
Collecting bits and pieces

– the development of methods for handling e-legal deposit of on-line news material at The National Library of Sweden

Pär Nilsson

Background on legal deposit in Sweden

- First legal deposit legislation in Sweden in 1661
- Part of a series of reforms of the political system
- Main focus on control, not on building a national collection of printed publications
- "It is deemed to be useful and necessary that Their Royal Majesties may have knowledge about what books and other writings are printed and brought to light in the realm and the provinces"

From control to collection building

- But two copies were to be delivered, to the National Archives and to the Royal Library and not only books, but also newspapers, magazines and ephemera.
- The law was amended in 1674 and 1707, including fines and documentation. Increased number of recipients, from 1707: universities of Uppsala, Lund, Åbo and Dorpat.
- First freedom of the press legislation in 1766; amended in 1809 and made more liberal; in 1812 a system of registered publishers (responsible for the content) of periodical publications.

Development of legal deposit legislation

- In 1949 legal deposit became a separate law; largely intact for 30 years
- Next revision in 1978: microfilming of newspapers and legal deposit for sound and moving images
- 1993-2004 further changes to keep up with technological development, e.g. electronic documents in fixed form
- 2012 a new law on e-legal deposit material (SFS 2012:492) after almost fifteen years of reports and proposals

The road to e-legal deposit - 1998

- E-legal deposit report of 1998 (SOU 1998:111): to preserve and provide access to the Swedish cultural heritage for posterity; large amounts of published electronic material that fell outside the legal deposit law
- Material “widely available in this country and related to Swedish conditions”, even behind paywalls, collected as completely as possible (like printed and audio-visual material); collection method: web harvesting
- Focus on publications produced by professional publishers and producers
- Private web pages, information from local associations only by selection, collected four times a year; databases once a year

The road to e-legal deposit - 2003

- E-legal deposit discussed in a broader government 2003 report (SOU 2003:129) about the work and future of the National Library
- The existing legal deposit legislation to include “remotely transmitted digital materials”, defined as “such materials that are made available to the public via remote transmission over a network”
- Material of permanent character, i.e. material not intended to change over time
- The producer or provider of web page content to deliver e-legal deposit material, if already in possession of a publication license (i.e. a certificate of no legal impediment to publication); thus mandatory for newspapers, municipalities, authorities, etc.

Web harvesting in the Kulturarw³ project

- No changes in the law after the proposals on e-legal deposit in 1998 and 2003
- But web harvesting in the Kulturarw³ project since 1997: all Swedish web pages were to be saved a couple of times per year
- Daily harvesting of 140 newspaper web sites since June 2002
- An almost complete collection instead of a careful selection because it cannot be known what material will be in demand in the future
- Some legal support from 2002 in a regulation (SFS 2002:287) concerning the processing of personal data

Proposed e-legal deposit legislation

- In February 2009 a new investigation concerning e-legal deposit legislation and in November 2009 the memorandum “Legal deposit for electronic documents” (Ds 2009:61)
- Proposed new legislation which picked up where the 2003 report had left off
- Government bill on e-legal deposit June 13 2012
- The new legislation (SFS 2012:492) effective July 1 2012; closely follows the ideas in the proposal from 2009

Publishers covered by the law

Three groups of publishers covered by the law:

1. Publishers that have constitutional protection (e.g. newspaper publishers or TV and radio companies)
2. Government and municipal agencies
3. Companies which professionally produce electronic documents, e.g. e-books, e-music and e-movies

Electronic documents produced or provided by private individuals not generally to be included, e.g. private blogs

Implementation of the law

The new law is implemented in two steps:

- From July 1 2012 to December 31 2014 only a limited number of publishers: the ten largest (printed) newspapers, the ten largest (printed) magazines and journals, a number of radio and TV companies, and a number of government agencies
- The second step in January 1 2015 with identification of and information to all publishers covered by the law, including “enterprises professionally producing electronic materials”

Materials covered by the law

- No web pages and similar dynamic material
- Only unchanging electronic documents: “a defined unit of electronic materials with text, sound or image that has a predetermined content intended to be presented at each use”, e.g. news articles, opinion pieces, reviews
- Material published only online, but “web unique” content is difficult to identify and publishers are allowed to deliver material even if it has also already appeared e.g. in print
- Material “related to Swedish conditions”: aimed at people who understand the Swedish language, includes works by a Swedish author or a performance by Swedish artist or otherwise mainly targeted at the general public in Sweden

Systems, methods and organization - 1

- Development of an in-house system (Mimer) for handling e-legal deposit and other types of digital material
- Slow in the beginning, but archiving 2 million pages of digitized newspapers pushed development
- Mimer follows the OAIS reference model and is integrated with other systems like LIBRIS, the joint catalogue of the Swedish academic and research libraries
- Fedora Commons is used as a repository to store metadata about the files and keep a structural representation of the data
- A combination of an HSM system and cloud storage platform EMC Atmos is used for storage

Systems, methods and organization - 2

- The e-legal deposit law states that the material should primarily be delivered on a physical carrier, but in reality this will be the last resort
- FTP used for some material and will perhaps mostly be used for larger files especially for audio-visual material; receipt to the publisher when the files have been processed and archived by the library
- RSS used for frequently updated web sites e.g. newspapers and radio/TV websites, with automated retrieval of new items through a custom RSS service (combination of Dublin Core and Yahoo's Media RSS) roughly every hour
- A third method under development: a web ingest form for uploading material through a web browser

Systems, methods and organization - 3

Development of a web based platform to guide all potential suppliers in 2015:

- check that the publisher is a supplier of e-legal deposit according to the legislation and that they meet the technical requirements
- recommend the right method of delivery depending on the size and nature of the material
- provide information about what material is to be included
- handle automated processes for the registration and connection of each supplier
- keep track of the contacts between the National Library and the publisher

Systems, methods and organization - 4

The Mimer system also has a user interface (Oden) for the library staff making it possible to:

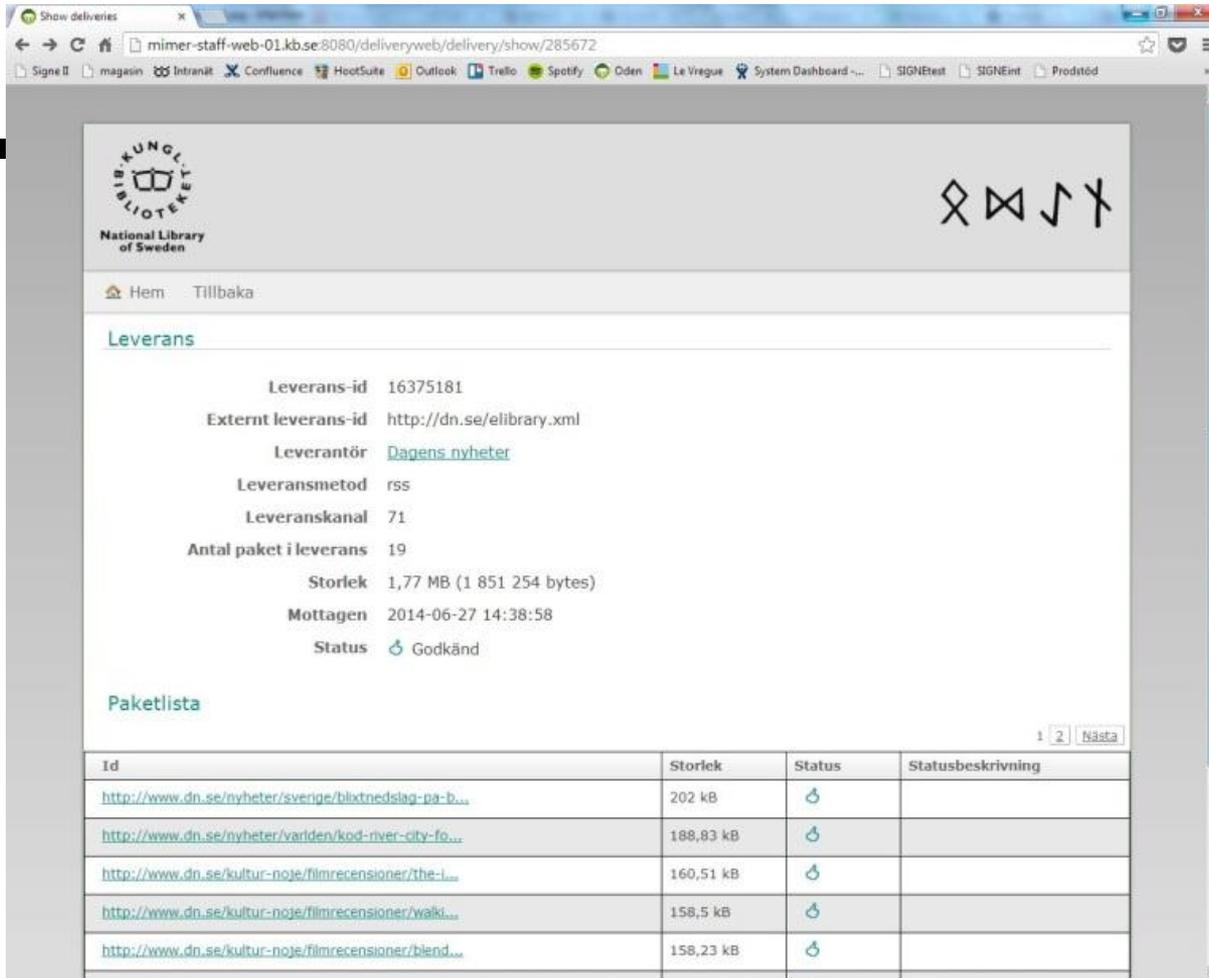
- monitor when and how much each publisher has delivered
- see the status of the material, i.e. if it was actually archived or if there is a need to investigate possible problems
- view the material itself by downloading the archival packet

The Oden interface – 1

The screenshot shows the Oden interface for the National Library of Sweden. The page title is "Leveranslista". The browser address bar shows the URL: `mimer-staff-web-01.kb.se:8080/deliveryweb/delivery/list?deliveryChannelId=&sort=id&max=50&order=desc&fromDate=Välj+från-datum&status=`. The page features the National Library of Sweden logo and a navigation menu. The main content area includes a search bar with "Dagens nyheter - DN.se" selected, a filter for "Antal träffar" set to 50, and a "Rensa filter" link. Below this is a table of suppliers.

Id	Extern leverans-id	Leverantör	Typ	Kanal	Mottagen	Status	Bearbetad
285678	http://dn.se/elibrar...	Dagens nyheter	rss	DN.se	2014-06-27 15:40:43	✖	✓
285672	http://dn.se/elibrar...	Dagens nyheter	rss	DN.se	2014-06-27 14:38:58	✓	✓
285669	http://dn.se/elibrar...	Dagens nyheter	rss	DN.se	2014-06-27 13:38:05	✓	✓
285658	http://dn.se/elibrar...	Dagens nyheter	rss	DN.se	2014-06-27 12:38:43	✓	✓
285657	http://dn.se/elibrar...	Dagens nyheter	rss	DN.se	2014-06-27 11:43:57	✓	✓
285649	http://dn.se/elibrar...	Dagens nyheter	rss	DN.se	2014-06-27 10:43:15	✓	✓
285629	http://dn.se/elibrar...	Dagens nyheter	rss	DN.se	2014-06-27 09:40:11	✓	✓
285592	http://dn.se/elibrar...	Dagens nyheter	rss	DN.se	2014-06-27 08:20:58	✓	✓
285336	http://dn.se/elibrar...	Dagens nyheter	rss	DN.se	2014-06-27 00:40:07	✓	✓
285335	http://dn.se/elibrar...	Dagens nyheter	rss	DN.se	2014-06-26 23:28:59	✓	✓
285333	http://dn.se/elibrar...	Dagens nyheter	rss	DN.se	2014-06-26 22:41:25	✓	✓
285328	http://dn.se/elibrar...	Dagens nyheter	rss	DN.se	2014-06-27 07:39:27	✓	✓
285327	http://dn.se/elibrar...	Dagens nyheter	rss	DN.se	2014-06-27 06:29:54	✓	✓

The Oden interface – 2



The screenshot shows a web browser window displaying the Oden interface. The browser's address bar shows the URL `mimer-staff-web-01.kb.se:8080/deliveryweb/delivery/show/285672`. The page header includes the logo of the National Library of Sweden (BIBLIOTEKET) and navigation links for "Hem" and "Tillbaka".

The main content area displays the following delivery details:

- Leverans-id:** 16375181
- Externt leverans-id:** <http://dn.se/elibrary.xml>
- Leverantör:** [Dagens nyheter](#)
- Leveransmetod:** rss
- Leveranskanal:** 71
- Antal paket i leverans:** 19
- Storlek:** 1,77 MB (1 851 254 bytes)
- Mottagen:** 2014-06-27 14:38:58
- Status:**  Godkänd

Below the details is a section titled "Paketlista" with a pagination control showing "1 2 Nästa". The table below lists the individual packages:

Id	Storlek	Status	Statusbeskrivning
http://www.dn.se/nyheter/sverige/blxtnedslag-pa-b...	202 kB		
http://www.dn.se/nyheter/varlden/kod-riker-city-fo...	188,83 kB		
http://www.dn.se/kultur-noje/filmrecensioner/the-l...	160,51 kB		
http://www.dn.se/kultur-noje/filmrecensioner/walki...	158,5 kB		
http://www.dn.se/kultur-noje/filmrecensioner/blend...	158,23 kB		

The Oden interface – 3

**LIBKUNGLI
BLIOTEKET**
National Library
of Sweden

Hem Tillbaka

Paket

Leverantör [Dagens nyheter](#)

Titel Blixtnedslag på Bråvallafestivalen - flera skadade

Publiceringsdatum 2014-06-27 14:38:58

Status Godkänd

Storlek 202 kB (206 843 bytes)

Externt paket-id <http://www.dn.se/nyheter/sverige/blixtnedslag-pa-bravallafestivalen-flera-skadade/>

Resursnyckel <http://www.dn.se/nyheter/sverige/blixtnedslag-pa-bravallafestivalen-flera-skadade/>

Paketinstans-id urn:nbn:se:kb:dark-package-instance-629212

Paketresurs-id urn:nbn:se:kb:dark-3847888

Leverans-id [16375181](#)

Libris ID [16620479](#)

Bearbetad Nej

Bearbetning startade 2014-06-27 14:50:03

Bearbetning avslutades 2014-06-27 15:14:06

Arkivpaket (AIP) [aip.mets.metadata](#), [Ladda ner: urn:nbn:se:kb:dark-package-instance-629212](#)

[Andra paket med samma resursnyckel](#)

Id	Externt paket-id	Typ	Status	Leverans

2014-08-16
Sidnummer 18

The Oden interface – 4

The screenshot shows a Windows Explorer window titled "Spara som" (Save as) in the foreground, overlaid on the Oden web interface. The Explorer window shows the "Skrivbord" (Desktop) folder with a list of files and folders. The file "um:nbn:se:kb:dark-package-instance-629212" is selected, and the "Filenamn" (Filename) field contains the same text. The "Filsformat" (File format) is set to "Compressed (zipped) Folder".

The background shows the Oden web interface. The top navigation bar includes "System Dashboard", "SIGNEvent", "SIGNEint", and "ProdStöd". The main content area displays the following information:

Bearbetad Nej
Bearbetning startade 2014-06-27 14:50:03
Bearbetning avslutades 2014-06-27 15:14:06
Arkivpaket (AIP) [aip.mets.metadata](#), [Ladda ner: um:nbn:se:kb:dark-package-instance-629212](#)

Andra paket med samma resursnyckel

Id	Externt paket-id	Typ	Status	Leverans
629140	http://www.dn.se/nyheter/sverige/blixtnedslag-pa-bravallafestivalen-flera-skadade/			16375.187

Fillista

Filnamn	Storlek	Status	Statusbeskrivning
http://www.dn.se/nyheter/sverige/blixtnedslag-pa-bravallafestivalen-flera-skadade/	199,97 kB		
http://www.dn.se/nyheter/sverige/blixtnedslag-pa-bravallafestivalen-flera-skadade/.rss.metadata	2,02 kB		

Version: 2.2.39

Systems, methods and organization - 5

The Oden interface will be developed further:

- more sophisticated report tools based on e.g. statistics about how much each publisher is expected to deliver
- the possibility to trigger alarms if the expected amount of material changes significantly
- more advanced viewing system for the content - more of a presentation system for the material (perhaps the first step towards an interface for researchers and users)

Systems, methods and organization - 6

- In the beginning: a new and (in retrospect) understaffed separate e-legal deposit division (with technical support from the IT department)
- After a re-organization of the library the e-legal deposit work is more integrated in different divisions under Digital Collections and Physical Collections
- Development of the different systems and technical IT support handled by the Information Systems Department in dialogue with Collections
- Legal support through the Corporate Services Department

E-legal deposit metadata

A very limited set of mandatory metadata accompanies the delivered files:

- where and when the files are first made available
- the format in which the files are first presented
- codes to open password protected files
- the relationship of the material with other material delivered by e-legal deposit, such as the relative order of the files in an article
- the relationship between the delivered files and analogue material delivered by legal deposit

Future development of the legislation

The National Library is expected to report back to the government about the implementation of the e-legal deposit legislation. Possible changes are:

- the prescribed method of delivery: on physical carrier; default method should be over the Internet
- a better definition (based on experiences 2015-) in the legislation of the rather vague “enterprises professionally producing electronic materials”
- legal support for making the e-legal deposit material available

Conclusion

- What the library now is able to collect with the help of the e-legal deposit law is to a large extent the bits and pieces that make up web sites, without context or structure
- It is really a necessity to tie together the traditional web harvesting process with the archive of the more complete content to give a reasonable picture of what is published on the web.
- The new law is in many respects a good start and makes it possible for the National Library to start preserving also the electronically published part of the Swedish cultural heritage for future research and studies.