Digidaily

A newspaper dream come true

Riksarkivet
MKC - Mediakonverteringscentrum

KUNGL. BIBLIOTEKET
National Library of Sweden

Mittuniversitetet
MID SWEDEN UNIVERSITY

SCHIBSTED
MEDIA GROUP

LÄNSSTYRELSEN
VÄSTERNORRLAND

En investering för framtiden

EUROPEISKA UNIONEN
Europeiska regionala utvecklingsfonden

Digidaily

The Collections at KB

- Swedish newspaper publishers deliver three legal deposit copies of all Swedish newspapers printed
- KB has 31 600 meters of newspapers or approx. 122 million pages
- 70 million pages are filmed on microfilm
- A major newspaper donation from a former journalist school

Digidaily

The newspapers

Activities at KB

- Requirement specifications
- Planning
- Preparation, registration and delivery
- Handling of rejected and damaged newspaper material

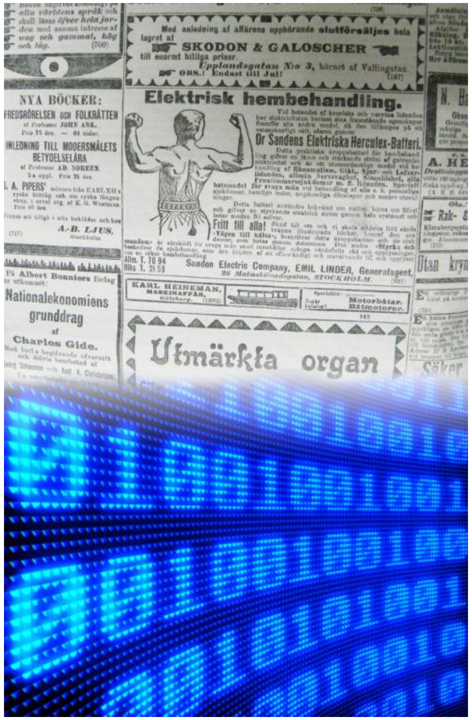
And much more…

Requirement specifications

- **End product?**
  - What needs do the end-users have?
  - What metadata is needed?
  - What file formats should be used?

  = many different requirement specifications

- **Infrastructure**
  - How should the files be transferred from MKC to KB?
  - Storage at KB?


SKODON & GALOSCHER
NYA BÖCKER:
Elektrisk hembehandling.
Dr Sandens Elektriska Hercules-Batteri.
Fritt till alla!
Sanden Electric Company, EMIL LINDER, Generalagent,
KARL HEINEMAN,
Nationalekonomiens grunddrag
Charles Gide.
Utmärkta organ

Planning

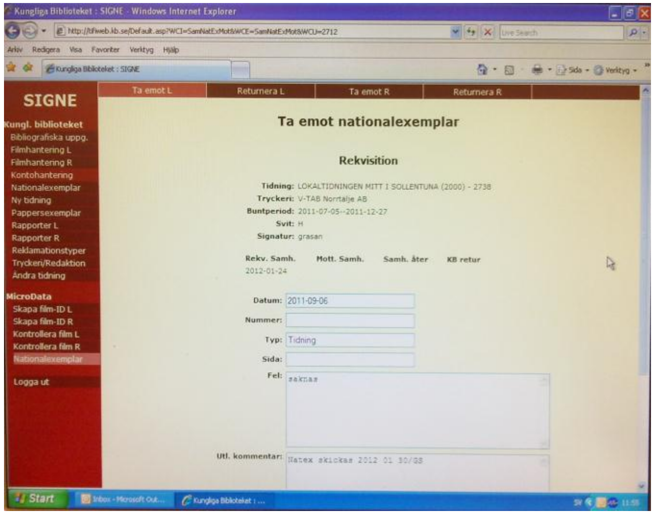
- **Overall Planning**
  KB assesses the newspapers' condition, volume, size, font type, if it’s already on microfilm, legal rights, scientific interest etc
- **Condition Survey**
  A condition check is done to get an overall picture of the newspapers' condition

Preparation

- **Review of the material**
  - Cleaning
  - Preparation on issue level
  - Comments on condition**,** supplements, parts, editions, missing issues
  - If necessary when the material is damaged, find a replacement copy

Registration

- Information on each bundle is recorded in KB’s database
- Data export/import including newspapers name, date, comments about supplements, editions, condition etc
- Lists of expected editions and attachments are exported/imported
- KB registers each bundle in the Workflow system



Ta emot nationalexemplar

Rekvisition

Tidning: LOKALTIDNINGEN MITT I SOLLENTUNA (2000) - 2738
Tryckeri: V-TAB Norrtalje AB
Buntperiod: 2011-07-05--2011-12-27
Svit: H
Signatur: grasan

Rekv. Samh.	Mott. Samh.	Samh. åter	KB retur
2012-01-24			

Datum: 2011-09-06
Nummer:
Typ: Tidning
Sida:
Fel: saknas
Utl. kommentar: Natex skickas 2012 01 30/GS

Delivery

- For transportation of the official national copies special transport boxes has been built.
- Each box is followed by a printed packing list from the Workflow system.

Damaged material

- **Rejects**
  What's the level for rejecting a page?
  Big holes? Small holes?
  Only loss of text?

- **Gaps**
  What gaps can be accepted?
  What's the priority? A complete material or large volume digitisation?

The goal is to find a rational way to handle rejects that suits both KB and MKC!

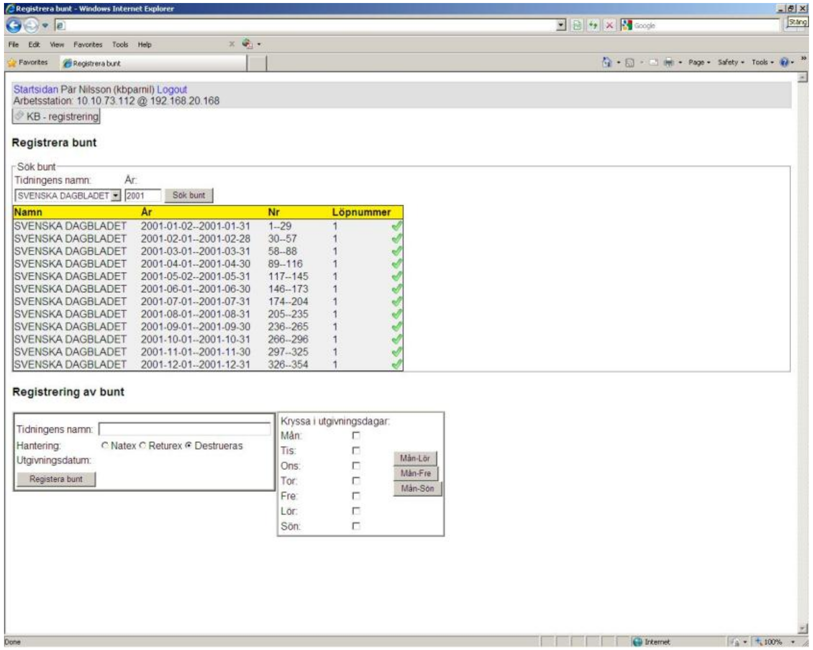
Activities at MKC

- Develop workflow system
- Tests, evaluation and procurement (scanners, OCR etc.)
- Production planning
- Preparation and registration
- Scanning, OCR, control features

And much more…

Workflow system

- Developed by a local team of developers at MKC
- Four major functions: database, add metadata through the process, keep track of and initiate the next process, collect data for statistics
- The workflow system is the backbone of the project
- Used by both MKC and KB



Startsidan Pär Nilsson (kbparnil) Logout
Arbetsstation: 10.10.73.112 @ 192.168.20.168
KB - registrering

Registrera bunt

Sök bunt
Tidningens namn: SVENSKA DAGBLADET
År: 2001
Sök bunt

Namn	År	Nr	Löpnummer	
SVENSKA DAGBLADET	2001-01-02--2001-01-31	1--29	1	✓
SVENSKA DAGBLADET	2001-02-01--2001-02-28	30--57	1	✓
SVENSKA DAGBLADET	2001-03-01--2001-03-31	58--88	1	✓
SVENSKA DAGBLADET	2001-04-01--2001-04-30	89--116	1	✓
SVENSKA DAGBLADET	2001-05-02--2001-05-31	117--145	1	✓
SVENSKA DAGBLADET	2001-06-01--2001-06-30	146--173	1	✓
SVENSKA DAGBLADET	2001-07-01--2001-07-31	174--204	1	✓
SVENSKA DAGBLADET	2001-08-01--2001-08-31	205--235	1	✓
SVENSKA DAGBLADET	2001-09-01--2001-09-30	236--265	1	✓
SVENSKA DAGBLADET	2001-10-01--2001-10-31	266--296	1	✓
SVENSKA DAGBLADET	2001-11-01--2001-11-30	297--325	1	✓
SVENSKA DAGBLADET	2001-12-01--2001-12-31	326--354	1	✓

Registrering av bunt

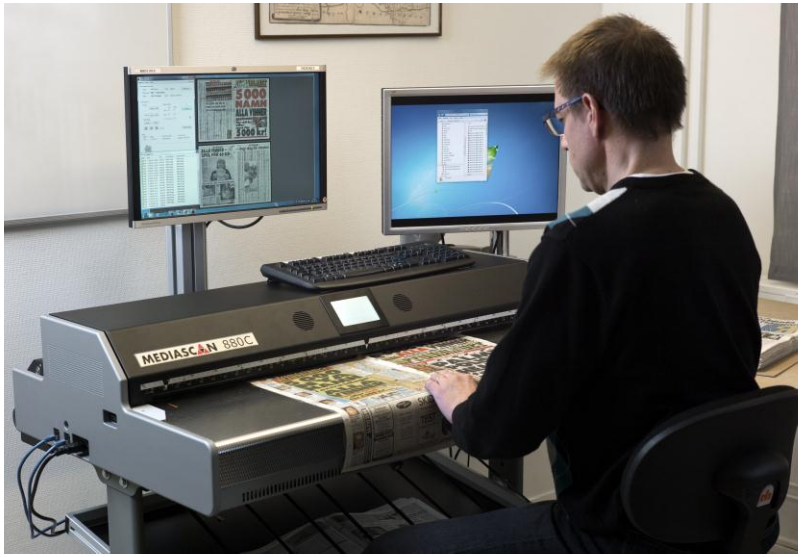
Tidningens namn:
Hantering: ○ Natex ○ Returex ● Destrueras
Utgivningsdatum:
Registera bunt

Kryssa i utgivningsdagar:
Mån: □
Tis: □
Ons: □
Tor: □
Fre: □
Lör: □
Sön: □
Mån-Lör
Mån-Fre
Mån-Sön

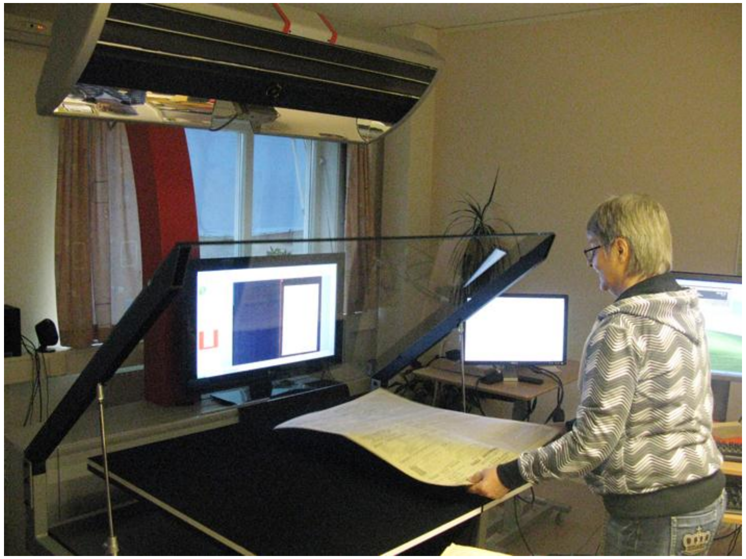
Preparering på MKC

- The preparation process: go through and take apart between 650 and 6000 pages per person and shift
- The workflow system is used to collect metadata

Scanning



SUPAG Mediascan 880c

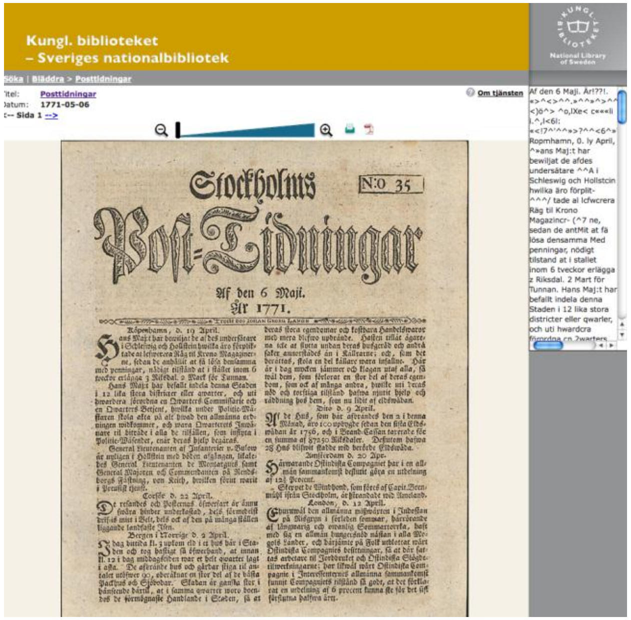


Zeutschel OS

OCR

Optical Character Recognition

- Content conversion software by Zissor
- Segmentation on article level
- Word lists used for OCR to improve word accuracy.
- No manual correction on headlines or text
- OCRed text and metadata in ALTO

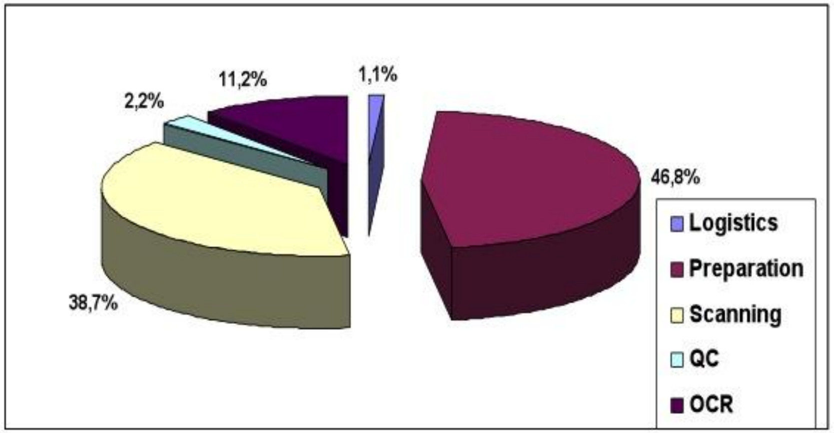

Kungl. biblioteket
– Sveriges nationalbibliotek
National Library of Sweden
Söka | Bläddra > Posttidningar
Om tjänsten
Stockholms N:o 35
Post-Tidningar
Af den 6 Maji. År 1771.

Titel: Posttidningar
Datum: 1771-05-06
<-- Sida 1 -->

```
Af den 6 Maji. Ar!??!.
«>^<>^^,»^^»^>^^
<)ö^> ^o,lXe< c«««li
i.^,l<6l:
«<!7^'^^»>?^^<6^»
Ropmhamn, 0. ly April,
^»ans Maj:t har
bewiljat de afdes
undersätare ^^A i
Schleswig och Hollstcin
hwilka äro förplit-
^^^/ tade al lcfwcrera
Räg til Krono
Magazincr- (^7 ne,
sedan de antMit at fä
lösa densamma Med
penningar, nödigt
tilstand at i stallet
inom 6 tveckor erlägga
z Riksdal. 2 Mart för
Tunnan. Hans Maj:t har
befallt indela denna
Staden i 12 lika stora
districter eller qwarler,
och uti hwardcra
förordna cn 2warters
```

£1 a page, one € a page or…

- Category 1 - bound, torn, fragile paper, large formats
- Category 2 - bound, where most can be taken apart and only a few are kept still bound, fair paper quality
- Category 3 – tabloids, stapled but not bound
- The price span today is between € 0.25–1.11/page incl. OCR



Digidaily

– ett utvecklingsprojekt där
Riksarkivet, Kungliga biblioteket och Mittuniversitetet tillsammans ska
utveckla rationella metoder och processer för digitalisering av dagstidningar.
Projekttid 2010-04-01--2013-03-31

Finansiärer är EU:s strukturfonder i Mellersta Norrland, Riksarkivet, Kungliga biblioteket,
Mittuniversitetet, Schibsted Sverige och Länsstyrelsen i Västernorrland

Riksarkivet
MKC - Mediakonverteringscentrum

KUNGL. BIBLIOTEKET
National Library of Sweden

Mittuniversitetet
MID SWEDEN UNIVERSITY

SCHIBSTED
MEDIA GROUP

LÄNSSTYRELSEN
VÄSTERNORRLAND

En investering för framtiden

EUROPEISKA UNIONEN
Europeiska regionala utvecklingsfonden