



Project Document Cover Sheet

Project Information			
Project Acronym	CAVA		
Project Title	Human Communication: an Audio-Visual Archive for UCL		
Start Date	01 April 2009	End Date	31 March 2010
Lead Institution	UCL		
Project Director	Paul Ayris		
Project Manager & contact details	Martin Moyle m.moyle@ucl.ac.uk / 020 7679 4351		
Partner Institutions	UKDA		
Project Web URL	http://www.ucl.ac.uk/ls/cava		
Programme Name (and number)	Capital Programme - Repositories: start-up		
Programme Manager	Amber Thomas		

Document Name			
Document Title	Project Plan		
Reporting Period			
Author(s) & project role	Martin Moyle, Project Manager		
Date	28 April 2009	Filename	
URL	<i>if document is posted on project web site</i>		
Access	<input type="checkbox"/> Project and JISC internal		

Document History		
Version	Date	Comments
0a	28 April 2009	First draft for PM approval



JISC Project Plan – Customised for Repository Start-Up

Overview of Project

1. Background

CAVA will establish a repository for UCL's audio-visual data on real-life human communication for spoken and signed languages. Historically, the study of communication has been based on highly-controlled experimental data, but a better understanding comes from examining natural audio-visual data. Such work, both qualitative and quantitative, involves in-depth study of video- and audio-recorded data of conversations and clinical encounters. The data mainly comprises videotaped interactions (conversations, interviews, assessments) between a person who has atypical communication (due to disabilities such as stroke, deafness, autism etc) and their spouse, teacher, parent or another 'typical' communicator, filmed in the home, at school or in the clinic. The duration of the videos ranges from 10 minutes to more than an hour per client. Some of the data is longitudinal with regular sessions filmed over a period of time.

Despite the advances of recent years in online resources for academics, there is no centralised data archive to support this work. It is not uncommon for vital and unique data to languish on VHS videotape in personal collections; researchers across the discipline waste time battling with increasingly inaccessible media and finding individual solutions to the challenges of editing and analysis. The resources of funders, researchers and subjects are wasted on the collection of new data rather than on the re-use of existing data.

CAVA will create a repository of re-usable video material to support the work of UCL and of the international human communication research community. The UCL team already holds a large body of content, in suitable formats, with appropriate permissions secured. Each video will be accompanied by a catalogue record based on a discipline-specific descriptive standard, IMDI¹. Transcripts and other supporting material will also be deposited wherever available. Access to the video content of the repository will be restricted to bona fide researchers, and the project will develop procedures and processes for managing access in accordance with subject consent. The CAVA repository will be housed within the UCL Library Services Digital Collections² service, which uses the Ex Libris DigiTool repository platform. A minimum of 600 hours of catalogued video will be made available during the lifetime of the project. The repository will be an international resource and will be publicised to the community of human communication researchers worldwide.

The project will also explore the potential for the collecting scope of the repository to expand to encompass the deposit of relevant material by non-UCL researchers, and it will investigate options for the managed archival storage of uncompressed master files in support of long-term digital preservation of the content of the CAVA repository.

There are no IPR obstacles to the deposit and distribution, to genuine researchers, of the recordings which it is proposed to place into the repository. The JISC funding seeding the CAVA project is primarily to be used to provide a staffing resource to accelerate the development of CAVA repository - in particular, to allow metadata creation, and to assist UCL in publicising the repository and in overseeing access to the content.

¹ IMDI: ISLE Meta Data Initiative. See <http://www.mpi.nl/IMDI/>

² UCL Digital Collections: <http://digital-collections.lib.ucl.ac.uk>

2. Aims and Objectives

The project will establish the CAVA repository for rights-cleared audio-visual material owned by UCL researchers. Specifically, the aims are as follows:

- To configure the CAVA repository as a discrete archive within the UCL Digital Collections service.
- To implement support for the IMDI metadata standard to an appropriate level of detail, and to ensure metadata interoperability between the CAVA repository and Dublin Core-based aggregation services.
- To populate the CAVA repository with a minimum of 600 hours of rights-cleared audio and video material in agreed dissemination formats, accompanied by transcripts and other supporting material where available.
- To ensure that the ongoing population of the CAVA repository is embedded within the relevant UCL research communities.
- To implement processes and technical procedures to support the management of access to the content of the repository.
- To publicise the CAVA resource to potential users and other stakeholders.
- To investigate the feasibility of the expansion of the remit of the repository to encompass deposit by non-UCL researchers in appropriate disciplines.
- To appraise the options for the managed long-term storage of uncompressed master files, to support digital preservation.

CAVA through these overall aims therefore addresses Call challenges (i) – increasing repository deposit, and (iv) – policy frameworks, specifically relating to subject consent and access management.

3. Overall Approach

The CAVA repository will be established to accommodate UCL's human communication digital video material, using the UCL Library Services DigiTool platform. The work will be delivered through 7 Work Packages, detailed at Appendix B and summarised here.

In a Repository Configuration Work Package, WP2, the work necessary to prepare the CAVA repository on top of the DigiTool platform will be undertaken. This will include an analysis of storage requirements, in terms of both volume and configuration. An analysis of the IMDI schema will be undertaken, and an IMDI subset appropriate for the description of the UCL human communication video corpus will be identified and implemented in DigiTool. A mapping between UCL's IMDI instance and Dublin Core will be written, and the DigiTool OAI Data Provider will be configured to support the exposure of the CAVA metadata records. The front-end of the repository – 'look and feel', search, browse, and mechanisms for the viewing and delivery of video and transcript material - will also be configured and tested under this Work Package.

WP3, Repository Population and Exploitation, will see 'real' video material uploaded to the repository. Some preliminary work will be carried out to determine preferred dissemination formats. Where material currently exists only in large, uncompressed file formats, such as AVI, conversion to at least one dissemination format will be carried out. Each resource will also be catalogued to the IMDI set agreed in WP2. Under WP3 the team will devise and test ingest processes so that video clips, transcripts (where available) and descriptive metadata can be uploaded to the repository in batches, in a way which maintains the relationships between the one or more versions of each video recording, its transcript, and the metadata which applies to each. The final ingest process will include the automatic generation of technical metadata and the creation of appropriate access restrictions.

Having established a satisfactory batch ingest process, the team will explore the enabling of a full "self-archiving" (that is, unmediated) model to enable UCL researchers to deposit material as it is collected in future.

Two sets of guides, one for depositors, and one aimed at end-users, will be written: these will be mounted on the project Web site and will also be available via the CAVA repository interface.

Project Acronym: CAVA
Version: 0a
Contact: Martin Moyle m.moyle@ucl.ac.uk
Date: April 2009

Towards the end of the project, a pilot will be undertaken to deliver content from the repository (where appropriate permissions are in place) directly through on-line courses in the UCL VLE, Moodle.

Access Management will be approached under WP4. The UCL team has full rights to place material in the CAVA repository and to make it available to researchers. Because of the sensitivity of the material, open access is neither possible nor desirable. WP4 will deliver the procedures and technical processes necessary to manage access to the content. The project team will design the application process for prospective users, and implement workflows through which requests are to be verified and authorised, and users registered and authenticated. The retention and presentation of rights information within the IMDI record, and the content of any hard-copy or click-through licensing agreements to be associated with particular categories of object, to indicate clearly and unambiguously what an authorised user may and may not do with the material, will be designed and implemented.

In the detailed planning for WP4, the team will draw on the advice and experience of the UKDA.

In WP5, Sustainability, CAVA will explore various issues relating to the long-term future of the repository. Firstly, the project will build on UCL's work on the standardisation of consent forms to draw up a Deposit Policy, for adoption by the UCL Centre for Human Communication, to ensure that all UCL human communication researchers submit copies of audio-visual data collected in future research projects to the CAVA repository. The UCL team will collaborate with UKDA in defining this policy, to ensure its close alignment with the data policies of research funders. Secondly, the costs and other implications of the extension of deposit privileges to researchers in appropriate disciplines from outside UCL will be examined, and a short feasibility report will be prepared. Finally, options will be appraised for the long-term curation of the uncompressed master files, the size of which (estimated at over 8 Terabytes for the 600 hours of initial CAVA content) makes it impracticable that they should be housed in the DigiTool repository, at least at the outset of the project. By default, the master files will be stored off-line at UCL. The partners will investigate potentially more suitable arrangements for these files, including the feasibility of depositing them with UKDA, to ensure the long-term preservation of their intellectual content.

The outputs from WP5 would be a formal Deposit Policy, and short feasibility reports (with, if appropriate, outline action plans), relating to:

- (i) opening up of repository deposit to non-UCL researchers in relevant fields.
- (ii) options for managed curation of archival copies, including possibility of deposit with UKDA.

The project team will be committed to disseminating the work of the project and its outputs, and dissemination will be managed under WP6: Dissemination. The work carried out by CAVA will be of interest to a number of stakeholders, including human communication science researchers internationally, and the global repository community. Opportunities for outreach to researchers will be sought, through mailing lists and at least one researcher-orientated conference, such as ICPLA (the Conference of the International Clinical Phonetics and Linguistics Association). Advocacy to depositors will also be managed under this Work Package, which will also be supported by the guides to be created under WP3. Opportunities for the publication of at least one paper in the human communication research literature on the genesis and development of the CAVA repository will be sought. A CAVA Web site will be created, to share information about the project. It will provide links to project reports and to copies of all publications and presentations deriving from the project. Such material will also be deposited in UCL's open access research repository, UCL Eprints, wherever applicable. A Dissemination Plan is included in this Project Plan at Section 18.

In WP7, Building Coherence, the JISC requirements on community membership will be addressed. The project overall will be managed by WP1: Project Management.

4. Project Outputs

- Analysis and implementation of IMDI metadata set to support a repository of audio-visual material.

- Availability of metadata for export, in a Dublin Core mapping, through an OAI gateway.
- Repository configuration: look and feel, discovery, and delivery.
- Analysis and implementation of appropriate file formats for the dissemination of video material in this field.
- On-line guides for users and depositors.
- Access management procedures and processes, relating to application, authorisation, registration, authentication and licensing, drawn up with the support of a UK Data Centre.
- A Deposit Policy, sponsored by the UCL Centre for Human Communication and applying to all UCL's human communication researchers.
- Pilot integration with the UCL VLE (Moodle).
- An options appraisal for the storage and preservation of uncompressed master files.
- Dissemination to prospective users and other stakeholders.
- Project Web site.

5. Project Outcomes

- Creation of a significant new communication research resource available to all researchers in the discipline.
- A precedent for the re-use of video data in this discipline.
- A precedent for the wide dissemination of "raw" data in this field, supporting different analytical methodologies without constraint.
- Expertise in the managed curation of video materials to be shared with other research communities, across disciplines.
- Opportunities for new interactions between video data and published research outputs: for instance, hyperlinks in citation lists.
- The CAVA Deposit Policy, in conjunction with the standardised consent forms already in use at UCL, will serve as a possible model for other research fields and other repository environments.

6. Stakeholder Analysis

Group	Interest	Value
UCL researchers	Greater exposure for work; greater security for and longevity of work.	High
Human communication researchers (worldwide)	Increased access to re-usable research, supported by discipline-specific cataloguing standards and researcher-orientated navigation.	High
Research funding agencies	Better compliance with data policies: greater value for money from grants awarded through an increase in visibility and re-usability of collected data.	Medium
UCL taught course students	Improved access to reliable, up-to-date research videos, a hands-on introduction to contemporary research methods.	Medium
Repositories community	Few exemplars of audio-visual repositories at present; learning outcomes of CAVA will be of wide interest.	Medium

7. Risk Analysis

Risk	P	S	PS	Management of risk
Research community fails to engage with CAVA.	3	4	12	Project will be widely publicised - CAVA has a dissemination WP. Evidence points to strong need for CAVA and similar undertakings within community.
UCL depositors fail to contribute.	2	5	10	Participating departments have signed letters of

				support and are committed to CAVA. CHC is committed to implementing Deposit Policy as a project output.
Project Officer leaves in course of project.	2	3	6	Recruitment. UCL Library Services will maintain CAVA repository services and continue to meet its obligations to project.
IPR Infringement	1	5	5	Unlikely because of stringent procedures already in place. All prospective content is IPR-cleared. Should complaint be received, access to digital objects in question will be suspended while rapid investigation takes place.
Technology failure undermines project aims	1	4	4	Unlikely - pilot work has already taken place - but support from Ex Libris (vendors of DigiTool platform) will be available if required.

8. Standards

Name of standard or specification	Version	Notes
IMDI	3.0.4	Local MD format
OAI-PMH	2	
Dublin Core		Export MD format
VIDEOMD		http://lcweb2.loc.gov/mets/Schemas/VMD.xsd technical metadata for video content

9. Technical Development

CAVA will use off-the-shelf software, which is already in production at UCL. The technical work of the project is primarily configurational, rather than developmental.

10. Intellectual Property Rights

The speakers whose communication is captured in these materials have given full informed consent for archiving and for the data to be accessed and used for future research and teaching by bona fide researchers with the appropriate authorisations. The material which is to be placed in the CAVA repository has been subject to rigorous data collection processes, which take full account of issues of copyright, data protection, and medical ethics and parental consent.

Copyright is held either by the researcher who collected the data, or by UCL. The UCL consent form requests participants (or their parent/s or carer/s) to assign the entire right, title and interest of the copyright existing in the data, with a Schedule of Detail about the type of data collected (for example assessments, interviews, conversations) .

All projects conform to the Data Protection Act (1998), and are registered with the UCL Data Protection Officer.³

All projects whose data will be deposited have been approved by an institutional research ethics committee (in the case of unimpaired speakers) or an NHS research ethics committee (in the case of speakers with disordered communication who are NHS patients or children in education). Where children are deemed too young to give informed consent to filming, and to the subsequent re-use of that video data – under strict access conditions - for research and teaching, this is obtained from

³ UCL Data Protection Policy: <http://www.ucl.ac.uk/efd/recordsoffice/policy/data-protection/>

Project Acronym: CAVA
Version: 0a
Contact: Martin Moyle m.moyle@ucl.ac.uk
Date: April 2009

parent/s or carer/s.⁴ At the age of 18, retrospective informed consent is sought from the subjects themselves, in line with NHS guidelines.

All published material produced by the project will follow current JISC guidelines for accessibility. Copyright of the appropriate author(s) will be asserted, but all such material will be covered by the appropriate Creative Commons licence allowing free non-commercial reuse.

Project Resources

11. Project Partners

CAVA has two partners:

- UCL
- UK Data Archive (unfunded – advisory role)

12. Project Management

Project Management has a Work Package – see below, WP1. This will help to ensure that all the Work Packages of the project are managed effectively and that all the project outputs are delivered to time and within budget. The Project Manager anticipates spending a minimum of 3 hours/week on CAVA, with responsibilities including project management, staff supervision and dissemination.

A Steering Group will be convened to oversee the development of the project. It will meet twice in the course of the project. Membership of the CAVA Steering Group will be as follows:

- Dr Paul Ayris (Chair), Director of UCL Library Services and UCL Copyright Officer
- Dr Suzanne Beeke, Head of Department, UCL Language and Communication
- Dr Libby Bishop, Manager, ESDS Qualidata, UKDA
- Dr Merle Mahon, Senior Lecturer, UCL Developmental Science
- Martin Moyle, Digital Curation Manager, UCL Library Services; Project Manager
- Stevie Russell, Site Librarian, UCL Language and Speech Science Library
- CAVA Project Officer, to be appointed

The role of the Steering Group will be to monitor, review, assess and sign off the outputs of the project, to sponsor all evaluation and dissemination activity, and to help to shape the work of the CAVA Project Officer.

Drs. Beeke and Mahon, and the Project Manager, will share responsibility for the day to day progress of the project, meeting fortnightly and communicating as and when by email.

13. Programme Support

No immediate needs.

14. Budget

See Appendix A. No change to originally proposed budget.

Detailed Project Planning

15. Workpackages

See Appendix B.

⁴ For the relevant guidelines, see UCL policy at <http://www.grad.ucl.ac.uk/ethics/forms/guidance1.pdf> and the MRC Ethics Guide at http://www.grad.ucl.ac.uk/ethics/MRC_framework.php

16. Evaluation Plan

Timing	Factor to Evaluate	Questions to Address	Method(s)	Measure of Success
Formative / summative	Progress against plan	Is project on time and on budget?	Project Manager to monitor plan, reporting to Steering Group and JISC PM.	Milestones achieved; final report published
Formative	Outreach to depositors	Is newly-created eligible material being deposited to CAVA?	Academic members of Project Team will monitor relevant UCL research projects	Eligible material is deposited to CAVA
Formative	Outreach to end-users	Is research community aware of CAVA resource?	Monitor repository usage	Stats show increasing usage, from a growing number of sources
Summative	Lessons learned	Are lessons learned in course of CAVA shared with repository community?	Lessons learned to be logged	Lessons learned published in openly available final report

17. Quality Plan

Output	Repository configuration				
Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
3	Fitness for purpose	Acceptance testing	Architecture is accepted	Project Manager	

Output	Metadata interoperability				
Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
6	Validity	External OAI-XML validation software	Validation tests passed	Project Officer	

Output	Ease of discovery and retrieval				
Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
3-6	Fitness for purpose	Feedback from UCL researchers	Repository meets researcher requirements, as far as is practicable	Project Officer	

Output	Access management procedures				
Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
5-6	Validity	Assessment by UKDA	Approval by UKDA and Steering Group	Project Officer	

Output	Final report				
Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
11	Fitness for purpose	Steering Group to approve	Steering Group approves final report	Project Manager	

18. Dissemination Plan

Timing	Dissemination Activity	Audience	Purpose	Key Message
1-12	Web site	All	Awareness, promote, engage, inform	Information about project; ready access to outputs
1-12	Conference papers and presentations in the fields of digital repositories and human communications research.	All stakeholder communities	Awareness, engage, promote, inform	Information about project, findings and outputs
4-12	Local outreach	UCL content holders	Awareness, promote, engage	Advantages of deposit; how to deposit
6-12	Publication in human communication research literature	Library and publishing communities	Engage, promote, inform	Information about project origins, intentions and progress
10-12	Publish preservation policy	Repository community	Building coherence	Exemplar policy
10-12	Add Building Coherence web page	Repository community	Building coherence	Collation of all policy information relevant to wider repository integration .

19. Exit and Sustainability Plans

Project Outputs	Action for Take-up & Embedding	Action for Exit
Project reports (both final and incremental (eg preservation, access management))	Dissemination	Will be publicly available for at least 3 years post-project to support dissemination plan.

Project Acronym: CAVA
Version: 0a
Contact: Martin Moyle m.moyle@ucl.ac.uk
Date: April 2009

Project Outputs	Why Sustainable	Scenarios for Taking Forward	Issues to Address
CAVA repository	Known demand for re-use of raw video data in the human communications research field	Access administration by UCL academic departments, publicity by UCL academic departments, repository management by Library	New resources may be required (of UCL) if repository success exceeds expectations. Long-term preservation of master videos (WP5 will begin to address)

Project Acronym: CAVA
 Version: 0a
 Contact: Martin Moyle m.moyle@ucl.ac.uk
 Date: April 2009

Appendix B. Workpackages

Project start date: 01 April 2009
 Project completion date: 31 May 2010
 Duration: 12 months

Project Team:
 PM - Project Manager
 PO - Project Officer
 AP - Academic partners
 SG - Steering Group

Workpackage and activity	Earliest/latest dates (project months)	Outputs	Milestones (where appropriate)	Responsible
WORKPACKAGE 1:				
Project Management				
1. Produce Project Plan	1-2	Detailed plan	Plan accepted by JISC	PM
2. Consortium Agreement	2-3	Signed formal consortium documentation		PM
3. Steering Group 1	3	Meeting		PO
4. Steering Group 2	8-9	Meeting		PO
5. Manage budget	2-12	Budget monitoring	Budget reports (SG and JISC)	PM
6. Evaluation activity as per Evaluation Plan	3-12	Evaluation activity	Evaluation incorporated in reports to SG and JISC	PM
7. JISC progress report	6	Interim report	Report delivered to JISC	PM
8. Draft final report	11	Draft report		PM

9. Final report	12	Final report	Final report delivered to JISC	PM
10. JISC completion report	12	Completion report	Completion report delivered to JISC	PM
WORKPACKAGE 2:				
Repository configuration				
11. Analyse storage requirements	2-3	Storage rules for CAVA files		PM, PO
12. IMDI analysis and implementation	3-4	Initial IMDI profile for CAVA's descriptive MD		AP, PO
13. Interface design and testing	2-4	Interface to CAVA repository	Repository is publicly accessible	PO
14. Create and test OAI metadata export	4-6	Valid OAI_DC output	Validated OAI responses	PM
WORKPACKAGE 3:				
Repository population and exploitation				
15. Agree dissemination formats	2	Agreement on initial access formats		PO
16. Catalogue repository content	2-12	Descriptive records created		PO
17. Devise and test ingest procedures	2-3	Processes for batch upload		PO, PM
18. Upload	4-12	Repository content uploaded	Repository is receiving content regularly; it is retrievable through CAVA interface	PO
19. Prepare guides for depositors and end-users	6-8	Online introductory / training documentation	Guides available through Web site	PO

20. Investigate feasibility of non-mediated deposit process	8-9	Agreement on feasibility and desirability of non-mediated deposit to CAVA		PM
21. Pilot delivery of CAVA content through UCL VLE	8-9	Demonstration of CAVA-VLE interaction	VLE-CAVA linkage is demonstrated	PO
22. Embed CAVA repository search functions in CAVA Web site	10-12	Enhanced discovery	CAVA Web site has repository search functions	PM
WORKPACKAGE 4: Access management				
23. Design access management procedures	4-6	Access management procedures		PO, AP, PM
24. Implement support for agreed access management procedures in CAVA repository	7	Repository supports for agreed access procedures	UKDA-approved AM processes are supported by CAVA repository	PM
WORKPACKAGE 5: Sustainability				
25. Options appraisal for curation of uncompressed master files	7-10	Report on long-term preservation	Report available through Web site	PM
26. Deposit Policy for relevant UCL research groups	8-10	Policy	Policy available through Web site	AP
27. Feasibility of repository deposit by non-UCL human communication researchers	8-10	Short report to SG on likely demand and resource implications		PO, PM
WORKPACKAGE 6:				

Project Acronym: CAVA
Version: 0a
Contact: Martin Moyle m.moyle@ucl.ac.uk
Date: April 2009

Dissemination				
28. Dissemination Plan	1-2	Plan	Dissemination Plan incorporated in Project Plan	PM
29. Ongoing dissemination activities	1-12	Various dissemination activities		PO, AP, PM
WORKPACKAGE 7:				
Building coherence				
30. Preservation policy for dissemination formats	10	Preservation Policy		PM
31. Build usage data reporting	9-11	Regular reports on usage, available (as a minimum) to Project Team and JISC		PM
32. Ensure item-level metadata is included in action 23	10	Item-level rights metadata in repository		PM
33. Ensure metadata policy included in action 13	10	Metadata policy		PM
34. Collate outputs from actions 29-33 into Building Coherence web page	10-12		Building coherence web page publicly available	PM