Cultural and scientific heritage production, archiving and sharing.

Example: The Program “Audiovisual Research Archive in Social Sciences and Humanities” (ARA)

Prof. DDr. PETER STOCKINGER

Institut National des Langues et Civilisations Orientales (INALCO), Paris

Equipe Sémiotique Cognitive et Nouveaux Médias (ESCoM)
Fondation Maison des Sciences de l’Homme (FMSH)

Trudeau Foundation – Summer Institute 2009
Kingston – Gananoque, Ontario, Canada

21st of May 2009
- The general context -
Launched in 2001/02, the Audiovisual Research Archive (ARA) program intends to:

1. to “capture” two main facets in scientific and cultural heritage:
   - the (scientific, social, …) “discourse”, the “speaking about” the world in form of interviews, conferences, workshops, etc.
   - “extracts”, examples, “illustrations” of the “object” of this discourse in form of audiovisual corpora dedicated to specific domains, themes,

2. to make this heritage available to any interested person or community;

3. to investigate and to develop the scientific, methodological and technical means enabling any institution, even any individual, to create and diffuse his/her own audiovisual resources in form of online libraries or archives;

4. to investigate the uses and exploitations/appropriations of AV data in given social contexts;

5. to investigate also new, “future” forms of knowledge capturing and appropriation.
The ARA programme actually is:

- A monthly growing corpus of digital audiovisual resources of actually some 5500 hours of online videos;
- A set of online video portals composed of thematically restricted online media libraries, specialised collections or republished videos (in form of thematic folders, bilingual folders, etc.);
- A complete digital working environment for the building, management and diffusion of online digital audiovisual archives ("ESCoM Suite 2009");
- A digital video production and publishing service;
- A computer service focussed on digital video management and processing;
- A R&D programme implied since 1996 in a variety of EC and French R&D projects (cultural heritage, digital libraries, e-learning, etc.);
- An interdisciplinary team of 12 to 20 persons in Paris and "abroad" and a small network of partner institutions.
General context motivating an initiative such as the ARA program:

- the “knowledge turn” in world economy
- the “semiotic turn” in digital archives and libraries
- i.e. the collection, constitution, diffusion, preservation and exploitation of knowledge through digital means

Example: Europe and Lisbon 2000 policy – the “European knowledge space” viewed and intended as:

- on the horizon of 2010, to become a world leading space in knowledge economy
- a common learning space appropriate to fulfil this objective (Bologna, Lisbon, Nice, Barcelona)
1st topic
The general context

- “Common European knowledge & learning space”

  - 8 specific areas of development (skills for knowledge society, teaching teachers, foreign languages, learning environment, learning attractiveness, mobility, …)

  - central perspective = LLL (Life Long Learning):
    - special issues: adult education, vocational training, “secondary alphabetisation”, science/society, (intercultural, artistic, …) sensitisation, …
    - challenge: promote employability

  - specific European-wide programmes and projects: framework programme, Leonardo, Socrates, Comenius, e-Culture, …

  - Principal requirements:
    - digital libraries,
    - learning spaces (broadly speaking)
Requirements for a European Learning space:

1/ Multimedia digital libraries: from digital objects to learning objects.

Challenges:

- interconnecting digital libraries, archives, information spaces, …
- “subjective authoring” of digital objects (texts, films, …) as pedagogical ones
- interoperability and standards: DC; LOM; MPEG7, …
- implementation of (multilingual) thesauri, terminologies, ontologies, …

Projects & Initiatives:

- The i2010 Digital Libraries Initiative (“Europe’s cultural and scientific heritage at a click of a mouse”): Europeana portal
- Delos Network of Excellence on digital libraries;
- Minerva a Thematic Network on Digitising Content.
2/ Virtual Campuses and (personal) learning spaces

- Learners or groups of learners should have the opportunity to create and manage their own learning spaces provided with personalised e-content, web services, tools, communication possibilities, … and also counselling and guidance facilities.

- Learning spaces are considered to be closely connected with digital libraries.
Some related on-line references:

- Peter Stockinger. Digital audiovisual archives in humanities. Problems and challenges. *Workshop on Multimedia in Digital Libraries*, Jun 2003, Chania, Greece. <hal-00130295>


- Peter Stockinger. La réalisation d’un entretien avec un chercheur invité dans le cadre du Programme des Archives Audiovisuelles de la Recherche (AAR). 2002. <hal-01304808>

- Peter Stockinger. Le Programme « Archives Audiovisuelles de la Recherche » : La génération d’archives audiovisuelles en ligne de la recherche.. [Rapport de recherche] ESCoM-AAR (Equipe Sémiotique Cognitive et Nouveaux Médias - Archives Audiovisuelles de la Recherche); INALCO. 2004, pp.63. <hal-01303666>
- A short presentation of the ARA digital data and web portal -
2nd topic
The ARA corpus and web portal

- **ARA corpus**: actually more than 5500 hours of videos and other digital resources.

- **Examples**: corpora in social history (900 hours); social and cultural anthropology (500 hours); language studies and linguistics (500 hours), archaeology (350 hours), urbanism (250 hours), …

- **Genres of online videos**: scientific events (interviews, conferences, seminars, workshops, …), cultural manifestations (exhibitions, concerts, performances, …), “terrain” (specific practices and metiers, “road movies, …), reportages, documentaries and audiovisual documentations

- **Languages**: videos available in 16 languages, French principal language

- **Authors**: International community of about 2100 people from 85 countries: scholars, writers, artists, politicians, people working in NGO, field workers, students, “ordinary witness”, …

- **Visitors**: 125 000 visitors from 170 countries since beginning of January 2009 (45% from France)
Specific points concerning the ARA corpus

- it is an open corpus: every month new contents produced by ESCoM and/or consigned by partner institutions or individuals to ESCoM;

- the selection of AV resources and production activities is decided cooperatively by a steering committee;

  - a significant part of filming of scientific events is determined through the general policies of FMSH;

  - ESCoM itself has an own production activity dedicated to specific knowledge domains and financed through special budget lines coming from EC and French projects, sponsoring, etc.

  - a more and more important part of the production is the result of individual or group initiatives, i.e. of people wanting to use the ARA publishing and diffusion environment for their own productions.
2nd topic
The ARA corpus and web portal

- Specific points concerning the ARA corpus

- All videos can be freely accessed and used in non-commercial contexts; ARA videos are licensed through “Creative Commons”;

- The authors of the videos keep their whole rights on the content, every author and every participant sign a contract permitting ESCoM/ARA to diffuse on their servers the audiovisual content;

- Each participant (author, filmed person, …) is informed about the uses of the filmic material diffused on the ARA servers (especially about the republishing/reuse of their content);

- Before the definitive publishing of a video, the author and – as far as possible – the concerned participants (filmed persons, …) are asked for to verify online the production and to give the final “green light” for diffusion.
Figure 1 shows the **home page** of the web site hosting the ARA audiovisual and multimedia resources.
3rd topic

- The production and publishing activities -
Three different types of sectors of activities, each one corresponds to a specific workflow:

1. The building and maintaining of an archive and a corresponding web portal

2. The "standard" production and publishing process of audiovisual resources documenting a knowledge domain, a scientific event, etc

3. The republishing process of already existing audiovisual resources
Production/publishing process of audiovisual resources: Complex process of which the “standard production/publishing workflow is the following one

1. Definition and preparation of a “terrain” (scenario for filming, management of all resources, participants included, technical and legal aspects, …)

2. Filming of the terrain (production of rushes, “real time” scripting, information & resources gathering, …)

3. Constitution of the corpus (selection of gathered/produced material, preparation, registration, …)

4. Basic processing of the corpus (technical processing; montage, postproduction, thematic indexing, translation/adaptation)

5. Publishing of the corpus (via standard publishing models)

6. Physical, digital, legal archiving and other related activities.
- The video repurposing challenge or how to make « useful » stored digital data -
List of challenges that the audiovisual and multimedia resources of the ARA are facing:

- transform the AV material in a useful tool for formal (secondary, higher) education
- transform the AV material in a useful tool for informal learning (life long learning, u-learning, vocational training, …)
- open the AV material to a multilingual and multicultural knowledge market
- use the AV material for other objective bounded purposes: political communication, social communication, …
- re-use the material in order to contribute to issues in social learning, LLL (adult learning, community training, vocational training, secondary alphabetisation, intercultural sensitisation, etc.)
- prepare the material for its critical use in the construction of new knowledge.
This means especially:

- the re-authoring or again the repurposing of a given source corpus

- with respect to given communication objectives, user groups and exploitation contexts

- as well as with respect to chosen genres, publishing genres that are more or less well attuned to the just quoted constraints.
Examples for video repurposing:

- from a single video registering a conference on public policy, extract the relevant sequences (i.e. relevant for a user), comment, enrich, ... them in order to use them for your university course and publish them as a thematic folder;

- from a corpus of videos documenting French immigration in Mexico, identify and extract the segments you want to have in order to produce a virtual narrative (a “virtual documentary”);

- from a set of documentaries, extract a corpus of sequences exemplifying documentarist scenario writing,

- from a French spoken conference video, extract some 20 minutes and comment them for English, German, … speaking users;

- from a bigger corpus of audiovisual resources dedicated to a given domain, extract a corpus of segments, organise them thematically and publish them as a kind of online hypermedia encyclopaedia.
Typical barriers between what we call the « source corpus » profile and a target user profile are, for instance:

- linguistic barriers,
- cognitive barriers,
- axiological barriers (i.e. barriers referring to diverging value systems),
- emotional barriers,
- barriers concerning the organisation and presentation of an information,
- barriers belonging to the lecture and appropriation routines (traditions) of the user community.
General picture of “repurposing”:

1/ a given (open) corpus of audiovisual (visual, textual, …) resources;

2/ a set of (static or dynamic) publishing models;

3/ the processing activity itself (extracting, categorising, indexing, commenting, versioning, …);

4/ resources:

- technical resources (working and publishing environment),
- semiotic resources (ontologies, description schemas, …),
- practical resources (guidelines, …).
6th topic

- Ongoing – an future - R&D challenges -
Key domains in R&D:

1/ the semiotics of audiovisual text and data
   - semantic rich access to audiovisual data
   - delinearized access
   - ontologies for description and indexing: themes, terminologies, description schemas, …
   - multilingual access

2/ the repurposing process as a “cultural translation” process
   - the deeper understanding of the semiotic but also historical nature of (audiovisual) texts and discourse in order to “simulate new and richer forms of “aggregating” delinearized videos
   - the “dynamic” aggregation of content with respect to specific profiles of user or user communities
6th topic
R&D challenges

- 3/ the archive for everyone: individual + institution ("ARA as a AV knowledge production, sharing and exploitation cooperative")

- 4/ the archive everywhere – u-archive

- 5/ the exploitation of an archive, the multiple uses of a video ("user scenarios")

- 6/ new forms in dealing with digital content:
  - immersive environments,
  - 3D videos,
  - the “ubiquitous capture”,
  - the “multiple worlds scenarios”.

Peter Stockinger: The ARA Program (Gananoque, 2009)