



UPPSALA  
UNIVERSITET

# Labour Gone Digital: Preservation of organizational activities in the On-line Era (DigiFacket)



RIKSBANKENS  
JUBILEUMSFOND

THE SWEDISH FOUNDATION FOR  
HUMANITIES AND SOCIAL SCIENCES

Jenny Jansson  
Katrin Uba  
Jaanus Karo

Department of Government  
Uppsala University

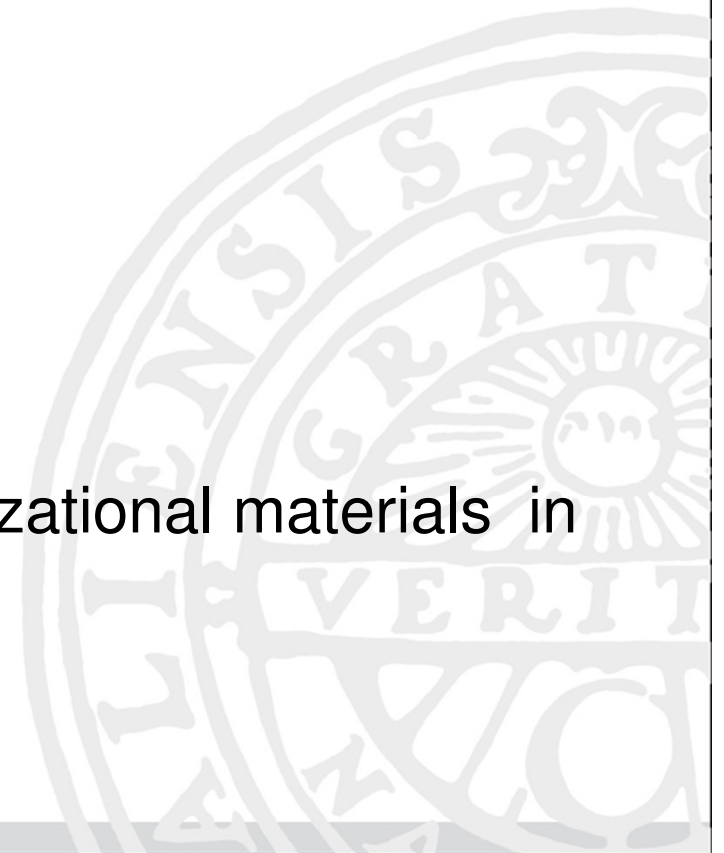


UPPSALA  
UNIVERSITET

# What happens with born-digital material on the Internet?

- Traditional social movement activities are now taken place on the Internet
- New activities have emerged

→ The challenge is to archive organizational materials in this new context





UPPSALA  
UNIVERSITET

# Aim of the Project

...to collect and archive material produced by Swedish trade unions online.

...to make materials available for scholars.

We download and index trade unions' webpages, Facebook pages, Twitter feeds and YouTube channels



UPPSALA  
UNIVERSITET

# What do we do that has not already been done?

## DigiFacket:

- Regular downloading
- The material is preserved in the movements' own archives

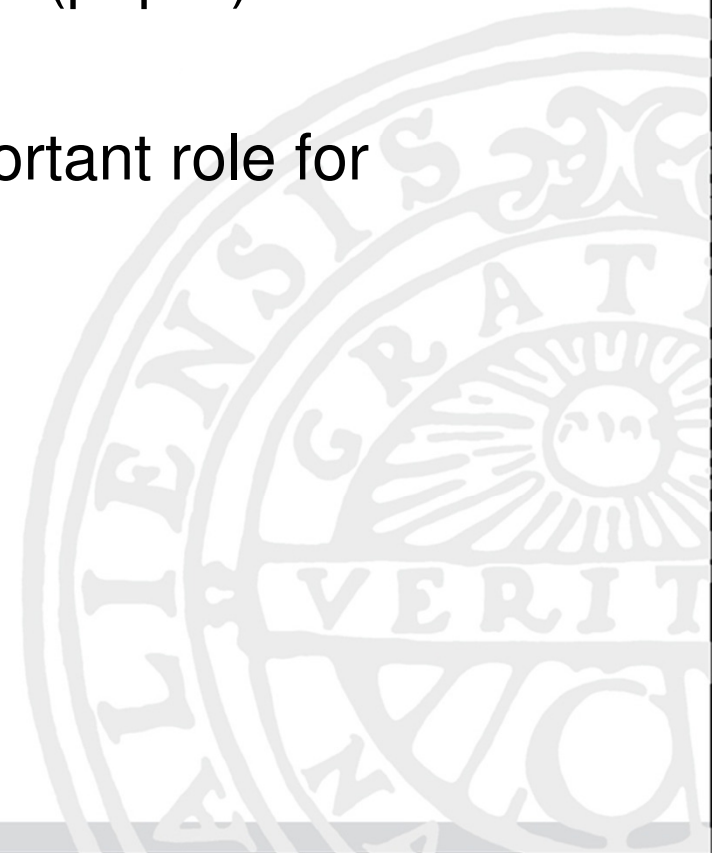




UPPSALA  
UNIVERSITET

# Why focus on trade unions?

- Old social movement with excellent (paper) archives
- Movement that has played an important role for democratization
- Easy to identify





UPPSALA  
UNIVERSITET

# What do we do?

## **For unions' webpages:**

Software (freeware based) that:

1. Harvest (collect and download material)
2. Storing
3. Indexing
4. User interfaces for maintenance and accessing data



UPPSALA  
UNIVERSITET

# Harvesting

- **NetArchiveSuite 5.6** combined with Heritrix3 (Internet Archive)
- Frequency:
  - The entire webpage: once in two months
  - The first page: once a week





UPPSALA  
UNIVERSITET

# Harvesting

Social media: different types of API

...legal greyzone

-> we have asked the unions to download the data for us  
(twitter history and facebook history)







UPPSALA  
UNIVERSITET

# Storage

- Downloaded files for one domain are packed in WARC format together with metadata (e.g., date of harvesting, domain url)
- For example, the Swedish Trade Union Confederation – LO – we have data in amount of **30 GB** (2015-2019 with a few gaps)

# Creating the index

## Two indexing databases

- Outback CDX (for OpenWayback history browsing)
- Apache Solr (for Solr wayback search)
  - Uses available metadata that comes with the downloaded files
  - Index created with thesaurus
- Time consuming





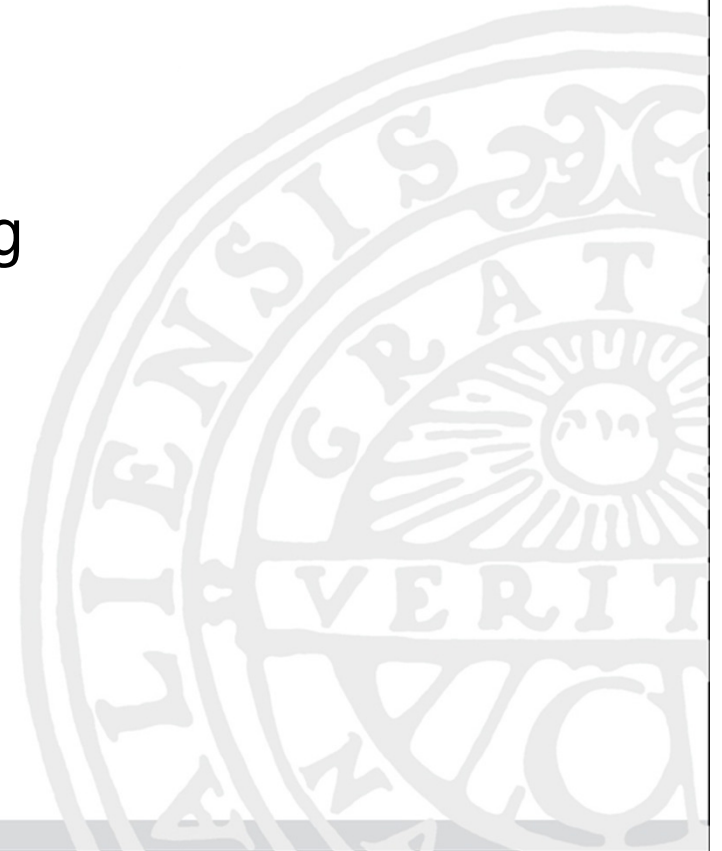
UPPSALA  
UNIVERSITET

# User interface in three sections

- NAS-UI → for administration (changing lists, log files etc., continuous maintainance)

Two interfaces for archive visitors:

- Open Wayback → history browsing
- Solr wayback → index search





# User interface: Solr WayBack search

SolrWayback Search - Mozilla Firefox

Kommunal x SolrWayback Search x Dåligt väder slår mot so... x +

← ⓘ http://localhost:9080/solrwayback/?query=solidari 67% 🔍 Search ☆ 📁 ⬇️ 🏠 🛡️ ☰

## SolrWayback Search

Grouping (slower response)

Image search  URL search

### Limit results

**Domain**

- lo.se : 20803
- arbark.se : 1895
- sac.se : 1477
- seko.se : 273
- ifmetall.se : 220
- suf.cc : 193
- mikromarc.se : 150
- abf.se : 133
- byggnads.se : 123
- facebook.com : 69


**Content Type Norm**

- html : 24879
- pdf : 423
- text : 216


Showing 1-20 of 25,760 hits

**Solidarity message to the people in Russia** 📄


**Content type:** text/html  
**Domain:** lo.se  
**Url:** https://www.lo.se/start/material/solidarity\_message\_to\_the\_people\_in\_russia  
**Score:** 119.2256  
**Highlighted content:** and a wage policy of **solidarity** Rapportserien Den svenska modellen Projektdirektiv En bättre värld LO



[Search for image](#)  
[Pages linking to image](#)



[Search for image](#)  
[Pages linking to image](#)

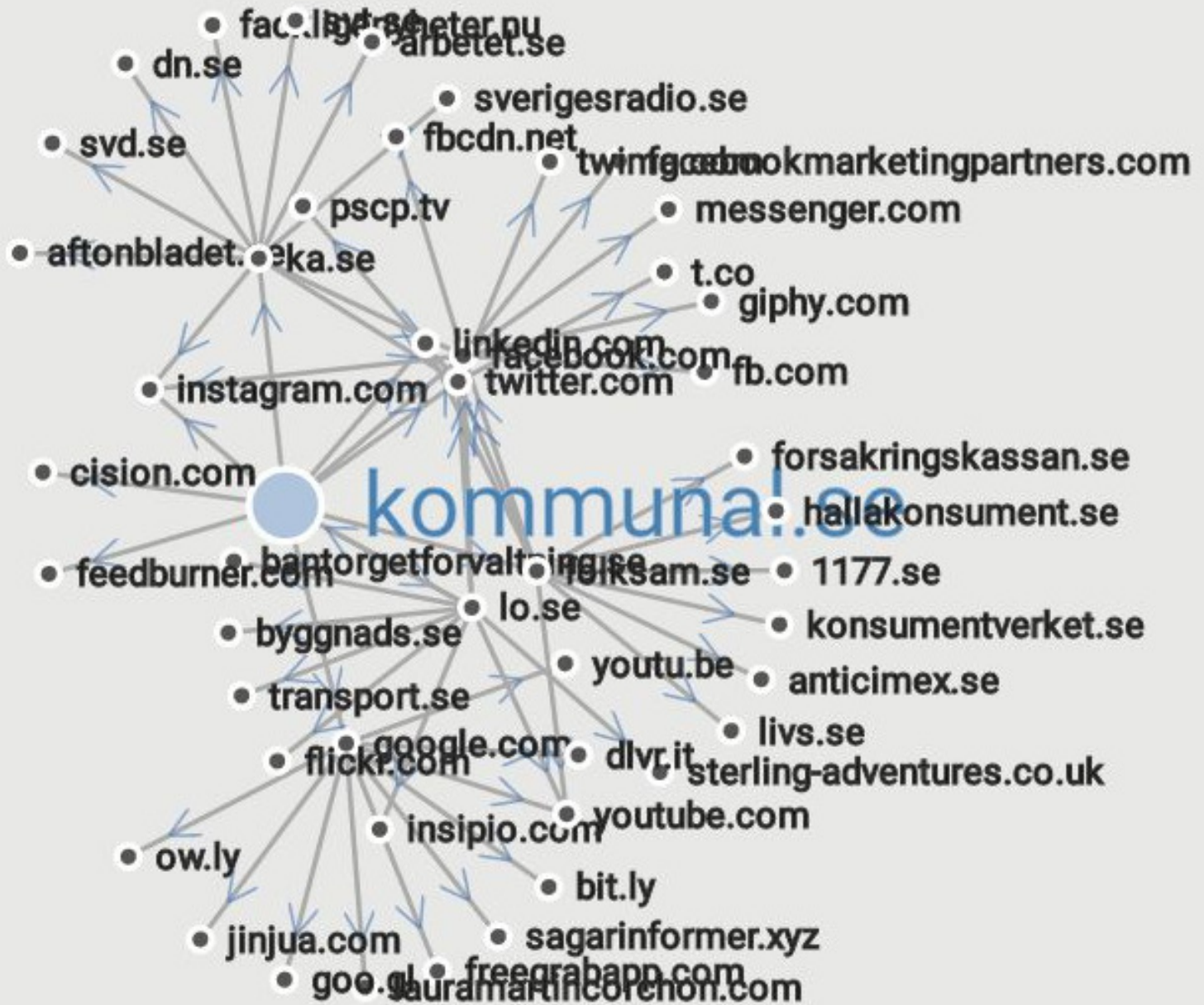


[Search for image](#)  
[Pages linking to image](#)

1 2 3 4 SolrWayback Search - Mo... 11:49:05 PM



UPPSALA  
UNIVERSITET





UPPSALA  
UNIVERSITET

More information available at:

[www.statsvet.uu.se/digifacket](http://www.statsvet.uu.se/digifacket)

[jenny.jansson@statsvet.uu.se](mailto:jenny.jansson@statsvet.uu.se)

