Fourth Winter School of the Estonian Graduate School of Culture Studies and Arts

Workshop 1: Digital archives and humanities: an explosive combination?

Digital audiovisual archives, semiotics and digital humanities

Prof. Dr. Peter Stockinger,
Fondation Maison des Sciences de l’Homme (FMSH)
Institut National des Langues et Civilisations Orientales (INALCO)

Baltic Film and Media School
Tallinn, Estonia, 23rd of January 2014
Topics of this lecture:

1) Short presentation and discussion of the notion « archive » in general.
2) Archives and culture.
3) Digital archives in particular and the central question of their exploitation.
4) Presentation and discussion of a concrete example: the Audiovisual Research Archives (ARA) program in Paris, at the Fondation “Maison des Sciences de l’Homme”
5) Semiotics and digital archives
« Archives » in general - a short presentation
« Archive(s) » - 2 related aspects ...

1. Archives = collection ("fonds") of data documenting activities, realizations, the history, ... of a social actor
   
   a) Data = objects of an archive (text, image, film, realia, ...) understood as documents;
   
   b) Record = data are processed (selected, "shaped", indexed, ...);
   
   c) (Social) actor = archived data as "documents" belongs to a social group, an institution, a company, an individual, ...

1. Archives = social practice of collecting, processing, preserving and, eventually, diffusion of archived data.
   
   a) Professions: archivist, curator, cultural heritage officer, custody provider, ...
   
   b) Skills: appraising (data), storing and preserving, cataloguing
   
   c) Procedures: (archival) administration (management)
Typical examples:

- The National Archives of (UK, Estonia, France, ... => national history)

- Company archives (=) corporate history of companies

- Community archives (=) history of local communities, diasporas, migrants, ...
  Example: Community Archives and Heritage (UK)

- Cultural heritage archives (intangible/tangible cultural heritage archives)
  Examples: European Heritage Portal

- Personal, family ... archives (=) genealogy, memory, identity transmission, ...

- etc.
Recurrent and central vocations of archives:

1) **Consignment of the past** of a social actor
   - “What has happened” and what is of “importance” for a social actor (what gives his life a “perspective”, an identity)

2) **Axiological (value) consignment** of the semantic (cultural) space of an social actor
   - Epistemic: what’s true, believable, …
   - Ethic/moral: what’s good, acceptable, …
   - Esthetic: …
   - Utilitarian: …
   - …

3) **Preservation and transmission mission**
   - What is “relevant” to be preserved and transmitted (to next generations) and how to do this?
Archives and culture
Particularly important aspect:

⇒ the *regulative* and *normative status* of an archive:

⇒ archived data constitute (are supposed to constitute) *references* ("standards") for the social actor

⇒ Simply speaking this means something like the following:

⇒ “In order to know/to believe/to decide, …

... what has happened in the past …
... what has been done (produced, believed, felt, ...) in the past …

... refer to the archive XYZ”.

----------------------------

⇒ “In order to know/to believe/to decide, …

... what is the “profound” identity of ....
... what is the artistic canon of ..., the artistic tradition of ...
... what is the truth of ...

... refer to the archive XYZ”.
Consider also the etymology of “archive”:

lat. “archivum” < Greek “ἀρχω”

1. govern, rule
2. and also “start”

M. Foucault: the archive forms a kind of « historical a priori » for a community

- Foucault refers here to the medical and criminal sciences and technologies used in the 18th century

J. Derrida: “... the question of the archive is not, ..., a question of the past, ..., but rather a question of the future, ..., of a promise and of a responsibility for tomorrow” (Mal d’Archive; 1995)
This intrinsic **normative (and regulative) status** of an archive is sometimes **explicitly defined by laws** ...

**Example of a law (France)**

- the French [Code de Patrimoine](#) (2004; 2011) – Cultural heritage Code – that defines, among others, the legal deposit (of intellectual realizations), general dispositions concerning archives, libraries, museums and monuments, etc.)

**Example of audiovisual archives (France)**

- INA = *Institut National de l'Audiovisuel* (legal deposit of the French public *TV productions*)

- **Archives françaises du film** (French Film Archives): legal deposit of the *French film productions*
But the regulative function is not restricted to its explicit, legal face.

It is inherent to each archival project.

Examples

- Memoro – The bank of memories (through life stories collected all over the world)
  => “bottom up” project of references to the “anonymous mass” who contribute to (regional, community, national, …) history but who don’t leave any traces in “academic history” ….

- MM – Mediterranean Memory – the audiovisual memory of the Mediterranean area (= Eurormed Heritage IV programme)
  => The countries of the Mediterranean region who create their common references …

- Archives en ligne – amateur films from the beginning (1896) till 1990 of Upper Normandy (France) and East Anglia (UK) – European INTERREG IV project
  => two adjacent regions that create their common historical and experiential references
Digital (audiovisual) archives and the central question of their exploitation
« Digital archives »:

- ... can be understood, basically, as a “combination” of 4 central dimensions:

1. as repositories of selected, pre-processed and recorded data
   a) Data can be either digitized data (i.e. of digital versions of non-digital originals) and/or of recorded genuinely born-digital data.
   b) Data are selected and recorded because they are considered to represent (to document !) a (cultural, historical, ...) heritage of a social actor

2. As repositories of the metadata, i.e. information identifying and describing the stored data
   a) Metadata identify the “profile” (the provenience) of a data.
   b) Metadata describe form, content and, eventually, uses of a data.
   c) Metadata describe legal rules with respect to and the archival history of a data

3. (Eventually) as the publication of a data and metadata in form of catalogues, structured “web pages”, portals, online libraries, ...
« Digital archives »:

- The **really new dimension** is the fourth one: digital archives include also at least potentially ...

4. ... **editorialized** and/or totally **user/community generated versions** of their data and metadata in form of:

- ... specialized **re-analyses**, new interpretations, “subjective” comments, ...

- ... **repurposed** (re-mediatised, ...) publications attuned to specific contexts of uses ...
« Digital archives »:

The interest of digital archives and their challenges for digital humanities in general and semiotics in particularly are concerned with this fourth dimension which requires:

a) A transdisciplinary approach of the notion of text broadly speaking

b) An appropriate “instrumentation” of this approach in form of working environments and software tools as well as new working methods

1) **Inathèque**

2) Ina – reediting 1: **Living arts**

3) INA.fr: re-editing 2: “**personalities**”

4) INA – **user generation**
Another example of reediting/repurposing of a digital AV data:

“Die baltische Hafenstadt Reval” (1920): a silent b/w film that ...

1) ... is originally stored (in its original form and through its identifying metadata) in a “traditional” archive (= Bundesarchiv-Filmarchiv Deutschland, Koblenz)

2) ... is selected, “processed” (=digitized) and integrated with other German films in a digital AV archive made available through a “national” web portal – the Filmportal.de (Frankfurt)

   2.1) Link: The virtual video archive
   2.2) Link: The web page publishing the video

3) ... is re-selected, re-analyzed and integrated with other European films in another – virtual - digital AV archive and made available through a central European access point - the European Film Gateway (Frankfurt – coordinated by the Deutsche Filminstitut)

   3.1) Note: EFG = access point to films, images and texts stored in European archives: CNC, Esti Filmiarhiiv, Bundesarchiv-Filmarchiv D., Cineteca di Bologna, Filmarchiv Austria, ...
   3.2) Link: The virtual video and image archive of EFG

4) ... is selected and republished with other thematically relevant films in one specific editorial project of the EFG: the 1914 (WWI) Project

   4.1) Link: The web page publishing this new version of the film
Other example: “Die baltische Hafenstadt Reval (1920) ...

a) [original] Data – place: Bundesarchiv D.
b) [digitized data] – place: Filmportal digital archive

(Meta-data/Description 1 of Bundesarchiv + 1st publication without video access)

(Metadata/Description 3 of EFG & WWI project + 3rd publication + link to Filmportal)

(Metadata/Description 2 of Filmportal.de + 2nd publication with video access)
General structure of a digital video (audiovisual) archive
General vision of the structure of a digital video archive

A) « Fonds of video (audiovisual) data »
1) Video data selection
2) Video data preparation
3) Video data storing

B) Video description/indexing
1) Current archive description
2) Legal deposit
3) Perennial archiving
4) Specialized issues
5) Interoperability: DC, EAD, OAI, LOM, SKOS, ...

C) Video Publishing
1) Only of metadata or of videos + metadata
2) Media choice (print, digital)
3) Genre choice (catalogue, portal, ...)

D) Exploitation
1) Isolated video/collection of videos (belonging to 1 or several “fonds”)
2) Reuse of parts (segments) or specific layers of a video
3) Editorialized reuse (new archives, virtual narratives, web-docu, ...)
4) User – generated reuse
5) Media: one, cross-, trans-, ...

Digital AV project
Digital AV reuse/repurposing
Example of a digital archive project: the Audiovisual Research Archives (ARA) Program in Paris
An example: ARA = “Audiovisual Research Archives” program:

- ARA is a French R&D (research and development) program, based in Paris, at the “Fondation Maison des Sciences de l’Homme” (appr.: Institut of Human and Social Sciences).

- Since 2002, ARA digitizes, publishes online, preserves and exploits the scientific and cultural heritage in social sciences and humanities.

- Main activities are:
  
  ✓ = **filming** of interviews with researchers, conferences, seminars, concrete fieldwork (in archeology, anthropology, ...)

  ✓ = **filming of cultural** (artistic, ...) **events**;

  ✓ = and **collecting audiovisual (video) data** produced by other actors (individual researchers, research labs, cultural associations, etc)
Example of a digital archive project:
The ARA video archive fonds
An example: ARA = “Audiovisual Research Archives” program:

- The **video archive fonds** to-day (2014):
  - about **6000 hours of online open access videos** (+ some 2000 hours waiting to be processed and published);
  - these **6000 hours** of online videos cover (more or less well) the principal disciplines in social and human sciences – for example:
    - some 650 hours in sociology
    - some 450 hours in archeology
    - some 600 hours in general history
    - etc.
  - **55%** of the videos are in **French**;
  - important quantity of videos in English and Spanish;
  - in total, **16 languages** are represented (among them: Russian, Chinese, German, ...).
Example of a digital archive project:
The publishing of the ARA video archive fonds
An example: ARA = “Audiovisual Research Archives” program:

- **Standard publishing of the ARA archive **fonds**:

  1) **Events**: Collected and processed videos are published in form of *dynamic web sites* called “events”, i.e.:
  - **Videos** are data used for *documenting* specific scientific “*events*” selected by their scientific committee of the ARA program
  - Events are classified in *collections* (topics, disciplines, ...)

  2) **Web portal** diffusing the videos
  - A **general portal** for accessing all events and videos

- Multilingual version of ARA portal: English, German, Spanish, Chinese, ...
2) The dynamic web site of an event (French version):

Example: La fabrication du pain domestique au Portugal (published on the ARA portal in June 2006)

- 1) (home page)
- 2) (content page)
- 3) (video lecture)
2 other examples of dynamic “event” web sites used for publishing an open hypermedia online documentations (related to South America)

Proyecto de memoria y educación en el Estadio Nacional de Chile: “Museo Abierto, Sitio de Memoria y Homenaje”

Example of a digital archive project:

The ARA archive production process. Methodology and procedures
The **ARA Production process** ("archive project")

1) The whole process of the production of this heritage is governed by a **production plan** composed of **6 main phases**;
   - each phase is defined through a **task list** and **explicit procedures**
   - the production plan is compliant with the main norm in the field of digital archives, i.e. the **norm OAIS**

2) A subset of tasks are performed by the **production team** (very small – 2,5 persons)

3) But many remaining tasks are performed by **third persons, external** to the production team (hence the **practical necessity** to possess an explicit production plan)
The ARA Production process ("archive project")

4) Tasks principally in charge of third persons (AV professionals, researchers, event organizers, ...)

a. The preparation of the filming and the filming itself of an event is in charge of external professionals (financed by the event organizers)

b. The collection of all legal documents (contracts) is under the responsibility of the researcher or other concerned persons;

c. The basic processing of the data is in charge of the external professionals;
The ARA Production process ("archive project")

5) Tasks mainly in charge of the production team ...

a. The montage can be assumed either by external professionals or by the production team

b. The "intellectual processing" (editing):
   1. description/indexing,
   2. pre-publishing,
   3. quality control,
   4. and definitive publishing

c. The legal deposit and the perennial deposit

d. The classification and management of all legal documents.
Example: The six central phases defining a **concrete ARA archival project** ...

| 1st phase: « Up stream” activities – preparation of the field work | 1) Definition and elaboration of the projects objectives and domain;  
| | 2) Team building;  
| | 3) familiarization; preparation of the filming/collection of data; ...  
| 2nd phase: Field work | 1) Setting-up, preparing, ... of the “terrain”  
| | 2) Filming of... : people, activities, events, ...  
| | 3) Collection of other relevant data (“documents”)  
| 3rd phase: Technical preparation and registration (deposit) of produced/collected material | 1) Derushing + technical enhancements  
| | 2) Digitizing of non-digital material  
| | 3) Recording of basic information in database  
| | 4) Uploading of files containing data  

The ARA production process
**Example:** The six central phases defining a **concrete ARA archival project** ...

| **4th phase:** **Editorial processing of the uploaded material** | 1) Video-segmenting  
2) Video description (content, audiovisual level, provenience, copyrights, ...)  
3) Video enriching (links, comments, ...)  
4) Video translation/adaptation  
[5) Video montage & postproduction] |
| --- | --- |
| **5th phase:** **Online publishing and diffusion** | 1) Web site creation of field work containing the previously edited material  
2) Pre-publishing and quality assessment  
3) Diffusion on web portal + via newsletter, social media, ... |
| **6th phase:** **Closing activities** | 1) “Physical on-the-spot“ archiving of all traces.  
2) Legal and perennial deposit.  
3) Archiving of contracts  
4) Follow-up of people database |
The exploitation/repurposing of digital audiovisual archives: Examples from the Audiovisual Archives Research (ARA) program
The exploitation of digital (audiovisual) archives:

1) An archive or a (published) data of an archive constitutes only a potential resource for a user.

2) In order to transform a data in an effective resource for a given user or user group it is necessary to “manipulate” it actively in order to attune it to specific usages, user contexts, etc.

Many different related terms in order to capture the notion of “exploitation” of data (collections of data) of digital archives:

- Segmenting/re-segmenting and re-configuring (video segments)
- Repurposing
- Republishing
- Remediating
- Collaborative and personalized archive building;
- User generated content for archive enhancing;
- Channelisation (of digital assets with respect to a user’s interest, preferences, ...);
- ...
Examples of exploitation/repurposing video content from the ARA program ...

- Experimentation with and use of series of republishing/repurposing genres ("templates")
  1. Specialized web portals for accessing thematically or otherwise circumscribed events and videos;
  2. Semantic video portals for accessing videos or segments (sequences) of videos with respect to their topics, their rhetorical specificity, their audiovisual specificity, etc.
  3. Bilingual folders
  4. Dynamic video corpus construction
  5. Video-books
  6. ...

= On-going research!
1st example of republishing (segments) of a video in the ARA program

- Data: Alto Bio Bio (doc. 02:15, Santiago Chile, 2007)

The exploitation/repurposing of AV archives – the example of ARA
2nd example of republishing/repurposing a sub-part of the video archive fonds of ARA in form of ...

Specialized web portals

Example: SCC – Semiotica, Cultura, Communicazione (cooperation between university La Sapienza of Rom and FMSH)
3rd example of republishing/repurposing a sub-part of the video archive fonds of ARA in form of ...

Dynamic subjects corpora (example – the “encomienda” in Argentina during the 18th century portal: AMSUR)
4th example of republishing/repurposing a sub-part of the video archive fonds of ARA in form of ...

Dynamic video encyclopedia (example here “religious practices”; source: PCM portal)

L'Encyclopédie Hypermédia

Les sujets de connaissance : Traditions et pratiques religieuses
Approfondissez vos connaissances du sujet "Traditions et pratiques religieuses" en explorant librement les segments qui lui sont consacrés

Elizabeth D. INANDIAK : L'histoire de deux sites sacrés, épargnés des laves du volcan Merapi

En évoquant l'activité du volcan Merapi de 2006, Elisabeth INANDIAK PRASETYO explique que la coulée de lave a miraculeusement épargné deux sites sacrés, le 'banian blanc' et la 'pierre éléphant'. Elle explique que la 'Pierre éléphant', a pour origine une coulée de lave, qui se serait arrêtée pour laisser passer une femme enceinte, prenant alors la forme d'un éléphant. Elle ajoute que cette histoire reste valable aujourd'hui puisqu'elle a permis aux sauveurs de se protéger de la coulée de lave en 2006.
4th example of republishing/repurposing a sub-part of the video archive fonds of ARA in form of ...

Bilingual folders (example here: selection of a sample of significant segments of an interview with Sabine Trebinjac from the French CNRS on the muqam genre in the Chinese Turkestan and translation-adaptation in English, Italian, Russian, Chinese, ...
5th example of republishing/repurposing a sub-part of the video archive fonds of ARA in form of ...

Geographic localization of topics (example here: research on Amerindian languages and language groups and families ...; source: AMSUR)

Accès au portail AMSUR
6th example of republishing/repurposing a sub-part of the video archive fonds of ARA in form of ...

Videos by usage (example here: videos proposed for BA, MA, D level as well as for post-doc students offered by the new AGORA portal)

The AGORA portal
The Studio ASA – an environment for building and repurposing digital audiovisual archives
The working environment for building archives, analyzing and publishing video corpora – The **Studio ASA** (Atelier de Sémiotique Audiovisuelle)
Le **Studio ASA** – a set of software tools which can be downloaded on a PC ...
The virtual video segmentation tool (INA-FMSH/ESCoM)
The video description tool (INA-FMSH/ESCoM)
The **video web publishing** tool **Semioscape** using the CMS Umbraco
The conception of the ASA Studio is based

- on **structural (text and discourse) semiotics**;
- and **other related approaches and disciplines**
  - such as: the rhetorical structure theory, discourse analysis, thematic and cognitive analysis of texts, knowledge representation, ...

The development of the Studio ASA has followed, as precisely as possible the “instructions” formulated from the side of human sciences

- => “**humanistic computing**”
The Campus AAR Project
• In an actually ongoing French (ANR funded) R&D project – **Campus AAR** - the objective is ...

  • ... to enable any interested person or group to conceptualize, produce, analyze, diffuse and preserves **his/her own archives**:

    • cf.: the notion of **user generated archives**

This means: everybody should have the possibility to **download for free**:

  • the **whole environment** (the ASA Studio)

  • all **procedures** and **guides** for building his/her own archives

• Three of the most struggling challenges here is:

  1. The definition and implementation of a **basic meta-language** on the top of specialized **domain ontologies**;
  2. The definition of a **library of procedures of video description**;
  3. The definition of a library of **video publishing/republishing templates**.

• Actually, a prototype is available at this url: [http://www.semioscape.fr/studio-asase/telechargement.aspx](http://www.semioscape.fr/studio-asase/telechargement.aspx)
Semiotics and archives
**Semiotics** - Definition (among many others, more or less compatible ones):

- Semiotics is a **theoretical framework** and a **methodology** for describing and analyzing ("expertizing") ...
  1. the production, exchange, sharing, conservation, reuse, ... of **messages**
  2. by the means of one or more **medias** and in form of **texts** (broadly speaking)
  3. in using a **language** (the "semiosphere" peculiar to the culture of a social actor)
  4. attuned to the specific the **context of use** ("social structure")

- Different traditions:
  - one important is the structuralism and **structural semiotics** going back to de Saussure, A.J. Greimas, R. Barthes, ...);
  - another is the **cultural semiotics** approach of J. Lotman and the Tartu school;
  - a third one is the **pragmatic** and **sign-theoretic approach** of Peircean semiotics, etc.
Personally, I think that semiotics can be used:

1. As a **methodology for the production of digital archives** (i.e. as a methodology for a “digital archive project”).
   - Definition of an archival project;
   - Specification of meta-language (ontology) and metalinguistic resources
   - Scenarios of describing and indexing
   - Scenarios of first publishing, preservation and legal deposit

2. As a methodology for **exploiting, reusing** digital archives.
   - Projects of authored republishing (narratives, web-documentaries, courses, ...)
   - Projects of creating user-based virtual archives
   - Projects of sharing and commenting digital archives
   - Projects in cross- and trans-media “storytelling”,
   - Etc.

3. As a methodology for **analyzing existing digital archives** and/or data stored in digital archives.
   - Audiovisual corpus analysis
   - “Virtual ethnography” of peoples productions, uses, shares, etc.
   - Information watch
   - “Cultural intelligence” (hunting of tendencies, ...)
   - Strategies of appropriation (remediation, ..)
   - ....
The central assumption here is:

- the **data** of an archive (as well as the **archive itself** as a complex data) ...
- ... is a text or can be approached as a **text** (cf. next slight)
- (cf. also the "**linguistic turn**" in archive studies)

Structural semiotics proposes an **appropriate theoretical framework** that enables us ...

- ... to conceive an **operational vision of the object "text"** (cf. the last two slights).

Naturally, this vision is **fallible** and **has to be revised**, if necessary ...
1) Semiotics as a **methodological and theoretical approach** for building digital audiovisual archives

**Data AND the archive = Text-object (video, photography, sound, …)**

**Archive Definition**
1) Archive definition and conception;
2) Specification of meta-language;
3) Web portal building

**Archive production**
1) Text Selection,
2) Text Production,
3) Text Processing (ex.: digitizing)
4) Text Recording

**Analyzing/ reanalyzing**
1) Basic analysis;
2) Specialized content analysis
3) Analysis of audiovisual level
4) Translation/ adaptation ...

**Publishing/Republishing**
1) Automatic/authored
2) Controlled/free
3) Media and genre
...
The semiotic theory of text – three complementary aspects:

1) The text as a **compositional entity** (= modularity assumption)
   - i.e. both, a video and a sequence of the video (a “segment”) are texts and share – in principle – the **same structural characteristics**

2) The text as a **layered entity** (= structural assumption)
   - i.e. a collection of videos, a corpus of videos, a single video or again a sequence of a video ... are constituted by several **meaning layers** such as a **referential layer**, a **rhetorical layer**, a **discursive layer**, an **expression layer**, etc.

3) The text as a **social practice** (= genetic assumption)
   - i.e. a video (corpus of, sequence of, ...) has **lifecycles** which are typically organized in phases such as the conception, the realization, the processing, the recording, the description, the analysis, etc.
A more detailed representation of the three complementary approaches to text as used in the ARA program ....

1. Segments, sequences, parts, regions, etc.
2. Functional (“named”) textual types.
3. Corpora, collections, etc.
4. Archives, libraries, textscapes, etc.

1. Referential thematic component: topical structure.
2. Discourse production component: discursive framing.
3. Discourse production component: discursive development.
5. Modalities of expression component: written, audiovisual, etc.
6. Component of the formal & physical organization of the text.
7. Component of the physical support of the text.

5. Practices of “consumption”.