time frame, we initiated a program setting in AXES RESEARCH that enables clips to start at an automatically preselected point. Firstly, this setting provided participants who either chose to or had to watch the fragments with the technical means to work through their search results quickly: instead of searching through entire television broadcasts, only specific fragments had to be scanned. Secondly, earlier studies determined that researchers in general prefer to work with audio-visual materials at fragment level (Huurnink, Hollink, Van den Heuvel and De Rijke, 2010).

2.2.3 Surveys

After each task, participants were asked to fill out a survey consisting of both multiple choice and open questions. At the conclusion of the experiment, a more general exit questionnaire was administered (for both surveys, see appendix A). It should be taken into account that the (additional) open questions were optional and therefore not always filled out.

The survey was performed by using an online data collection tool called Qualtrics. The results of the survey, as well as the observations made by the researcher present, were then exported for analysis to SPSS (Statistical Package for the Social Sciences) and Excel. All questions were specifically formulated with the purpose of gaining more insight into the search behaviors and experiences of the participants.

Questions task surveys:

- A general estimation of the retrieved search results.
- Relevance criteria related to the evaluation of the video clips.
- The usefulness of the associated archivists' annotations (only for the task with display setting 'detailed view').
- The relative importance of the content of the video clips for making relevance judgments.
- An assessment of the level of difficulty for each task.
- A comparison of the level of difficulty between both tasks.

Questions exit survey:

- Demographics.
- The ways in which the participants generally use audio-visual material for their research and teaching.
- The general functionality of AXES RESEARCH (would they consider using this search system for their own research?).

2.2.4 Comments and ecological validity

To ensure a high ecological validity, both queries were related to a broader, more realistic research question (see table 3). The ecological validity of an experiment relates foremost to the degree to which the context, search tasks, and setting of the experimental study approximates actual research (Schmuckler, 2001). We ensured subjects, queries, research questions, and source materials were similar in nature with real ones used in humanities research.

Regarding the results of our study, two important factors should be taken into account. First, AXES RESEARCH is still a prototype under construction. Second, for this experiment, a beta version of the AXES RESEARCH system was used in which a limited data set was made available for testing purposes. The dataset included 3092 hours of

material in 4520 video files from the archives of the Netherlands Institute for Sound and Vision. All documents are specifically unlocked and annotated for use in education and science. Therefore, the choices of tasks and queries was necessarily directed and limited by the available dataset. However, great care was taken to create two equivalent assignments for the participants, e.g. similar research questions, similar queries, and a similar amount of relevant and irrelevant results.

2.3 Participants

For the purposes of AXES, all participants in this user study can be defined as humanities researchers. Before scholars were approached, they were all carefully preselected on their specific academic backgrounds. For the final experiment, a geographically heterogeneous group of 27 scholars was enlisted, focusing on historical, media and/or communications research in which videos are frequently used as (source) materials. 23 out of 27 participants reported to make use of audio-visual materials in their research and/or teaching often or very often. To further ensure unimpeded test results, we ascertained that all users were also well acquainted with both Dutch politicians and national television programs. Finally, all participants reported prior experience with audio-visual databases.

The scholars were divided into two as uniformly as possible user groups, according to their age, gender, and academic level of seniority. Of the total user group, 14 participants were female, and 13 were male; 13 scholars were classified into user group A, and 14 scholars into user group B.

User group A	User group B
7 female	7 female
6 male	7 male

Table 4: Gender distribution of the participants.

Age	Frequency
25-34	11
35-44	7

45-59	7
60+	2

Table 5: Age distribution of the participants.

2.4 Results

2.4.1 General comparison of the information retrieval systems employed

First, we determined whether the participants were able to complete simple research tasks using CBIR by comparing the number of videos seen and saved (added to 'favorites') between the traditional text-based search and the visual search task. This comparison purely involved the information retrieval systems used, meaning that the results of user group A's first task were contrasted with those of user group B's first task, and likewise, the results of user group A's second task were compared with those of user group B's second task (see table 2). The variable 'presentation of the search results' (whether accompanying metadata was shown or not) was left out of the equation and will be extensively addressed in the next paragraphs.

The results show that the humanities scholars generally completed both their text-based and CBIR task (see result tables 1 and 2). Out of 27 test subjects in total, only 3 participants were not able to select any relevant video clips at all on one or both tasks. There were no discernable differences between the results of the two information retrieval systems used: the participants achieved similar scores regarding videos seen and saved in the text-based search as the CBIR search.

2.4.2 Q1: How is the browsing behaviour of SERPs affected by a lack of annotations?

To address the first sub-question, we analyzed the influence of the search results' presentation and compared the number of videos seen and saved by the humanities researchers between the two given tasks. Based on our observations, we were able to detect no dissimilarities between the results of the two tasks: the participants were able to complete both the task where the program information was displayed ('detailed view') and the task where no archivists' annotations were shown ('thumbnail view'). There was also no significant difference in the rate at which they obtained relevant fragments. The tables below show the averages of videos seen (clicked on and watched) and saved (added to their 'favorites', i.e. deemed relevant):

	Task 1 (detailed view)	Task 2 (thumbnail view)	Total		Task 1 (detailed view)	Task 2 (thumbnail view)	Total
--	-------------------------------------	--------------------------------------	--------------	--	-------------------------------------	--------------------------------------	--------------

Group A	5	11,1	8,0	Group A	3,5	4,0	3,8
Group B	5,5	12,9	9,2	Group B	2,9	12,9	3,8
Total	5,3	12,0	8,6	Total	3,2	12,0	8,6

Result table 1: Videos watched (mean) by participants per task.

Result table 2: Videos saved (mean) by participants per task.

The participants managed to add at least as many or more videos to their favorites list with (mean: 3,2 clips) as without (mean: 4,4 clips) the archivists' annotations displayed with the search results. We also did not find any significant differences between user group A and user group B, **meaning that for the scholars it was of no importance whether the queries matched the presentation of search results**. I.e. for the evaluation of the search results list in both tasks, the query did not necessarily have to be compatible with the way in which the search results are displayed.

Another observation is that **the number of videos clicked on and watched was more than twice as large for the second task where no metadata was displayed**. Whilst performing the task with program information, the participants watched an average of 5,3 videos in ten minutes; without any metadata to go by, users looked at an average of 12 clips within the same time frame. Interestingly enough, prior to the start of the visual search task some of the scholars had informed the researcher present about their expectations of the experiment: these participants feared that watching video clips instead of scanning the associated program information like they were used to would be too time consuming and probably be very difficult.

Results Q1
How is browsing behaviour of SERPs affected by a lack of annotations?
<ul style="list-style-type: none"> ➤ The participants added at least as many or more videos to their favorites list with ('detailed view') as without ('thumbnail view') the archivists' annotations displayed with the search results. ➤ For the evaluation of the search results list in both tasks, the query did not necessarily have to be compatible with the display of the search results. ➤ The number of videos clicked on and watched was more than twice as large for the second task where no metadata was displayed.

2.4.3 Q2: How is relevance of AV search results assessed?

Choi and Rasmussen (2002) have composed a clear definition of relevance criteria: 'Relevance is viewed as a concept that expresses a value judgment about the quality of a relationship between information and information need [...] at a certain time in an

information-seeking and use situation'. For our study, depending on the given task, we focused in particular on practical relevance criteria such as the various components of the associated program information, interface and the content of the clips seen.

After performing the task in detailed view, 25 participants judged the overall search results as somewhat to very relevant. After the second task in thumbnail view 24 scholars deemed their retrieved list somewhat to very relevant. However, out of 27 participants, 10 pointed out that – were it actual research – they would at a later date have reassessed the usefulness and relevance of their list of 'favorite' videos. These remarks are in line with the findings of studies by i.a. Tang and Solomon (1998) and Wang and White (1999), who each observed that relevance criteria change during the different stages of the information retrieval process (Choi and Rasmussen, 2002).

Whilst performing the first task in detailed view, ***all participants made use of the textual program information displayed***. 22 scholars indicated that they had quickly scanned the associated metadata, and 5 participants even expressed to have read the text carefully. In order to obtain more information, 11 users clicked on the extended archivist annotations during the detailed view task, which opened in a separate screen. Also during the detailed view task, 9 people added clips to their favorites list unseen one or more times, based solely on the program information. In contrast, during the second task only 2 scholars added fragments unseen (based exclusively on the thumbnail picture). Generally, however, a (small) part of the fragment was watched before it was added to a collection of relevant findings.

For the thumbnail as well as the detailed view task, the importance and usefulness was tested of five general aspects in relation to the relevance evaluation of audio-visual material: the visual appeal of a clip, the thumbnail displayed (still frame from the clip), the ranking of the clip in the results list, the content of the clip and the accompanying program information (in thumbnail view this annotation consisted only of a program title). The importance of key frames was not taken into account due to slowness of the AXES RESEARCH prototype on some locations. Participants were asked to rate the usefulness of the aforementioned aspects on a 5-point Likert scale (see appendix B, figure 20 and 21). As it was possible in the survey to tick multiple criteria as e.g. 'important', we also requested from users to create a ranking, or in other words, to make a top 3 list of the given aspects (see appendix B, figure 22).

Results show that for the task in detailed view, users judged ***the accompanying program information*** as the most important criterion for evaluating the relevance of a particular clip. In second place on the scale of importance came ***viewing the content***

of the clip. For the thumbnail view test, **watching the content of the clip** was judged as the most important aspect. Note that the displayed metadata in the latter experiment consisted only of a program title. **The thumbnail picture** was placed in second position.

For the thumbnail as well as the detailed view task, the importance and usefulness was tested of four general aspects in relation to **the content** of video clips: what was said by people in the clip (statements made), the setting of the broadcast (e.g. a talk show), people represented in the video (entities such as politicians and celebrities) and events (significant occurrences, e.g. a demonstration) (see appendix B, figure 23 and 24). We again requested a ranking on a Likert scale of the usefulness of the aforementioned content aspects (see appendix B, figure 25).

Results show that in both tasks, participants judged the **words spoken** in a specific clip as most important in relation to relevance judgments. In second place for both searches, the scholars considered the **people who were seen in the fragment** as important. The researcher present was told by multiple users that some fragments were almost immediately discarded as non-relevant after e.g. particular celebrities appeared on the screen: "I cannot imagine this particular talk show host having anything interesting to say." (on the importance of context and background knowledge, see 2.4.5).

Results Q2	
How is relevance of AV search results assessed?	
➤	In both the first and second task, a large majority of participants judged their overall search results as somewhat to very relevant.
➤	Whilst performing the first task in detailed view, all participants made use of the textual program information displayed.
➤	For the task in detailed view, users first judged <i>the accompanying program information</i> as the most important criterion for evaluating the relevance of a particular clip and second, viewing <i>the content of the clip</i> .
➤	For the thumbnail view test, <i>watching the content of the clip</i> was first judged as the most important aspect for evaluating the relevance of a particular clip and second, <i>the thumbnail picture</i> .
➤	In relation to the content of video clips, participants in both tasks first judged <i>the words spoken</i> in a specific clip as most important in relation to relevance judgments and second, the <i>people who were seen</i> in the fragment.

2.4.4 Q3: To what extent do humanities scholars consider a lack of annotations an issue?

The aforementioned results leave out the extent to which humanities researchers consider a lack of archivists' annotations a problem, and also whether they would personally contemplate working in thumbnail view for their own research. To address this sub-question, the participants were requested to estimate and compare the difficulty levels of both tasks:

Level	Detailed View (number of participants)	Thumbnail View (number of participants)
Very easy	2	0
Easy	13	12
Neutral	6	4
Difficult	4	9
Very difficult	0	0

Result table 3: Answers to the following question (asked after each task): 'How would you rate the overall difficulty of the previous test?'

Level	Number of participants
Much easier	1
Somewhat easier	6
The same	3
Somewhat harder	7
Much harder	10

Result table 4: Answers to the following question (asked after the final task): 'How would you rate the overall difficulty of the first task (in detailed view) in comparison with the second task (in thumbnail view)?'

Table 3 shows that the results on the difficulty levels of both tasks lie relatively close together, although more than twice as many participants ticked **difficult** after the thumbnail view task than after the detailed view task. After being asked to compare both tasks at the end of the experiment, 17 out of 27 participants considered the thumbnail view task to be **somewhat to much harder than the detailed view task**.

Nearly all test subjects searched through their SERPs in a systematic way (especially from top to bottom and in the thumbnail view task, from left to right as well). For the thumbnail view task, the picture displayed also played a role of importance in the decision where and how to start searching: 13 people decided to click on a specific clip purely because the image 'looked promising' (in most cases, a well-known picture of the politician in question from e.g. newspapers). This result is in accordance with our previous observation presented in paragraph 2.4.3 that the thumbnail picture functions as a significantly more important relevance criterion in the thumbnail view task than in the detailed view task.

To determine whether the scholars would consider making use of a relatively extensive but poorly annotated audio-visual collection in their own future research, we composed an open question: 'Suppose you could only search in 'thumbnail view' (i.e. without the associated program information), would you consider it worthwhile to digitize relevant audio-visual material for your research?' The reactions to this question were mixed:

Answer option	Number of participants
---------------	------------------------

Yes, digitize AV materials, I am able to find my way	9
No, I prefer a carefully unlocked and annotated archive	11
Compromise	5
No opinion	2
Total	27

Result table 5: Whether participants were willing to work in thumbnail view during their own research.

9 out of 27 participants indicated to rather work with a larger albeit poorly annotated digital collection. 11 humanities researchers preferred a smaller, but carefully unlocked and annotated audio-visual archive. Below are some quotes of the participants who preferred to work the traditional (extensively annotated) way:

Answer option No:

- "I sorely missed the program information. A minimal description at the least is absolutely necessary, consisting of persons in the clip, the subject, and date of the broadcast. Otherwise too much time is lost on scanning irrelevant materials
- "I am used to working with metadata and all its possibilities, so I wouldn't want to search in thumbnail view for my own studies. I think it will be some time before visual search has been sufficiently developed and operationalized to actually be of use to most humanities researchers."
- "It's problematic that in thumbnail view many generic images are displayed [e.g. a well-known image of a politician]. From this picture, it's impossible to derive any clues about the contents of the clip. Therefore, it's necessary to add program information, otherwise you have to click on each individual clip to see if it pays off."

Below are some quotes of users who indicated that they would be willing to work with poor archivists' annotations provided that larger amounts of audio-visual materials could be digitized:

Answer option Yes:

- "I would be able to work in thumbnail view. The number of relevant clips this way of searching yielded was pleasantly surprising."
- "A slightly easier way of searching does not outweigh the risk to just miss that one crucial fragment for your research. In my opinion, visual search has the future, researchers working in the traditional way should better get used to this new technique."
- "I prefer to work with the widest possible selection of digitized audio-visual materials, so I will manage to find my own way through the archive. I don't have much confidence in program descriptions produced by archivists."

As can be seen from result table 5, five participants out of 27 chose 'a middle ground', meaning that they preferred a combination of visual search with some basic form of annotation, whether produced by an archivist or not:

Answer option Compromise:

- "I might work with visual search, if the metadata was incomplete or inadequate anyway. But I'd rather search the traditional way if the program information was extensive and produced by an educated and experienced archivist."
- "I am hopeful about new techniques such as speech recognition and image search, as well as the development of systems in which users can tag and produce their own metadata."

Results Q3

To what extent do humanities scholars consider a lack of annotations an issue?

- More than twice as many participants ticked 'difficult' after the thumbnail view task than after the detailed view task.
- After being asked to compare both tasks at the end of the experiment, 17 out of 27 participants considered the thumbnail view task to be somewhat or much harder than the detailed view task.
- 9 out of 27 participants indicated to rather work with a larger albeit poorly annotated digital collection, as opposed to 11 humanities researchers who preferred a smaller, but carefully unlocked and extensively annotated audio-visual archive.

2.4.5 Context and background

An important factor to take into account with all results is that in general the scholars, whilst making their relevance judgments, relied heavily on their own previous experiences and knowledge. For more than four fifths of all users, use of their background knowledge was crucial in the decision-making process. Upon encountering a particular program title, almost all participants immediately connected the name to a specific setting, program type, reputation and sometimes even a show host. Interestingly enough, 19 out of 27 participants also voluntarily commented on evidently non-relevant search results, which were then immediately eliminated from their evaluation process. For purposes of illustration, see some quotes below:

Background knowledge:

- "*Nova* [Dutch current affairs show] is a suitable program for the task, as I am looking for shows in which interviews are conducted."
- "I do not hold *DWDD* [Dutch current affairs show] in high regard, Matthijs van Nieuwkerk [the host] always comes across as very superficial."
- "I always find *Nova College Tour* very informative, and furthermore, I know that Ayaan Hirsi Ali was once a guest on this show."
- "*Boer zoekt Vrouw* [Dutch dating show] cannot have anything to do with it."

2.5 Conclusions

Using image and voice recognition, CBIR systems such as AXES RESEARCH have now made manually unannotated video archives searchable by means of automatic linking to 'known entities', implying that it is useful to digitize thousands of hours of undescribed and unused audio-visual material currently being stored in archives. However, the technical potential of AXES RESEARCH to generate unannotated search results does raise questions regarding their usability and usefulness for scholars. Therefore, the objective of our user study was to help archivists, librarians and ICT-researchers and developers to gain more insight into the various ways humanities scholars navigate and seek in audio-visual archives, and their ability to make relevance judgments when presented with poor archivist annotations. Our main question was: ***How does a lack of accompanying archivists' annotations influence the ability of humanities researchers to make relevance judgments about audio-visual documents?***

Based on our collected data, we may conclude that humanities scholars in general are able to work with poor archivists' annotations when searching through an audio-visual archive. They have shown to be able to locate useful documents and make well-considered relevance judgments. However, navigating without archivists' annotations is perceived as more difficult than the traditional method wherein extensive associated metadata can be consulted.

First, we determined that the ***humanities researchers were able to find relevant audio-visual material for their queries using the CBIR option in AXES RESEARCH.*** No discernible differences could be detected between the results of the two information retrieval systems used: the participants achieved similar scores regarding videos seen and saved in the text-based search as the CBIR search.

As for the visual representation of the search results, the participants were also able to complete both the task where the program information was displayed ('detailed view') and the task where no archivists' annotations were shown ('thumbnail view'). Our results show that the scholars ***managed to add at least as many videos to their favorites list without the program information than with the metadata displayed.*** In our experiment, a 'favorite' equalled a relevant video. Interestingly enough, the number of videos clicked on and watched in the exact same time frame was more than twice as large for the task where no metadata was displayed.

Context proved king during the experiment: a large majority of the humanities scholars **relied heavily on their personal prior experiences and background knowledge** whilst making relevance judgments. Most program titles, settings and faces were automatically connected to detailed information regarding specific broadcast types, (network) reputations and entities involved. A quick scan of the archivists' annotations and/or the contents of a particular clip therefore proved all the scholars needed to make a well-measured relevance judgment. It should, however, be interesting to consider what would be the outcome when extensive background knowledge is missing, e.g., when scholars are searching through unknown television archives of other countries.

In our study, we also tested the importance and usefulness of several relevance criteria in relation to the audio-visual documents: the visual appeal of a clip, the thumbnail displayed, the ranking of the clip in the results list, the content of the clip and the accompanying program information. Whilst performing the first task in detailed view, the participants had the most use for **the textual program information displayed**, confirming our earlier assumption that humanities researchers – 'expert users of text' (Duff and Johnson, 2002) – relatively are (still) very text-oriented and dependent on the associated archivists' annotations. Also considered important was viewing the content of the clip. For the thumbnail view test, **the content of the clip** and the associated thumbnail picture were judged as the most important relevance factors.

The experiment included the importance and usefulness of four variables related to the content of video clips as well: what was said by people in the clip, the setting of the broadcast, and the people represented in the video and events. We found that in both the thumbnail and the detailed view task, participants judged **the words spoken** in a specific clip as most important for making relevance judgments. Second, the scholars considered **the entities seen** in the fragment as important for both tasks. However, a critical comment on these results is necessary, as this outcome is strongly query-related. In both tasks, participants were specifically asked to look for opinions about two politicians. Very probably as a result thereof, the presence of these persons or other associated politicians in the clip was judged as a very important relevance criterion. We might safely assume that if the given query had been 'Eiffel Tower' instead of a person's name, the influence of 'people' would have been considerably diminished.

With regard to the user friendliness of thumbnails instead of detailed program information, no general consensus was reached. Although the participants succeeded to add as many relevant clips to their favorites with as without the annotations

displayed, the majority indicated to have perceived the task in thumbnail view as **more difficult**. As for their willingness to use a larger but poorly annotated audio-visual collection in their own future research, the feedback was also mixed. Half of the scholars felt that a higher ease of searching does not outweigh the benefits of a more extensive collection of digitized audio-visual. The other half prefers a necessarily smaller, yet easily searchable database.

Conclusions user study

- Humanities researchers were able to find relevant audio-visual material for their queries using the CBIR option in AXES RESEARCH.
- The scholars managed to add at least as many videos to their favorites list without the program information as with the metadata displayed.
- A large majority of the humanities scholars relied heavily on their personal prior experiences and background knowledge whilst making relevance judgments.
- Whilst performing the first task in detailed view, the participants had the most use for the *textual program information displayed* as a relevance criterion; for the second task in thumbnail view, the *content of the clip* was judged as the most important relevance factor.
- Regarding relevance criteria in relation to *the content* of the clips, we found that in both the thumbnail and the detailed view task participants judged *the words spoken* in a specific clip as most important for making relevance judgments.
- The majority of participants indicated to have perceived the task with absent archivists' annotations as more difficult than the task where the program information was displayed.

3 EVALUATION AXES HOME

During the last year of the AXES project the AXES HOME system has been delivered. This system is specially designed for home users. In November 25 users at the BBC tested the system, focussing on two aspects:

- ***The overall acceptance of the AXES HOME system;***
- ***An assessment of potentially interesting models for users to access audio-visual archives.***

3.1 Defining the home user

Earlier in the project a requirements study was conducted to determine the features and functions of AXES HOME, see D1.2 – User Requirement Report V1. This group is defined as ‘users who engage with the system for their own personal entertainment or knowledge seeking activities.’ While this definition is broad, it is important to encompass everyone. Also the term ‘home user’ is itself of course misleading, as personal users can be interacting with AXES elsewhere, not just in the home.

Inevitably there is a crossover in common requirements for all groups such as clarity, efficiency, and usability of whatever services they are interacting with. However, home users have distinctive characteristics that drive their Internet usage. Obvious ones are **entertainment** and **enjoyment** and facing challenges such as in the gaming and general leisure sphere. Time is not necessarily as much of a constraint as it would be for someone using AXES tools in a professional capacity (such as a journalists that are catered for in AXES RESEARCH or media professionals that are catered for in AXES PRO). Elements like these are distinct features of the home user and indicate the



Figure 3: The photo that was used to visualise the personas of the AXES home users

difference with the other user groups. A **requirement** following from these profile matrices that were developed for the requirements studies (see D1.2) is that the AXES HOME system should be suitable for people who are inexperienced, yet goal-directed in their search and consumption of online video (Kemman, Kleppe and Beunders 2012). Although small differences were found between different groups of people, this is true for all home users as surveyed.

The user requirements stage gave us information with which we could formulate personas for the types of people we envisaged would use AXES HOME. A goal of personas is to support the developing team to 'evaluate ideas from a user's point of view'. For each persona, a person is created that mirrors a real group of users enriched with imaginary facts to make it more personal (personalisation). For the AXES HOME system we chose to create a persona family consisting of a father, mother, son and daughter. For a full description of the home user persona, see D1.2, User Requirements Report v1, pages 112 – 113.

3.2 AXES Home Digital Library System

Taking into account the results from the Requirements Study (D1.2) the AXES HOME system was developed as a system that can be used on tablets. The design of the user interface for the Home system comes from several AXES sources, including the earlier AXES RESEARCH interface, the related mock-ups and specifications described in D7.1, D7.2, D7.4, and on the user requirements, personas as described in D1.2. It allows users to search the system but also offers users new pathways to enter the archive. For this goal, the sections **Browse** and **News Feeds** were designed. In the Browse section users could see video items that were automatically retrieved based on a random selection of topics. In the News Feeds section, textual representations of recent news items were shown in different domains (e.g., news, culture, celebrities). After the user clicked on a news item, the system automatically retrieved relevant video items for this item from the archive.

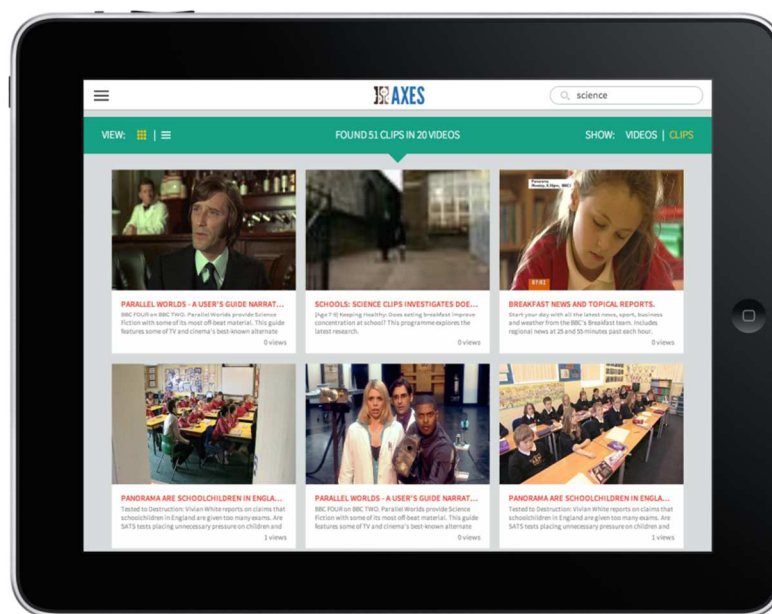


Figure 4: The AXES HOME system on an iPad

3.3 Goal of the evaluation

The goal of the evaluation of the AXES HOME system was twofold. First we wanted **to test to what extent home users could actually use the current system**. Five aspects stood central to this assessment:

- **Performance** – Did the system and underlying techniques work adequately?
- **Navigation and functionality** – User interaction with the system. The ease of use and ‘learnability’.
- **Discovery** of tools and understanding of what they offer.
- **Impediments** – Noting any factors that have a negative effect on using the system.
- **Requested Features and improvements** – User may identify things that would improve user experience or efficiency.
- **Entertainment function** – The main goal of AXES HOME is to entertain the user.

These aspects were aimed at giving concrete feedback to the developers of the AXES HOME system.

The second goal of the evaluation focussed on a more generic question: **Can we think of alternative access models that are interesting for users to access audio-visual archives?** Since AXES HOME was aimed at entertaining the home users and allowing them new pathways to explore digital audio-visual archives, this matter was incorporated in the set-up of the evaluation.

3.4 Set-up of the evaluation

The AXES HOME system was introduced to and tested by members of the public at the BBC Visitor Centre in London on 26 - 28th November 2014. The BBC has a busy visitor centre and is well attended by tourists visiting London. It offers a ‘window into the news room’ where is it possible to see news production taking place as well as live broadcasting. There are tours of the building, including studios, and an interactive session where visitors can read the news and make a radio recording. As each tour came to an end, the tour guides invited the public to come over to our evaluation session. The large volume of people ensured we had a good opportunity to show AXES to a wide age group. For these sessions we



Figure 5: The entrance of the Visitors Centre of the BBC, where visitors were pointed to the AXES HOME Evaluation

focussed on people who were eighteen and over. The set-up involved having an AXES table with members of the project team available to sit down with participants and go through the various steps in the evaluation. We used iPad tablets with wireless capability to link to the AXES server at BBC Centre House. Large display boards and promotional material was also available to disseminate the project around the visitor centre (see appendix G).

We aimed at setting up a qualitative evaluation. In total 25 persons participated ranging in the ages from 18 till 60+. At around twenty participants we reached the point of theoretical saturation (Morse, 2007), meaning that no new data or insights appeared and no additional data was needed at that point. However, we continued to talk to 5 more persons to guarantee that all age groups were covered, thus reaching a total of 25 participants. The interviews lasted between 30 minutes up to one hour. At the end each interviewee filled out an electronic survey. See appendix E for the full questionnaire.

During the preparation phase a workflow was developed to organize the evaluations as smoothly as possible. During the testing phase the workflow was adapted into the final format as included in appendix D. This workflow enabled us to efficiently get responses of the audience and gave us an impression. After people finished the testing their reactions were recorded by a survey that was prepared beforehand and follows the same aspects. Each user was assigned a unique user ID both in the AXES HOME system as well as the exit survey. The next paragraph describes the observations of the researchers and the results of the exit survey.

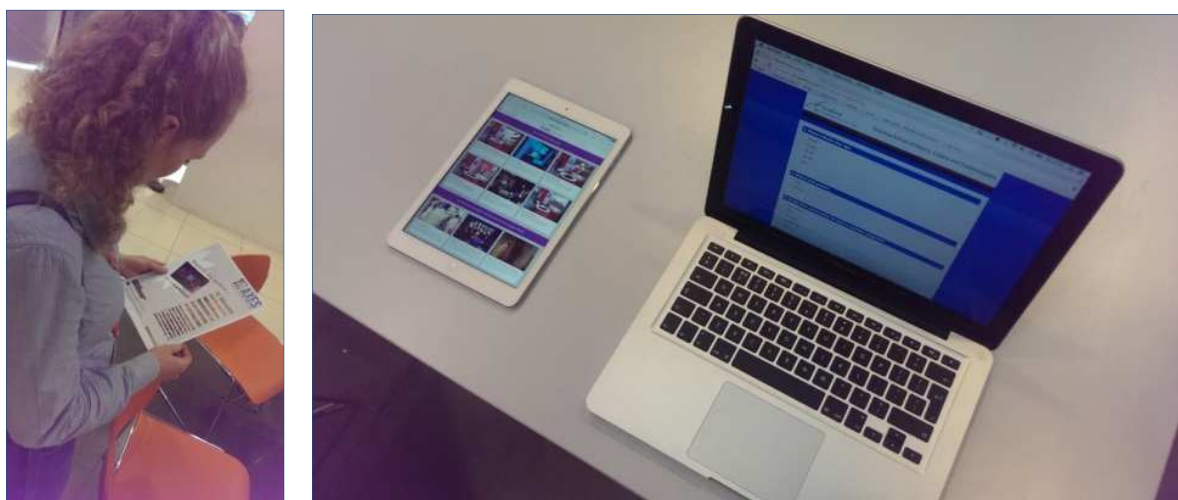


Figure 6 and 7: Participants were handed a leaflet with information about the AXES project. The tests were performed on an iPad. The exit survey was filled out on a laptop.

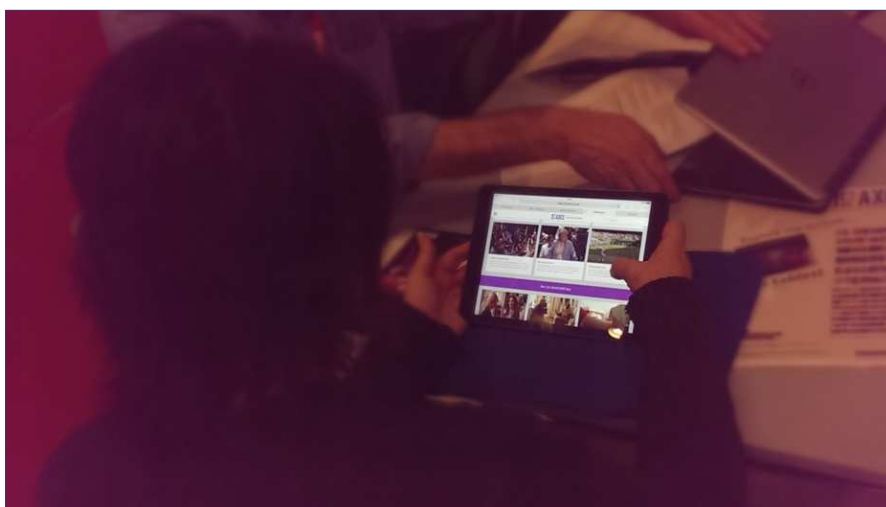
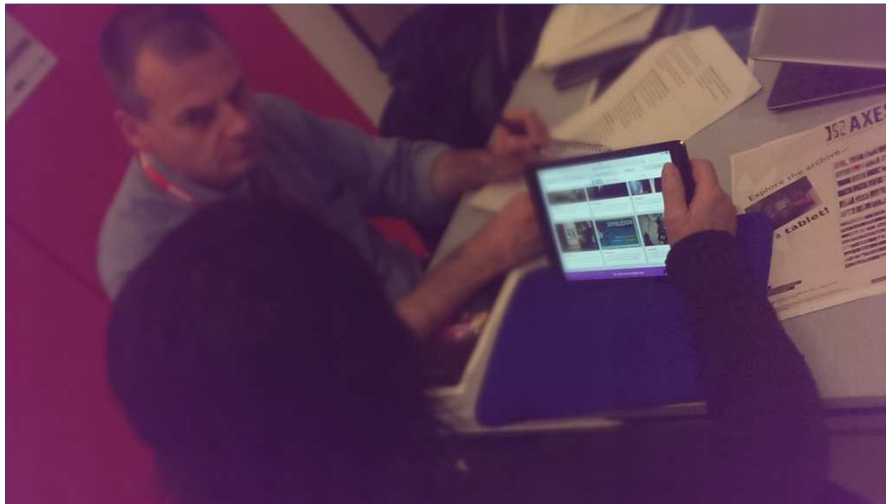


Figure 8, 9 and 10: Participants trying the AXES HOME system.

3.5 Results: Acceptance of the AXES HOME system

This paragraph describes the general results of the user testing and follows the first goal of the evaluation: the acceptance of the AXES HOME system. The second goal – what kind of access models are interesting for users to access audio-visual archives? – is discussed in the conclusion paragraph of this chapter.

All detailed results of the exit survey are included in appendix F. A selection of the results are included in the text when they illustrate the main argument. A similar approach is followed with the quotes of the interviewees and their answers to the open questions of the survey. A list of suggested improvements, based on the outcome of this chapter is included in appendix C.

In general, most users were very positive about AXES HOME. Even though they needed some instruction on how to use the system, most users grasped the idea of the system and were able to use it. The answers to the open question on ‘what did you like most about the AXES System and why?’ range from ‘easy to navigate’ to ‘the transcript search is very useful’ and ‘it provides new ways to explore the collection’. Answers to the closing question of the exit survey – ‘would you use this system at home in your free time?’ – also indicate users are positive about AXES HOME.

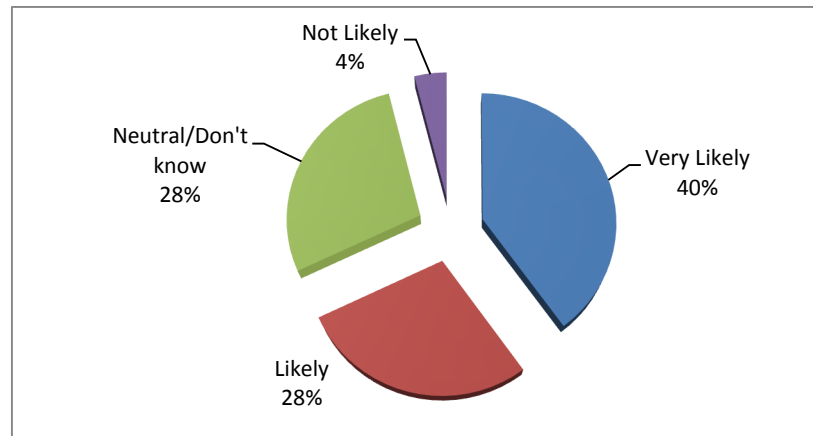


Figure 11: The results to the question ‘would you use this system at home in your free time?’

Furthermore, we asked users what they did *not* like about the system. These comments are incorporated in the following paragraphs.

3.5.1 Performance

A system like AXES HOME only functions well if the underlying techniques work correctly and when a large dataset is incorporated. Although in the BBC archive as a whole there may be a lot of relevant results given a specific query (e.g. based on a specific location), results may be sparse in the relatively small data set of a couple of

thousand hours of BBC content in the AXES system. Especially the sections Browse and News Feeds were interesting in this respect, since these sections offer pathways to the user to explore the archive by automatically finding relevant videos based on the topics or news item that were chosen by the user.

However, quite often the system did not show relevant video clips. At this point we are not able to state whether the ***underlying techniques did not work*** or whether the system ***did not contain sufficient videos***. Nevertheless, we noticed that users did not like it when (apparently) irrelevant results were shown. These results are now shown because the system always gives results even when the relevance score is low. To improve the usability of the system it is advised to look carefully at strategies for thresholding: better provide the user with no results rather than bad results. Or alternatively, make more clear why the system thinks something is relevant. After close inspection of a – deemed irrelevant – result, it often appears that from an observers' point of view it is 'imaginable' that the system indicates the item as relevant, for example because certain keywords are mentioned in one of the metadata fields, or because there is a certain visual similarity (e.g. glasses and a drawing with circles). Very often, however, this is outside the scope of relevance from the perspective of a user.

Performance
<ul style="list-style-type: none">➤ The performance of the system was criticized but given the type of evaluation the cause cannot be assessed properly and it therefore remains unknown if this can be explained by the techniques or the available data.➤ Users prefer the system to state it found 'No relevant results' instead of always showing results, even when they are not relevant.

3.5.2 Navigation and functionality

Users that were handed the iPad with the AXES HOME system were given a brief instruction on how to use the system. We trialled two approaches in the initial experiment setup. In the first the user was not given an instruction. In the second instruction was given. This set-up would have enabled us to gain insight into the intuitiveness of the system. However, during the initial trials of this evaluation set-up we learned almost immediately that the innovative functionalities of the AXES HOME system are relatively unknown making it difficult for users to grasp the essence of AXES in such a restricted time period. We therefore decided to give all users a brief introduction.

Our observations show the users understood the set-up of the interface, which is shown in the results of the exit survey. On the question how easy or difficult it was to use the system, no one stated it was difficult.

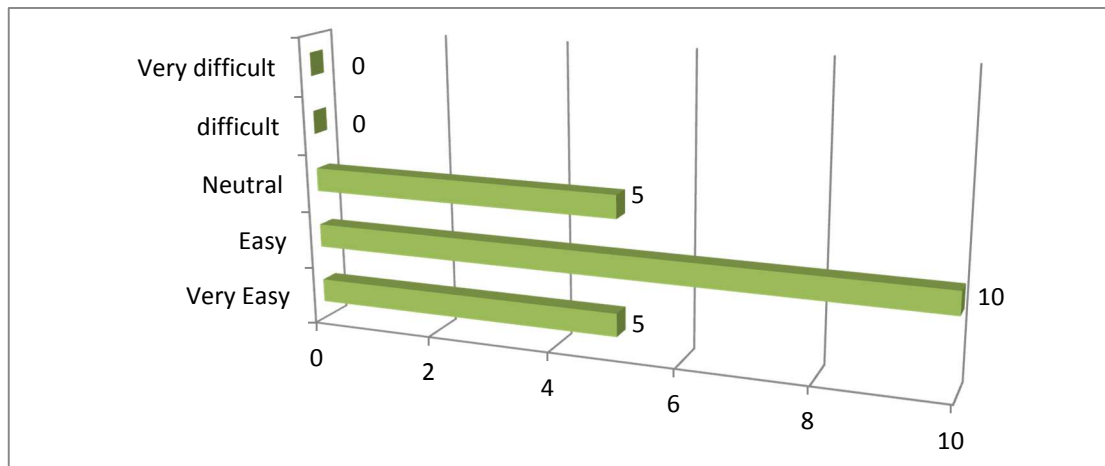


Figure 12: The answers to the question 'How easy or difficult was it to learn the system for you?'

However, this result does not mean everything was immediately clear to the user. Two observations contradict the outcome of this survey question. First, several users did not see **the search bar** on top of the page. Even though this is a requirement for the home users (See D1.2), the way it is currently designed does not help the user to see it right away. Second, the symbol of the three lines, referring to **the menu** is not clear to most users. The interviewer had to point to it for several times after which the interviewee stated he did not understand what the symbol stands for. It is advised to replace this icon by a tab that contains the words 'Menu'.

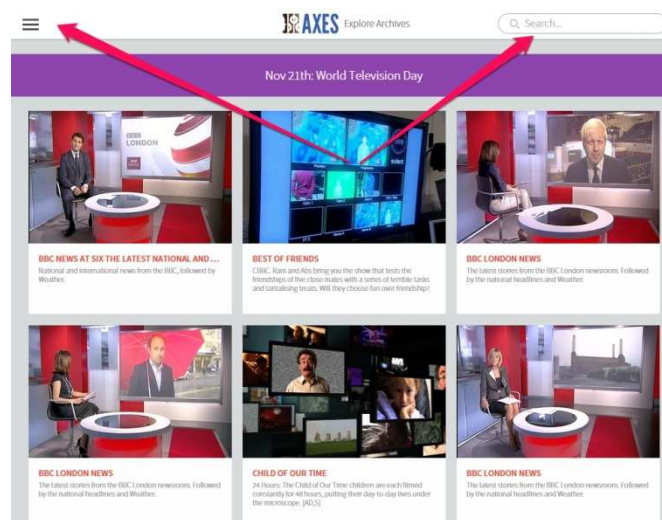


Figure 13: Both arrows point to the elements in the interface that were unclear to the user: the search box and the Menu pictogram.

The observers furthermore noted that several users were annoyed when scrolling both the Browse page as well as the results page of the News section. As the user goes down the page, the button 'Fetch More Random Topics!' disappears off the screen. The same happens at the results section of the News page. When one scrolls down, the original news feed item is off the screen, just like the options to show a detailed or thumbnail view.

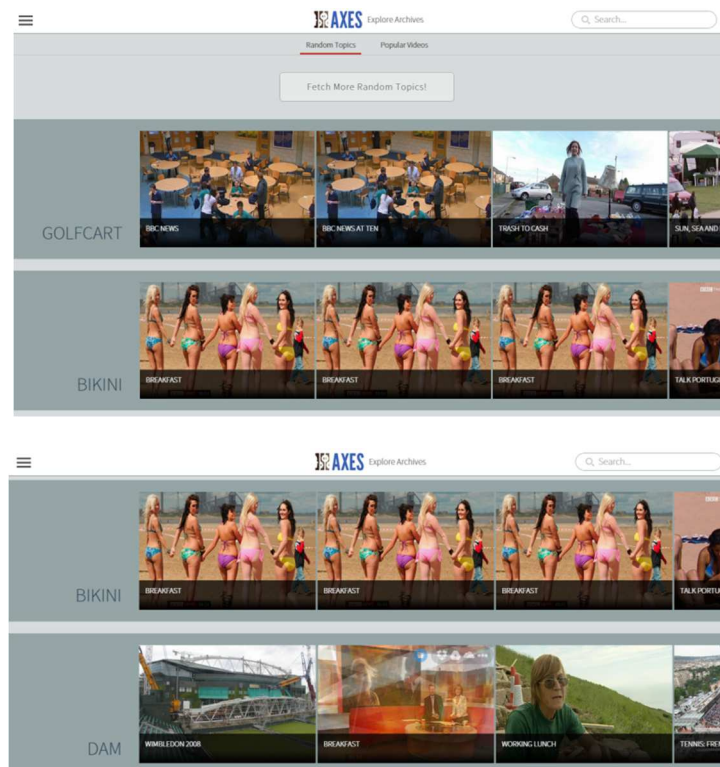


Figure 14: As the user scrolls down the Browse section, the button 'Fetch More Random Topics' disappears

Navigation and functionality

- The general impression of the interface is clear to the user.
- However, the *search bar* is not prominent enough.
- The *pictogram for the menu* is unclear. It is advised to replace this pictogram by a tab with the word 'Menu'.
- The *navigation tools* at the Browse Section and Results page of the News feeds disappear when scrolling and should stay at the top of the page.

3.5.3 Discovery of tools and understanding of what they offer

AXES HOME has partly been developed to help the user explore the archive. Two sections are specially designed for this: the News feeds section and the Browse section.

During the evaluation these sections were evaluated intensively together with the user feedback tools and the elements on the homepage.

News Feeds

The News feeds allow the user to go over recent news items and by clicking on an item, the system automatically searches for relevant video items. Respondents reacted positively to this feature. 71% of the users found it fairly useful to very useful (see appendix F) to find archive content. In general they agreed and were neutral towards the statement that the news feed browser allowed them to find interesting videos without the need for searching.

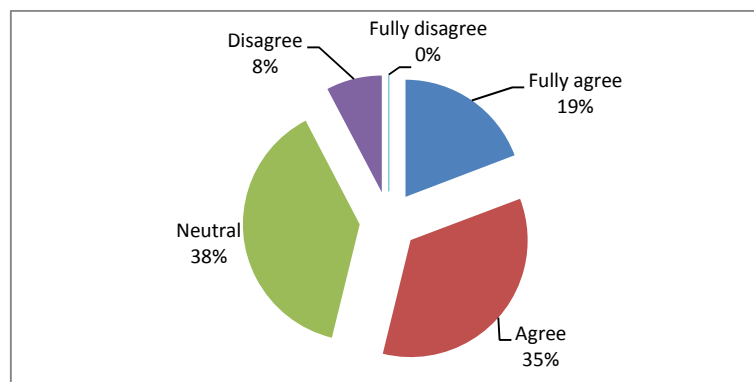


Figure 15: Answers to the question to what extent do you agree with the following statement: 'The news feed browser allows me to find interesting videos without the need for searching'.

Browsing

Similar reactions were observed with the Browse section. 69% of the testers considered this to be useful and 81% (fully) agreed with the statement 'The browsing options allow me to find interesting videos without the need for searching'.

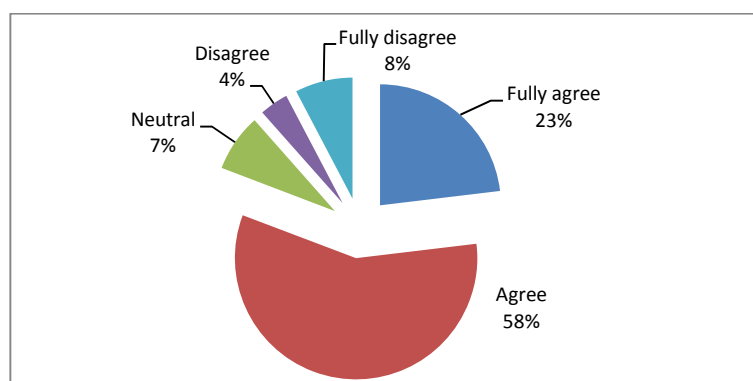


Figure 16: Answers to the question to what extent do you agree with the following statement: 'The browsing options allow me to find interesting videos without the need for searching'.

However, several testers noted that the terms used in the Browsing section were confusing and not true topic terms. This may not come as a surprise since the terminology that was used to show the topics stem from the concept-vocabulary choices made within the visual analysis algorithms (uberclassifiers). However, it is advised to change this into a standardized topic list or ideally, base the concept-vocabulary upon existing taxonomies. In either case, the goal should be to match the vocabulary of topics to the perception of the home user.

User feedback tools

Since the technology behind AXES HOME is relatively unfamiliar to the home user, user feedback tools play an important role in the system in order to help the user understand why certain results are retrieved. However, during the evaluations several testers commented they did not understand why certain results were retrieved. To pre-accommodate this comment, symbols are shown indicating which technique has been used to retrieve the results. During the tests, it became clear these symbols were too small to be observed by the testers. Making these symbols bigger is not a simple solution since it contradicts the previously formulated requirements of the home user to have a clean interface.



Figure 17: The arrow points to the small pictograms representing the techniques used to indicate how the results were retrieved.

We also noted that the 'Like' option underneath each video did not give a visual feedback after a user clicked on it. Below the video, the number of 'likes' do go up but ideally the heart would turn red after a user clicked on it.



Figure 18: The 'Like' button underneath each video.

Homepage

The homepage of the AXES HOME system is the first screen a user will visit. Given the exploration function of the system, in the current situation videos were offered referring to international days, such as 'World Television Day'. However, most users found the homepage **overwhelming, containing too much clutter**. Therefore, it did not show them the potential goal of the system in a direct manner.

In the exit survey we asked our testers which elements they prefer on the homepage. The answers vary a lot, although an instruction, recently viewed videos and popular videos are mentioned most (see appendix F), suggesting that the current set-up of the homepage is not the most appropriate.

Discovery of tools and understanding of what they offer

- The News Feeds and Browsing options are considered to be useful to find interesting videos without the need for searching.
- User feedback tools could be improved, though this might contradict previous requirements on a clean interface.
- The homepage needs to be redesigned to give the user a clear-cut impression of the potential of the system. However, the evaluation did not provide clear answer as to what should actually be on the homepage.

3.5.4 Impediments

While the users were trying out the system, the observer noted down positive and negative feedback by the users, including factors that had a negative effect on using the system. Our main findings concern terms that are incorporated into the system. In paragraph 3.2.3 we already mentioned that the topics shown in the Browse section are not clear enough for the users. The same is true for the terms that are used in the Advanced Search page.

The screenshot shows the AXES Advanced Search interface. At the top, the AXES logo and 'Explore Archives' text are visible. Below, there are five search categories, each with a search input field and a list of examples:

- Q Metadata:** search in titles, descriptions, and other metadata. Search input: 'Search metadata...'. Examples: Doctor Who, Top Gear.
- Q Speech:** search in subtitles, transcripts, or automatically detected speech. Search input: 'Search spoken words...'. Examples: (none listed).
- Q Visual categories:** search visually for broad terms and concepts. Search input: 'Search visual categories...'. Examples: cars, forest, landscape, skyscraper.
- Q Visual faces:** search visually for faces and facial attributes. Search input: 'Search faces...'. Examples: Tony Blair, Jeremy Clarkson, sunglasses, moustache.
- Q Visual instances:** search visually for specific places and logos. Search input: 'Search instances...'. Examples: London Bridge, bbc logo.

Figure 19: The current terms, used in the Advanced Search.

The terminology that is now used stems from the techniques that are behind AXES HOME. However we propose to make the following changes:

Current terms	Suggested terms
Metadata	Search the description (titles, descriptions, etc. E.g. Doctor Who, Top Gear)
Speech	Search spoken words (subtitles, transcripts or automatically detected speech)
Visual Categories	Search visual (faces, places, logos or concepts)
Visual Faces	
Visual Instances	

The most fundamental change we propose is to combine visual categories, visual face and visual instances into one option called 'Search visual'. We noted that not only does the advanced search page look overwhelming, these three options are also not clear to the user even though he/she does like the visual search option.

Impediments
<ul style="list-style-type: none"> ➤ In the Advanced Search section, terminology should be reconsidered to help the user understand the gains per option.

3.5.5 Requested features

During the tests we received several requests of users that knew quite well how they could use the system to find fragments of videos that relate to their hobbies. Often they want to perform a known-item search (instead of the browse functionality AXES HOME mainly offers). To accommodate this, they would like to:

- ***Search by date, e.g. see all programs that were broadcast on a date of birth.***

We also noted several additional ways of browsing the archive that are not included at this point:

- ***Search or browse on program titles, e.g. see programs of the users' childhood;***
- ***Search or browse on region, e.g. the place someone lives or used to live;***
- ***Search or browse on genre;***
- ***Search or browse on (famous) people.***

Requested features
<ul style="list-style-type: none"> ➤ To accommodate users' need to perform a known-item search, searching and sorting results on date could be included. ➤ Allow searching and browsing on program title, genre, region and names of (famous) people.

3.5.6 AXES HOME as entertainment tool

The main goal of AXES HOME is to have a system that accommodates an entertainment function for the home user. One of the main questions of the user evaluation was to what extent the current system is indeed a tool for entertainment. Given the answers to the question if people would use the system in their free time

(see paragraph 3.2.1), one could say 'Yes'. We indeed observed many users who were using the 'Browse' section to see what they could watch.

However, we also observed several users who immediately wanted to look for a television program that they knew of and that they wanted to see to satisfy their curiosity need. Two examples illustrate this nicely:

- One participant mentioned a gardening program of which she knew that was broadcast and which discussed a garden she would like to know more about. She started searching for it right away because she would love to find the fragment with that particular garden.
- Another participant said she was a teacher and searches for specific videos all the time to show in her class. While using AXES HOME she got very enthusiastic since she understood how easy it would be for her to find small fragments.

Both examples show real-world use cases that go beyond the browsing option of AXES HOME but refer to searches that support a research need. These users would probably benefit more of the AXES RESEARCH system given the focus of that system on finding specific video fragments for educational or research use. However, both users would not consider themselves researchers and would probably consider their searches as entertainment as well, indicating the difficulties when designing a system that needs to accommodate the needs and requirements of a wide range of home users.

AXES HOME as entertainment tool
➤ There is no clear-cut definition of 'entertainment', making it complicated to design a system that allows all sorts of entertainment needs.

3.6 Conclusion: Home user access models to audio-visual archives

Apart from the acceptance tests, the evaluation of AXES HOME also provided us with valuable information focussing on the question: ***what kind of access models are interesting for users to access audio-visual archives?*** AXES HOME is one of the first systems worldwide aiming at opening up archival audio-visual content that is centred around entertaining the home user without a clear-cut information need. Therefore, it is worthwhile to end this chapter with a reflection on the lessons learned. To do so, we build upon the results as presented in the previous paragraphs.

- ***Home users are interested in audio-visual archives.***

Given the very positive responses of the users, we can firmly state there is an interest of home users to explore (historical) audio-visual archives.

➤ ***Users need pathways.***

Given its entertaining function, users need to be guided into the archive via pathways that provide a journey by which users are navigated through the archive. When the performance of the system is up-to-date, sections such as Browse and News Feeds provide the pathways desired. These pathways can be further extended by allowing users to browse or search by date, region, genre or (famous) person. The design of such specific entry-points is challenging but should at least focus on guidance, simplicity and high performance.

➤ ***Users need to understand the concept.***

Given the novel techniques that are used, users need to be guided towards the possibilities of the system to grasp its potential. This can be achieved not only by creating a clear-cut instruction manual but also by redesigning the homepage in a way that shows the potential of the system immediately.

4 REFERENCES

- Auffret, G. and Y. P. Prié (1999), 'Managing Full-Indexed Audiovisual Documents: A New Perspective for the Humanities', *Computers and the Humanities* 33.
- Balatsoukas, P., and I. Ruthven (2012), 'An Eye-Tracking Approach to the Analysis of Relevance Judgments on the Web: The Case of Google Search Engine', *Journal of the American Society for Information Science and Technology* 63:9.
- Buchanan, G., and S.J. Cunningham, A. Blandford, J. Rimmer (2005), 'Information Seeking by Humanities Scholars', *Proc. ECDL2005*.
- Choi, Y., and E.M. Rasmussen (2002), 'Users' Relevance Criteria in Image Retrieval in American History', *Information Processing & Management* 38.
- Chung, E. (2012), 'Analysis of Multimedia Needs and Searching Features: An Exploratory Study', *ASIST2012*.
- Duff, W.M. and C.A. Johnson (2002), 'Accidentally Found on Purpose: Information-Seeking Behavior of Historians in Archives', *Library Quarterly* 72:4.
- Fickers, A. (2012), 'Towards a New Digital Historicism? Doing History In The Age Of Abundance', *Journal of European Television History and Culture* 1:1.
- Gibbs, G.R. *Analyzing Qualitative Data* (London 2008).
- Hollink, L., G.P. Nguyen, D.C. Koelma, A.T. Schreiber and M. Worring (2005), 'Assessing User Behaviour in News Video Retrieval', *IEE Proceedings* 152:6.
- Hollink, L., G. Schreiber, B. Huurnink, M. van Liempt, M. de Rijke, A. Smeulders and A. de Jong (2009), 'A Multidisciplinary Approach to Unlocking Television Broadcast Archives', *Interdisciplinary Science Reviews* 34:2.
- Huurnink, B., L. Hollink, W. van den Heuvel, and M. de Rijke (2010), 'Search Behavior of Media Professionals at an Audio-visual Archive: A Transaction Log Analysis', *Journal of the American Society for Information Science and Technology* 61:6.
- Jessop, M. (2004), 'Computing or Humanities? The Growth and Development of Humanities Computing', *Ubiquity* 41:5.
- Jong, F. de, R. Ordeman and S. Scagliola (2011), 'Audio-visual Collections and the User Needs of Scholars in the Humanities; a Case for Co-Development', *Proceedings of the 2nd Conference on Supporting Digital Humanities*.
- Kemman, M., M. Kleppe and H. Beunders, 'Who are the Users of a Video Search System? Classifying a Heterogeneous Group with a Profile Matrix' in: *International Workshop on Image Analysis for Multimedia Interactive Services* (Dublin 2012).
- Morse, J., 'Theoretical saturation', in: M. Lewis-Beck, A. Bryman, & T. Liao (eds.), *Encyclopedia of social science research methods* (Thousand Oaks 2004).
- Schamber, L. (2000), 'Time-line Interviews and Inductive Content Analysis: their Effectiveness for Exploring Cognitive Behaviors', *Journal of the American Society for Information Science* 51:8.
- Schmuckler, M.A. (2001) 'What is Ecological Validity? A Dimensional Analysis', *Infancy* 2.
- Vajcner, M. (2008), 'The Importance of Context for Digitized Archival Collections', *Journal of the Association for History and Computing* 11:1.
- Van den Heuvel, W. (2010), 'Expert Search for Radio and Television: a Case Study amongst Dutch Broadcast Professionals', *EuroITV*.
- Van der Kreeft, P., K. Macquarrie, M. Kemman, M., Kleppe and K. McGuinness (2014), 'AXES-RESEARCH — A User-oriented Tool for Enhanced Multimodal Search and Retrieval in Audiovisual Libraries', *12th International Workshop on Content-Based Multimedia Indexing*.
- Van Someren, M.W., Barnard, Y.F. and Sandberg, J.A.C., *The Think Aloud Method. A Practical Guide to Modelling Cognitive Processes* (London 1994).
- Yang, M., B.M. Wildemuth and G. Marchionini (2004), 'The Relative Effectiveness of Concept-based Versus Content-based Video Retrieval', *MM'04*.
- Yang, M., & Marchionini, G. (2004), 'Exploring Users' Video Relevance Criteria - A Pilot Study', *Proceedings of the American Society for Information Science and Technology* 41:1.
- Zarro, M. (2007), 'The Information Seeking Needs and Habits of Humanities Scholars in the Digital Age', *Proc. ECDL 2005*.

Appendix A **SURVEYS USER STUDY AXES RESEARCH**

This appendix contains the three AXES RESEARCH surveys of which the results are described in chapter 2.

Part I (administered after the task with archivists' annotations):

1. To which user group have you been assigned? *(For the interviewer)*
 - A
 - B
2. You have selected several video clips to watch. Please describe in a few sentences which aspects made you decide to click on these particular videos. In other words, why did you think they might be relevant to your search query?
(Open)
3. Were you able to find clips that were (somewhat) relevant to your search query?
 - Yes
 - No (*> Go to question 8*)

4. Please indicate if, and to which extent the following aspect(s) of a particular video clip made you decide it was useful or relevant:

The video clip was high ranked.

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

The accompanying program information.

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

The thumbnail.

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

The content of the clip.

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

The video was visually interesting and appealing to me.

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

Please indicate additional factors or comments which have been important in your judgment, if any.
(Open)

5. Please prioritize the three aspects mentioned in the previous question you consider most important for making your relevance judgments.
(Ranking)

6. Overall, I found the retrieved search results:
- Very relevant
 - Partially relevant
 - Not relevant
7. Overall, finding video clips (partially) relevant to the query was:
- Very easy
 - Easy
 - Moderate
 - Somewhat hard
 - Very hard
8. Have you read the accompanying program information next to the retrieved video clips?
- Yes, I read the information thoroughly
 - Yes, but I only scanned the information quickly
 - No, I have not paid attention to the accompanying information
9. Which accompanying program information, if any, did you find helpful while determining the relevance or usefulness of a particular retrieved video clip?

☐ None of the accompanying program information was helpful (> *Go to question 11*)

Program title:

- ☐ Not at all helpful
- ☐ A little helpful
- ☐ Somewhat helpful
- ☐ Very helpful
- ☐ Extremely helpful

Short program description:

- ☐ Not at all helpful
- ☐ A little helpful
- ☐ Somewhat helpful
- ☐ Very helpful
- ☐ Extremely helpful

Air date:

- ☐ Not at all helpful
- ☐ A little helpful
- ☐ Somewhat helpful
- ☐ Very helpful
- ☐ Extremely helpful

Category:

- ☐ Not at all helpful
- ☐ A little helpful
- ☐ Somewhat helpful
- ☐ Very helpful
- ☐ Extremely helpful

Keywords:

- ☐ Not at all helpful
- ☐ A little helpful
- ☐ Somewhat helpful
- ☐ Very helpful
- ☐ Extremely helpful

Genre:

- ☐ Not at all helpful
- ☐ A little helpful
- ☐ Somewhat helpful
- ☐ Very helpful
- ☐ Extremely helpful

Contributors:

- ☐ Not at all helpful
- ☐ A little helpful
- ☐ Somewhat helpful
- ☐ Very helpful
- ☐ Extremely helpful

Entities:

- ☐ Not at all helpful
- ☐ A little helpful
- ☐ Somewhat helpful
- ☐ Very helpful
- ☐ Extremely helpful

Transcript:

- ☐ Not at all helpful
- ☐ A little helpful

- ☐ Somewhat helpful
- ☐ Very helpful
- ☐ Extremely helpful

10. Please prioritize the three aspects mentioned in the previous question you consider most important for making your relevance judgments.
(Ranking)

11. Please indicate if, and to which extent the following aspect(s) of *the content* of a particular video clip made you decide it was useful or relevant:

People:

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

Places:

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

Spoken words:

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

Events:

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

12. Please prioritize the three aspects mentioned in the previous question you consider most important for making your relevance judgments.
(Ranking)

Survey Part II (administered after the task without archivists' annotations):

13. You have selected several video clips to watch. Please describe in a few sentences which aspects made you decide to click on these particular videos. In other words, why did you think they might be relevant to your search query?
(Open)

14. Were you able to find clips that were (somewhat) relevant to your search query?

- Yes
- No (> Go to question 19)

15. Please indicate if, and to which extent the following aspect(s) of a particular video clip made you decide it was useful or relevant:

The video clip was high ranked.

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

The accompanying program information.

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

The thumbnail.

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

The content of the clip.

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

The video was visually interesting and appealing to me.

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

Please indicate additional factors or comments which have been important in your judgment, if any.
(Open)

16. Please prioritize the three aspects mentioned in the previous question you consider most important for making your relevance judgments.
(Ranking)

17. Overall, I found the retrieved search results:

- Very relevant
- Partially relevant
- Not relevant

18. Overall, finding video clips (partially) relevant to the query was:

- Very easy
- Easy
- Moderate
- Somewhat hard
- Very hard

19. In the previous task, you were able to read the accompanying program information while determining the relevance or usefulness of a particular video clip. In this task, you only had the thumbnails and the program title to go by. Which information, if any, did you miss when making your selection of relevant videos?

- ☐ I did not miss any program information

A short program description:

- ☐ Not at all
- ☐ To a small degree
- ☐ To a moderate degree
- ☐ To a considerable degree
- ☐ To a great degree

What kind of program it is:

- ☐ Not at all
- ☐ To a small degree
- ☐ To a moderate degree
- ☐ To a considerable degree
- ☐ To a great degree

The genre of the program:

- ☐ Not at all
- ☐ To a small degree

The air date of the program:

- ☐ Not at all
- ☐ To a small degree
- ☐ To a moderate degree
- ☐ To a considerable degree
- ☐ To a great degree

Keywords:

- ☐ Not at all
- ☐ To a small degree
- ☐ To a moderate degree
- ☐ To a considerable degree
- ☐ To a great degree

The contributors to the video:

- ☐ Not at all
- ☐ To a small degree

- ☐ To a moderate degree
- ☐ To a considerable degree
- ☐ To a great degree

The entities in the video:

- ☐ Not at all
- ☐ To a small degree
- ☐ To a moderate degree
- ☐ To a considerable degree
- ☐ To a great degree

- ☐ To a moderate degree
- ☐ To a considerable degree
- ☐ To a great degree

The transcript of the video:

- ☐ Not at all
- ☐ To a small degree
- ☐ To a moderate degree
- ☐ To a considerable degree
- ☐ To a great degree

20. Please prioritize the three aspects mentioned in the previous question you consider most important for making your relevance judgments.

(Ranking)

21. Please indicate if, and to which extent the following aspect(s) of *the content* of a particular video clip made you decide it was useful or relevant:

People:

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

Places:

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

Spoken words:

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

Events:

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

22. Please prioritize the three aspects mentioned in the previous question you consider most important for making your relevance judgments.

(Ranking)

23. In comparison with searching the list of results in 'detailed' view, I consider searching with the 'thumbnail' view:

- Much easier
- Easier
- Comparable
- Somewhat harder
- Much harder

24. If you were only able to search the way you did in this task ('thumbnail' view), would you consider it worthwhile to digitize audio-visual material relevant to your own research, lacking any accompanying program information?

(Open)

Part III (exit survey, administered after the experiment):

1. Please provide your age:

- ☐ 18-24
- ☐ 25-34
- ☐ 35-44
- ☐ 45-59
- ☐ 60+

2. What is your gender?

- ☐ Male
- ☐ Female

3. Please state your occupation:

(Open)

4. University / research institution:

(Open)

5. How often do you make use of audio-visual material as a source?

- ☐ Very often
- ☐ Often
- ☐ Occasionally
- ☐ Rarely
- ☐ Never

6. Do you usually know, before starting a search in an audio-visual archive, what you are looking for?

- ☐ Strongly agree
- ☐ Agree
- ☐ Undecided
- ☐ Disagree
- ☐ Strongly disagree

7. Please indicate if, and to which extent the following aspects of a video clip make you decide if it will be relevant or useful as a source to your research:

The video is relevant to my topic.

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

The video is an accurate representation of what I am looking for.

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

The video represents the time periods of my study.

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

The video gives me new ideas or new insights.

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

The video is new to me.

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

The video contains the kinds of details I can use to clarify important aspects of my research area.

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

The video is free or easy to obtain.

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

I like the video.

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

The technical quality of the video is high.

- ☐ Very important
- ☐ Important
- ☐ Moderately important
- ☐ Of little importance
- ☐ Unimportant

Please indicate additional factors which are important in your relevance judgments, if any:
(Open)

8. Please prioritize the three aspects mentioned in the previous question you consider most important for making your relevance judgments:

(Ranking)

9. As a final effort, can you explain how you would evaluate your audio-visual materials through source criticism when using a system such as AXES RESEARCH?

(Open)

10. In your opinion, does AXES RESEARCH work better than the search systems for audio-visual materials you normally use?

- Yes
- No

Appendix B **GRAPHS USER STUDY AXES RESEARCH**

How is relevance of audio-visual search results assessed by humanities scholars? For the thumbnail as well as the detailed view task, the importance and usefulness was tested of five general aspects in relation to the relevance evaluation of audio-visual material. The given response options were: 1. Very important / 2. Important / 3. Fairly important / 4. Unimportant / 5. Very unimportant (Scale 1-5, the lower the number, the higher the importance). Since there were no discernable differences between the choices of user group A and B, their results were added together.

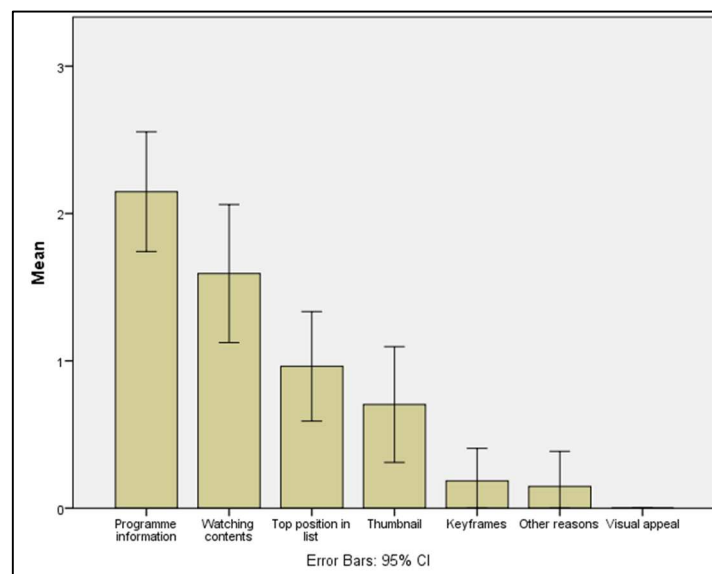


Figure 20: The importance of general relevance criteria for judging audio-visual material in detailed view.

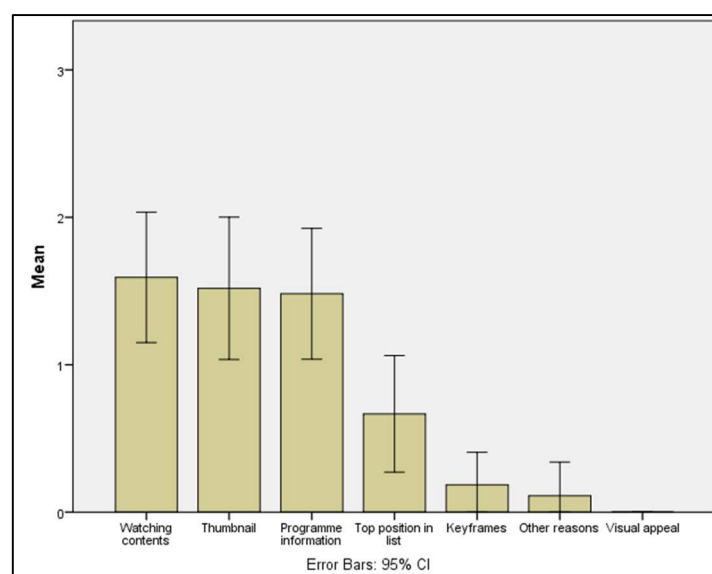


Figure 21: The importance of general relevance criteria for judging audio-visual material in thumbnail view.

As it was possible in the survey to tick multiple criteria as e.g. 'important', we asked users to create a ranking, or in other words, to make a top 3 list of the given aspects. The following results emerged after the scholars assigned positions to the provided general relevance criteria. The graph below is sorted from important (down) to unimportant (up). The given response options were: 1. Position one / 2. Position two / 3. Position 3 (Scale 1-3, the higher the number, the higher the importance):

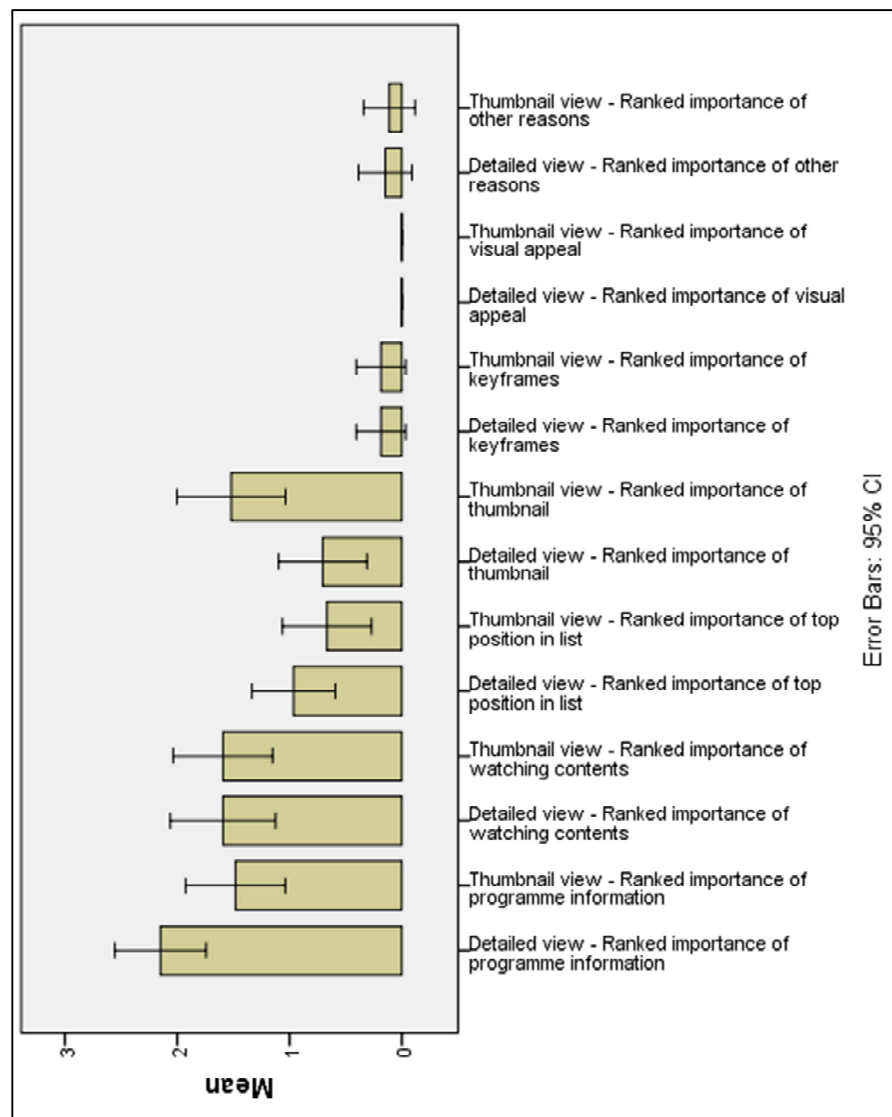


Figure 22: The aforementioned relevance criteria for judging audio-visual material ranked.

For the thumbnail as well as the detailed view task, the importance and usefulness was tested of four general aspects in relation to *the content* of video clips. The given response options were: 1. Very important / 2. Important / 3. Fairly important / 4. Unimportant / 5. Very unimportant (Scale 1-5, the lower the number, the higher the importance).

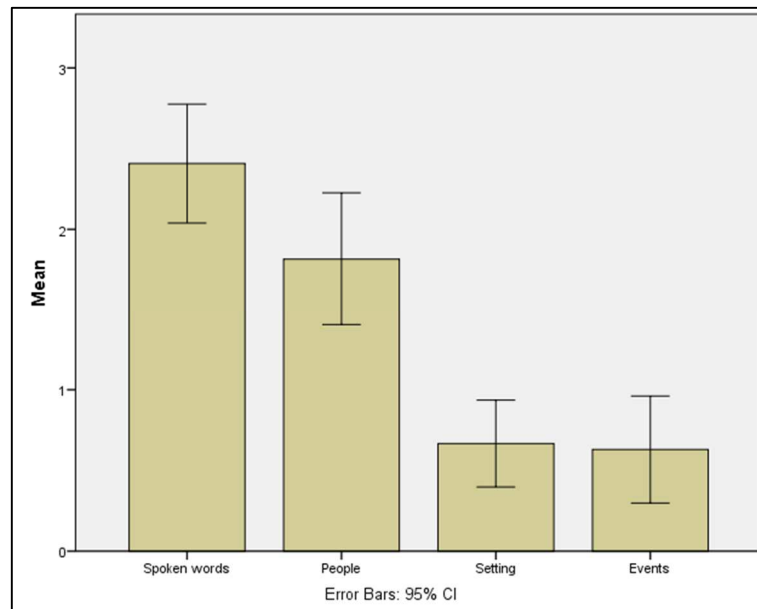


Figure 23: The importance of the *content* of video clips in detailed view.

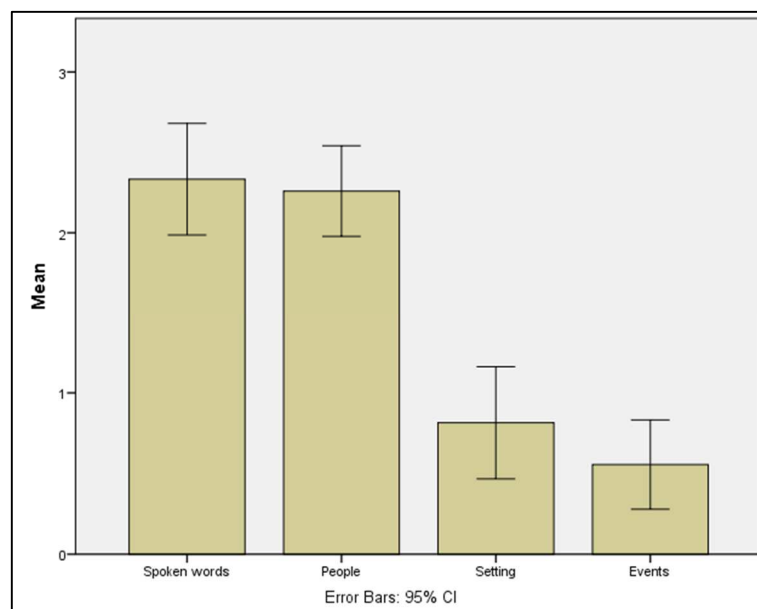


Figure 24: The importance of the *content* of video clips in thumbnail view.

We again requested the participants to rank the usefulness of the aforementioned aspects of *the content* of the clips on a Likert scale. The results are presented below: the graph is sorted from important (left) to unimportant (right). The given response options were: 1. Position one / 2. Position two / 3. Position 3 (Scale 1-3, the lower the number, the higher the importance).

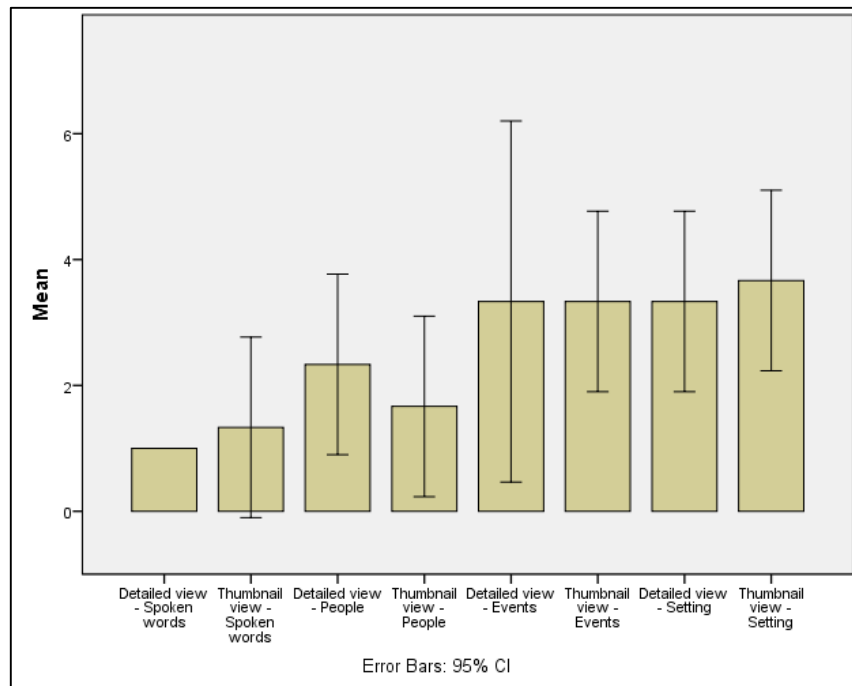


Figure 25: The importance of the *content* of video clips ranked.

Appendix C IMPROVEMENTS AXES HOME

Based on the evaluations of AXES HOME, the participating interviewers and observers jointly made a list of improvements that should make the system better for the user. We divided this into two sections: small improvements that could be taken up within the scope of the project and larger improvements that might go beyond the current status of the project.

Small improvements

Interface related

1. The search bar should be made more prominent.
2. The Menu icon should be changed into a tab with the words 'Menu'.
3. When scrolling functionality is provided (e.g. at the News Feeds and Browse section), interfaces should be aware of the fact that buttons or information on top of the pages should remain visible during scrolling.
4. When scrolling the Browse section the button 'Fetch Random Topics' should stay on top.
5. The homepage needs to be reworked. However, the requirements for the homepage vary a lot.

User feedback

6. Logos on technique used to retrieve results can be enlarged.
7. There should be a visual validation when users press the 'Like' button.

Large improvements

1. The topics in the Browse section should be changed to reflect real-world topics.
2. The words in the 'Advanced Search' section should be reformulated into more understandable terms:
 - a. Metadata > Search the Description (cataloguing)
 - b. Speech > Search speech
 - c. Visual Categories + Visual Instances + Visual faces > Search visual
3. Users prefer the system to state it found 'No relevant results' instead of always showing results, even when they are not relevant.
4. To accommodate users' need to perform a known-item search, searching and sorting results on date could be included.
5. Allow searching and browsing on program title, genre, region and names of (famous) people.

Appendix D **WORKFLOW AXES HOME EVALUATION**

Introduction to participant:

- a) The BBC is part of an EU project developing tools to search television archives in new ways, it is called AXES.
- b) We have with us today a system in development and we would like members of the public to try it out and give us their impressions.
- c) We will take note of your verbal feedback as you go along, which will help us refine the system and bring improvements.
- d) Please note we are not testing you – rather we are evaluating this version of the AXES system with a wide range of people.
- e) We can stop at any time.
- d) The data we collect is anonymous and is only used for the purposes of this project.
- e) Introduction to the 'observer', who will be taking notes.

Evaluator:

- 1) Gives the introduction and answers any questions the participant may have during the session. (Time tbc)
- 2) Encourages participant to talk through experience as they go along.
- 3) Ensures the evaluator or observer fills out the online questionnaire.

Participant:

Is handed the tablet with AXES in active mode, set up with a registered user ID.

Observer:

Makes a record of observations and verbal feedback from the participant and completes the online questionnaire.

Evaluation steps

- Home Page Introduction: AXES home page is offering topics based on International days. As we can see, 21st November is World Television Day. Please scroll and take a look.
- Search Bar: existing search techniques, for example enter 'sheep'. It has returned results even if the content is not catalogued. It has searched the video for images of sheep, which is novel.
- Menu Options.
- Newsfeeds: gathered from different news sources for navigation into the content, therefore linking current content to archive content e.g. programs such as Doctor Who.
- Sort results: there are two types of view, thumbnail and list, please try both views.
- Browse-Topics function: it is also possible to browse the archive by random topics. Let's take a look. The topic is listed on the left, please scroll up/down and left/right and take a look. There is an added feature to fetch more random topics. Please have a go and access some archive content. As the results are returned, you can see new topics appearing. It is also possible to search for the most popular videos – please have a go.
- Advanced Search: it is possible to carry out advance searches. Let's take look and I will show you an option.
- Face search: e.g. Jeremy Clarkson or Bill Oddie. AXES has searched images on the internet for the person named and retrieved similar faces from the archive content. Please select one to play.
- Thank you, that has given you a quick overview of AXES system. We will now go through a short questionnaire.

Appendix E **SURVEY AXES HOME EVALUATION**

This appendix contains the survey on the general acceptance of the AXES HOME system. For the detailed results of this questionnaire, see chapter 3 and/or appendix F.

Survey Evaluation AXES Home

1. What is the user ID? (*For the interviewer*)
 (*Open*)
2. Did the respondent get an instruction on how to use AXES HOME? (*For the interviewer*)
 - A. Yes
 - B. No
3. Please indicate your age
 - 18-24
 - 25-34
 - 35-44
 - 45-59
 - 60+
4. What is your gender?
 - Male
 - Female
5. Do you often search online for video or television programs?
 - Very often
 - Often
 - Regularly
 - Sometimes
 - Never
6. (*Dependency Question 2A*)
 You were handed the iPad with an instruction on how to use the system.
 How easy or difficult was it to learn for you?
 - Very Easy
 - Easy
 - Neutral
 - Difficult
 - Very difficult

7. *(Dependency Question 2B)*

You were handed the iPad without any instruction on how to use the system.

How easy or difficult was it to learn for you?

- Very Easy
- Easy
- Neutral
- Difficult
- Very difficult

8. *(If answered difficult/very difficult in question 6 or 7)*

Why was it difficult for you?

(Open question)

9. How attractive was the user interface to you?

- Very attractive
- Attractive
- Neutral
- Unattractive
- Very unattractive

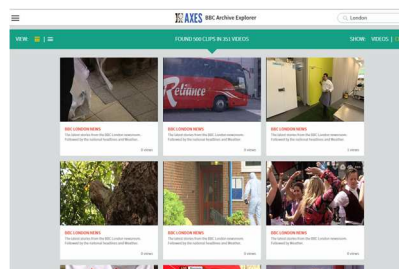
10. How relevant were the results shown when you performed a search via the search bar?



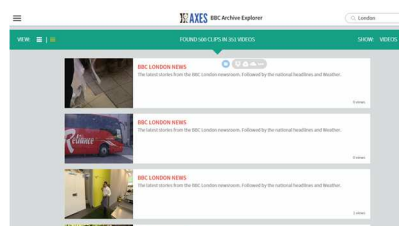
- Very relevant
- Relevant
- Neutral
- Irrelevant
- Totally irrelevant

11. The results of your search are shown in thumbnails or in a list. Which view do you prefer?

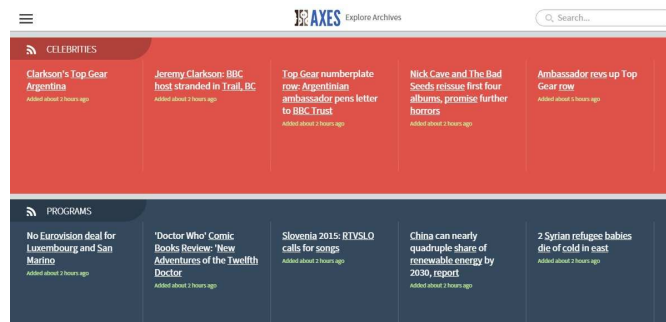
A. Thumbnails



B. List



12. How useful were the News feeds to find archive content?



- Very useful
- Fairly useful
- Neutral
- Hardly useful
- Not useful
- I haven't used it (> Go to question 14)

13. Do you agree with the following statement:

The news feed browser allows me to find interesting videos without the need for searching.

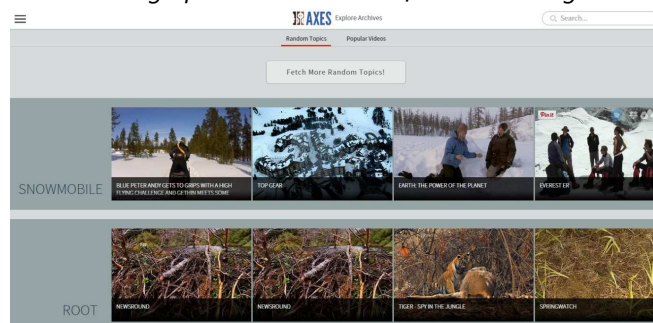
- Fully Agree
- Agree
- Neutral
- Disagree
- Fully disagree

14. How useful were the Browsing options?

- Very useful
- Fairly useful
- Neutral
- Hardly useful
- Not useful
- I haven't used it

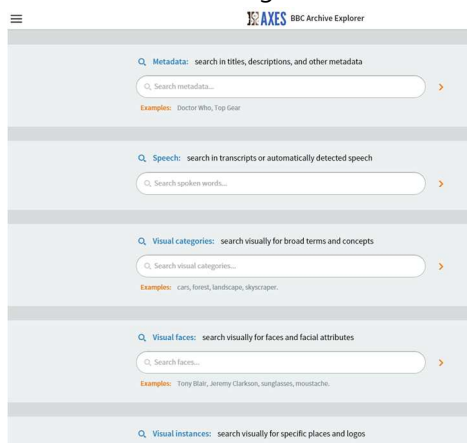
15. Do you agree with the following statement:

The Browsing options allow me to find interesting videos without the need for searching.



- Fully Agree
- Agree
- Neutral
- Disagree
- Fully disagree

16. Which of the following advanced search options did you use? (*multiple answers*)



- ☐ Metadata
- ☐ Speech
- ☐ Visual categories
- ☐ Visual faces
- ☐ Visual Instances
- ☐ I didn't use the advanced search options (> go to question 16)

17. Does the system produce useful results per advanced search option?

	Very relevant	Relevant	Neutral	Irrelevant	Totally irrelevant
Metadata (<i>depend 16.1</i>)					
Speech (<i>depend 16.2</i>)					
Visual Categories (<i>depend 16.3</i>)					
Visual Faces (<i>depend 16.4</i>)					
Visual Instances (<i>depend 16.5</i>)					

18. Which features would you like to have on the home page? (*multiple answers*)

- Recently viewed videos
- Popular videos
- Videos selected by an editor
- An instruction on how to use the system
- Other, namely ... (*Open*)

19. Are there things you did not like about AXES system?
(*Open*)

20. What did you like most about the AXES system and why?
(Open)
21. Would you use this system at home in your free time?
- Very likely > *Go to question 22*
 - Likely> *Go to question 22*
 - Neutral/Don't know> *Go to question 23*
 - Not likely > *Go to question 23*
 - Not at all> *Go to question 23*
22. Why would you be use the system?
(Open)
23. Why wouldn't you use the system?
(Open)
24. Do you have any other remarks about this system or other things you would like to share with us?
(Open)

Appendix F RESULTS AXES HOME EVALUATION

This appendix contains all detailed results from the user tests regarding the acceptance of the AXES HOME system (for the extensive associated survey, see appendix E).

1. Please indicate your age:

Answer	Frequency	Percentage
18-24	2	8%
25-34	7	28%
35-44	0	0%
45-59	7	28%
60+	9	36%
Total	25	100%

2. What is your gender?

Answer	Frequency	Percentage
Female	10	53%
Male	9	47%
Total	19	100%

3. Do you often search online for video or television programs?

Answer	Frequency	Percentage
Very often	7	29%
Often	4	17%
Regularly	9	38%
Sometimes	3	13%
Never	1	4%
Total	24	100%

4. You were handed the iPad with an instruction on how to use the system. How easy or difficult was it to learn for you?

Answer	Frequency	Percentage
Very easy	5	25%
Easy	10	50%
Neutral	5	25%
Difficult	0	0%
Very difficult	0	0%
Total	20	100%

5. How attractive was the user interface to you?

Answer	Frequency	Percentage
Very attractive	2	8%
Attractive	17	71%
Neutral	4	17%
Unattractive	1	4%
Very unattractive	0	0%
Total	24	100%

6. How relevant were the results shown when you performed a search via the search bar?

Answer	Frequency	Percentage
Very relevant	5	19%
Relevant	18	69%
Neutral	2	8%
Irrelevant	0	0%
Totally irrelevant	1	4%
Total	26	100%

7. The results of your search are shown in thumbnails or in a list. Which view do you prefer?

Answer	Frequency	Percentage
Thumbnails:	16	70%
List:	7	30%
Total	23	100%

8. How useful were the News feeds to find archive content?

Answer	Frequency	Percentage
Very useful	5	19%
Fairly useful	14	52%
Neutral	3	11%
Hardly useful	2	7%
Not useful	3	11%
I haven't used it	0	0%
Total	27	100%

9. Do you agree with the following statement: 'The news feed browser allows me to find interesting videos without the need for searching'.

Answer	Frequency	Percentage
Fully agree	5	19%
Agree	9	35%
Neutral	10	38%
Disagree	2	8%
Fully disagree	0	0%
Total	26	100%

10. How useful were the Browsing options?

Answer	Frequency	Percentage
Very useful	7	27%
Fairly useful	11	42%
Neutral	5	19%
Hardly useful	3	12%
Not useful	0	0%
I haven't used it	0	0%
Total	26	100%

11. Do you agree with the following statement: ‘The Browsing options allow me to find interesting videos without the need for searching’.

Answer	Frequency	Percentage
Fully agree	6	23%
Agree	15	58%
Neutral	2	8%
Disagree	1	4%
Fully disagree	2	8%
Total	26	100%

12. Which of the following advanced search options did you use? (multiple answers possible)

Answer	Frequency	Percentage
Metadata	8	30%
Speech	7	26%
Visual categories	5	19%
Visual faces	21	78%
Visual instances	4	15%
I didn't use the advanced search options	0	0%

13. Does the system produce useful results per advanced search option?

Answer	Very relevant	Relevant	Neutral	Irrelevant	Totally irrelevant	Total
Metadata	0	1	2	1	4	8
Speech	0	2	1	2	2	7
Visual categories	0	1	0	0	3	4
Visual faces	3	9	5	3	1	21
Visual instances	1	0	1	0	2	4

14. Which features would you like to have on the home page? (multiple answers possible)

Answer	Frequency	Percentage
Recently viewed videos	11	44%
Popular videos	17	68%
Videos selected by an editor	9	36%
An instruction on how to use the system	13	52%
Other, namely...	10	40%

Other, namely...	
<ul style="list-style-type: none"> ➤ Dates, items, organizations ➤ One random topic ➤ Menu of option visible ➤ Visual instruction, download ➤ List of personalized topics 	<ul style="list-style-type: none"> ➤ More specific to location or region or topic ➤ Text size larger and adjustable ➤ Date, time, new editions ➤ Breaking news ➤ Sport, better titles, year

15. Are there things you did not like about AXES system?

(NB. The answer transcripts from this question were carefully reviewed, after which homogeneous content categories were distinguished in the responses of the participants. Similar comments per task were structured and clustered).

No comment:
<ul style="list-style-type: none"> ➤ I did not use it enough to comment. ➤ I have not really played with it.
Certain specifics:
<ul style="list-style-type: none"> ➤ The search bar should be more prominent. Now there is a high risk of using the Google search bar instead, thus one is no longer using the AXES tool, but Google instead. ➤ Poor video quality. ➤ Autocorrect. ➤ Limited database content make the menu option more accessible. ➤ Too many results. ➤ Some of the results were not relevant. AXES is too slow and the interface is buggy, especially the key frames. The system must be easier to use than YouTube.
The amount of information:
<ul style="list-style-type: none"> ➤ It ought to be in an app rather than a browser. Although I could figure out the results, I need more info on why these results come up. If it's not sure if it's relevant, why does it come up with it: e.g. the dog is not a panda, give me info what the relationship is to be clear. Certainly for an advanced user. ➤ A bit random in places. ➤ I want to know why results are relevant, and what they're for, research or browsing. ➤ A bit vague, needs refining. ➤ Not enough information on what I am looking at. ➤ It has to be more specific. ➤ The massive amount of results to look through takes you off in all directions. Sometimes you want one specific topic.

16. What did you like most about the AXES system and why?

(NB. The answer transcripts from this question were carefully reviewed, after which homogeneous content categories were distinguished in the responses of the participants. Similar comments per task were structured and clustered).

General comments:

- Really focuses on images, not a lot of text is displayed.
- Handy tool for researchers and teachers.
- AXES is more about iPad, instant, quick and touch.
- I like the lay out and the different search options.
- You can find unrelated, unexpected clips, which is fun.
- The fact that you can find videos on tv programmes.
- Good visuals.

Certain specifics:

- Transcript search, visual search, faces and other visual elements.
- The transcript search is very useful. It provides new ways to explore the collection.
- Very impressed by the face searching and the object searching. That's brilliant. Tradeoff: the output. I can live with the result images which do not match, because they are compensated by all the other relevant ones. To be able to find the image in the collection I want through that click, that's the needle in the haystack. It's acceptable that I am also presented other images that are false hits.
- The results are spread out in categories, the look is presentable.
- Advanced search works perfectly.
- There are plenty results to choose from, which makes it perfect.
- The ability to watch the programmes at your own time at the end of a day.
- All the different ways you could research something were great, as well as the speed of the system.

Ease of the system:

- The way the system is organized: simple, interesting content and easy access.
- It's easy to use, intuitive, and has a familiar layout like YouTube.
- Another easy way of getting information and viewing programmes: clips are downloadable for when no wifi is available, possibility to connect to tv for parties and games.
- The public access thumbnail navigation is reasonably easy.
- The newsfeed makes it easy to find topics.
- The system is easy to navigate.

17. Would you use this system at home in your free time?

Answer	Frequency	Percentage
Very likely	10	40%
Likely	7	28%
Neutral/Don't know	7	28%
Not likely	1	4%
Not at all likely	0	0%
Total	25	100%

18. Why would you use the system?

'Very likely' or 'Likely':

- AXES offers a lot to explore.
- Because I am a curious person. I like to do a bit of research on the news: I go on the internet to find out some additional info on what I've heard. I would also use AXES to find programs I'd like to watch.
- To get to archive material from different access points.
- If it were available as an iPad app, I would use it to search the archive: I have an interest in broadcast-related material.
- To look for videos on tv programmes.
- For entertainment purposes.
- The system is easy to use and organized.
- To find archive programmes and additional information.
- To check the news and to look at tv programmes I want to catch up on.
- AXES provides something different, I have never seen any system like this before. It's also easy to find videos.
- Mainly for education.
- To get to see something new, off the beaten track.
- Because the content is so vast.
- Films, sport: another variant for looking up stuff. AXES is a useful addition to the BBC website.
- I think this system is easier for us to follow. AXES is a useful thing to have, e.g. to find clips that you have recently viewed. Certainly better than watching adverts.
- I want to find out more about different issues, broaden my horizons. I am looking for objective and efficient journalism.

19. Why wouldn't you use the system?

'Neutral', 'Not likely' or 'Not at all likely':

- Not clear what to use it for.
- No time, I would only use AXES on my phone.
- Access to other content.
- I would only use it for specific research.
- Cumbersome interface.

20. Do you have any other remarks about this system or other things you would like to share with us?

Comments:

- Is it open, can I use it?
- You can double the license fees if I can get this. So, great!
- Adding an instruction would be good.
- I'd like to access the system now!
- The interface needs improving: topics in news e.g. the arts and culture section load too slow. I lost interest after a couple of seconds. Also, the results are too noisy, not targeted enough.
- Would be nice to try it out at home.
- Looks too blockish and squash, needs more softness.
- Homepage needs redesign, it doesn't tell you anything. System is too slow.
- I think it is a very useful idea.
- I would like to browse through documentaries and music concerts.
- I can't wait until old archive materials come available.
- A good idea in theory.



AXES
ACCESS TO AUDIOVISUAL ARCHIVES

Explore the archive...



... on a **tablet!**



TOOLS OFFERING
NEW ENGAGING WAYS TO INTERACT
WITH AUDIOVISUAL LIBRARIES.



January 2011 - December 2014 | FP7 - IP project 269980 | Digital Libraries and Digital Preservation
Project partially funded by the European Commission.

CONTACT
Philippe Rohou | AXES project coordinator ERCIM | www.ercim.eu | E-mail: philippe.rohou@ercim.eu
Tinne Tuytelaars | AXES scientific coordinator Katholieke Universiteit Leuven | www.esat.kuleuven.be/psi/visics
E-mail: tinne.tuytelaars@esat.kuleuven.be

WWW.AXES-PROJECT.EU

Project Partners:

