

Att göra forskningsdata sökbara  
- från Europaprojekt till lokal implementering

---

Sven Strömqvist & Nils Holmberg  
Lunds universitet

[www.ling.lu.se/projects/echo/contributors](http://www.ling.lu.se/projects/echo/contributors)

**ECHO**  
EUROPEAN CULTURAL HERITAGE ONLINE

**European Cultural Heritage Online**

Contributor's Web Site

LUND  
UNIVERSITY

[Dear Colleague](#)  
[Presentation of ECHO](#)  
[Partners](#)  
[Charter](#)  
[Statement of Purpose](#)  
[Announcements](#)  
[To Contribute](#)

**Presentation of ECHO**  
--an echo of our cultural past to our future culture



Prospectival representation of a royal palace, reconstructed by the Bibliotheca Hertziana (MHF)

The ECHO consortium has given itself a charter to ensure its basic values, goals and restrictions. These general propositions include the free availability of tools and content (in particular European Cultural Heritage) on the Internet, the support of open standards, measures on long term archiving and the provision of a common portal.

The charter serves as a guideline for an extension of ECHO into its second phase ECHO 2, as proposed in an Expression of Interest to the 6th Framework Program of the

There is a growing understanding that in order to

"Cultural Heritage encompasses material culture, in the form of objects, structures, sites, landscapes, as well as living (or expressive) culture as evidenced in forms such as music, crafts, performing arts, literature, oral tradition and language. The emphasis is on cultural continuity from the past, through the present and into future, with the recognition that culture is organic and evolving."

World Bank, Washington, 1988

The vision of the ECHO initiative is to create an encompassing infrastructure for the historical heritage of European science in the Internet, constituted by a self-organizing network of open digital libraries. This infrastructure will, at the same time, represent a test bed for exploring strategic problems of science in the Internet age, such as archiving and knowledge organization. Its realization will generate new technology, but also involve harvesting the results of advanced pilot projects. The initiative will create and implement living standards and tools capable of sustaining a widespread development of digital libraries, and establish a culture of connectivity between them.

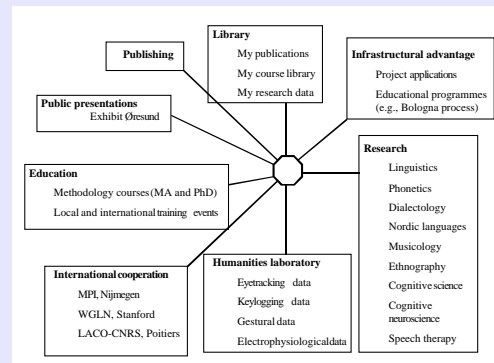
## Educating the Humanities for e-science

- caring, sharing and creating added values

Sven Strömqvist

[http://www.kb.se/dokument/Bibliotek/utbildning/educating\\_the\\_Humanities\\_e-science.pdf](http://www.kb.se/dokument/Bibliotek/utbildning/educating_the_Humanities_e-science.pdf)

## Eight contexts for the integration of DAM-LR at Lund University



## Added values

- facilitate research cooperation
- enrich education
- facilitate mobility of researchers and students (e.g., the Bologna process)
- solve potential conflicts between cultural institutions and research groups
- enrich presentation projects (museums, encyclopaediae, popular science, media)
- create an infrastructural advantage

## Översikt

1. **Forskningsdata på Humanistlaboratoriet**
  - forskningsområden och datatyper
  - produktion av forskningsdata
2. **Impementering av sökbara forskningsdata**
  - syften till sökbarhet
  - implementering av sökbara data (Richthoff-corpus)
3. **Länka forskningsdata och publikationer**
  - LUP, SND, MPI

## Forskning på Humanistlaboratoriet

- **Språkforskning**
  - språkinläring
  - språk och gester
  - dialekter, artikulografi
  - corpus-lingvistik
- **EEG och GSR**
  - begreppsbildning
  - syntax, pragmatik
- **Eye-tracking**
  - visuell perception
  - läsning
  - kognition
- **Motion-tracking / Virtual reality**
  - digital arkeologi
  - religionspsykologi

## Produktion av forskningsdata



Typiskt experiment vid Humanistlaboratoriet: Ett 40-tal försökspersoner får läsa autentiska dagstidningar i 20 minuter. Läsprocessen registreras med hjälp av en huvudburen eye-tracking-kamera.

- **Insamling av rådata**
  - => sifferdata, videodata
- **Dataanalys**
  - => exkludering, aggregering
- **Bearbetad data**
  - => statistisk analys
- **Resultat**
  - => publikationer
- **Arkivering**
  - => spårbarhet, replikering
- **Alla steg är forskningsdata?**
- **Alla steg ska göras sökbara?**

## Syften med sökbara forskningsdata

Syften relaterade till Humanistlaboratoriets verksamhet

- Samarbete mellan forskargrupper
- Forskare vill replikera experiment
- Finansiärer vill kontrollera resultat
- Insyn för allmänheten
- Skapa corpusar som växer över tid
- Arkivering för redundans/backup

## Implementering av sökbara forskningsdata

Exempel: Workflow i Richthoff-corpusen

1. Digitalisering och transkribering av insamlad data
2. Generera och validera metadata (**IMDI-XML**)
3. Indexering och reglering av åtkomst (**ARBIL**)
4. Publicering av data och metadata (**LAMUS**)
5. Använda sökgränssnitt för att hitta data (**MPI**)



## Sökgränssnitt för att hitta data



- Metadata används för att söka efter forskningsdata
- All metadata registrerad i centralt söksystem
- Metadata innehåller index till lokalt lagrad forskningsdata

## Länka forskningsdata till publikationer

- LUP
- Svensk Nationell Dataservice
- Lund Corpora (MPI)

## Workflow i Richthoff-corpusen



- Rådata  
(videoinspelningar)



- Dataanalys  
(digitalisering och  
transkribering)



- Generera metadata  
(imdi-xml)



- Validering av  
rådata och metadata  
(Arbil)



- Indexering och  
publicering av  
forskningsdata  
(Lamus-MPI)