

*Digitisation of historic newspapers and voluntary
digital deposit of newspaper pre-print files in the
the National Library of Estonia*

Krista Kiisa
Digitisation Coordinator



Activities to be presented

- Digitisation of historic newspapers
- Harvesting newspapers from the web
- Voluntary deposit of newspaper pre-print files deposited by newspaper publishers



Workflow for current newspapers – why do we need pre-print files?

- Legal Deposit Act doesn't apply to the electronic pre-print files in Estonia
- Because of reduced budget, microfilming the current newspaper volumes stopped during the time of economic slowdown
- Content published on the web and the content on paper are two totally different things



Negotiations with newspaper publishers voluntary deposit of pre-print files

- As a starting point, the biggest daily newspaper *Postimees* was contacted in the end of 2007
- Positive feedback - all pre-print files, up to the beginning of 2006 were retrospectively sent to the NLE's FTP server
- volumes gathered retrospectively need extra manual work (file management, additional efforts for sorting)



← Tagasi otsingusse

Ajalehed (8)

Postimees ; 2010

| Ilmumisaeg: 2010

Postimees ; 2009

| Ilmumisaeg: 2009

Postimees ; 2008

| Ilmumisaeg: 2008

Postimees ; 2007

| Ilmumisaeg: 2007

Postimees ; 2006

| Ilmumisaeg: 2006

Salvesta Trüki Kontrast Vähenda Suurenda 50% Täiskuva Kuva 2 Indeks Pööra TagasiEdasi Lk Märjista Otsi

25



Postimees



Tartu Postimees

2007

< Detsember

E	T	K	N	R	L	P
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

Informatsioon

Postimees ; 2007

| Kirjastaja: Postimees
| ISSN 1406-0981

| Laad : ajaleht
| Ilmumisaeg : 2007

| Kogum : digar
| Keel : et

Vaata veel

| Postimees : Eesti päevaleht

Märksõnad

| ajalehed

Teenused

| Tagasiside

| Abi

Painful process of sorting the data gathered from publishers

- limited IT support for a period of time
- deposit of pre-print files took place absolutely irregularly – no expected time or structure for automated script-based archiving
- example of the leading newspapers' behaviour, crash of one's publishers server (leading to the loss of data) and coming Legal Deposit Draft Act - starting point for negotiations



Appetite grows by eating

We wanted:

- more titles
- content to be deposited not some time, not after some reminders, but early in the morning
- change the inhouse workflow - compile the Database of Estonian Articles (<http://ise.elnet.ee>) from pre-print files and not from extra paper originals, we had to buy



2012 – negotiations with Estonian Newspaper Association (EALL)

- EALL – a non-profit organisation working in the common interests of newspapers.
- Unites 40 newspapers published in Estonia
- Daily circulation ~ 510,500 copies
- EALL's interest - run a business model and involve private media monitoring companies and NLE into the same value chain



The aim of NLE:

- Archiving newspapers in its digital archive DIGAR (<http://digar.nlib.ee>)
- Flexible user access
- Arrange inhouse workflow of cataloguing the articles to the Articles Database from pre-prints
- Save money from microfilming and from further digitisation – using digital pre-print instead of them



The aim of EALL:

- archive all it's newspapers in one safe place
- have right to get the original deposited pre-print files back any time the publishers need
- use NLE's server as intermediate station where licenced media monitoring companies can download pre-print files every morning

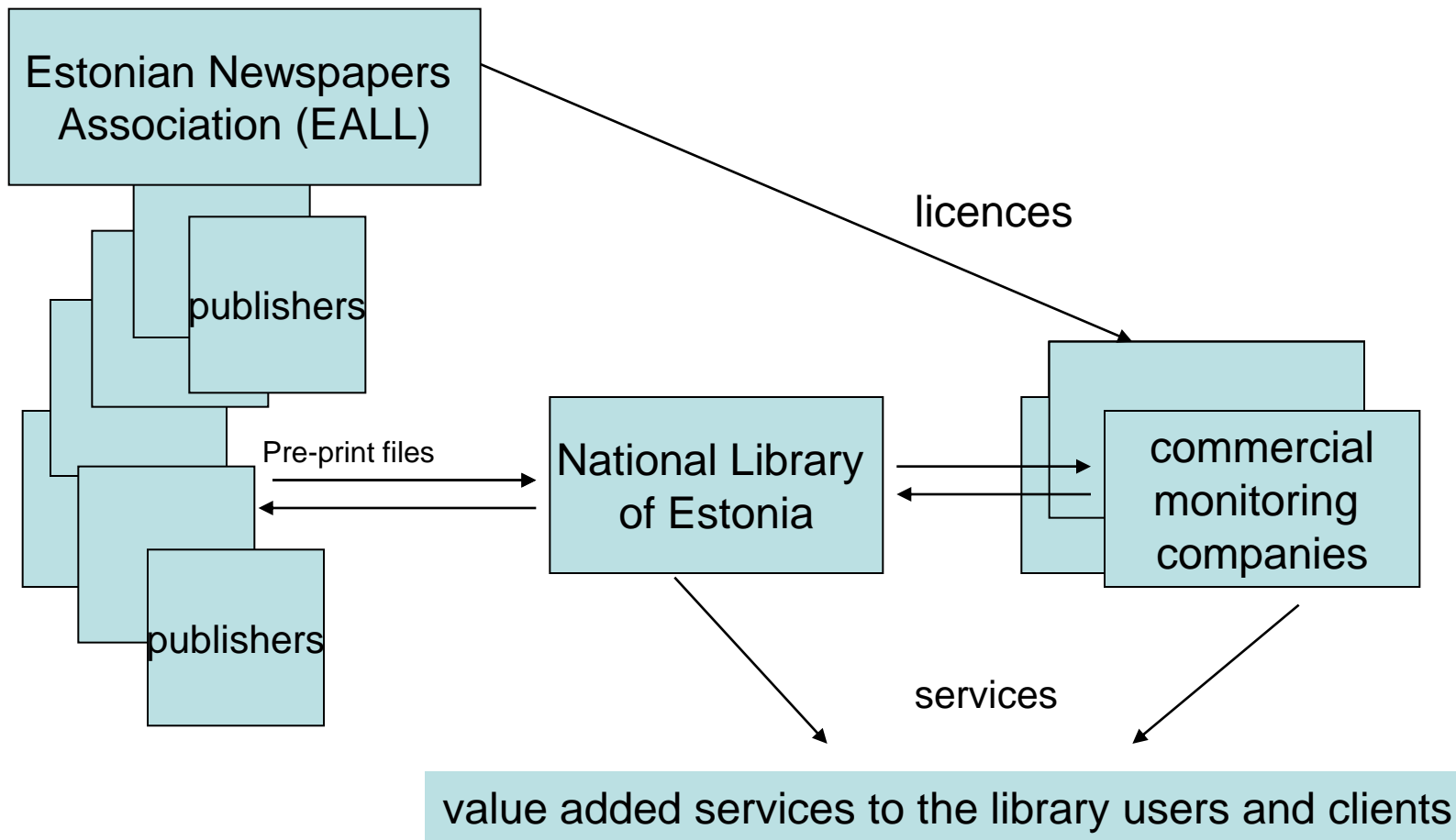


Newspaper publishers interest:

- to have safe place for archiving their content
- possibility to get original files back any time they need
- reduce preservation costs and get more visibility to their trademarks
- right to define the access restrictions in digital archive according to their business interests



Value Chain



Statistics

- 41 publishers
- 110 titles (makes ~99% of production in Estonia)
- 342 titles of so called small scale newspapers are in the waiting list (newspapers of schools, churches, institutions, organisations, different societies and companies...)



Harvesting newspapers from web

- activity based on Legal Deposit Act (2006)
- selective method used
- criterias fixed by the Web Archiving Experts Group (advisory body comprised of members from leading memory and research institutions)
- access to archived websites is open unless right holders impose a restriction
- complete harvest of the estonian web – 2014
- online newspapers will be harvested according the frequency of publishing



Digitisation of historic newspapers

- microfilming 1993-1998
- digitisation from microfilm starting 2003 (the only complete collection in Estonia)
- selection based on demands of long-term preservation (priority to items which are physically in bad condition)
- advantage - out of copyright material
- agreements with publishers
- access via image database <http://dea.nlib.ee>



- 380 titles
 - 1,3 digitised pages
 - 300 single visits
- per day

The screenshot shows a web browser window with the address bar displaying "http://dea.niib.ee/index.php?list". The page content is a list of digitized newspaper titles, each with a list of years and a link to "Isainfo" (Full information).

Title	Years	Link
Oma Maa : teaduste ja tuntuude ajakiri (Tartu : 1884-1891)	1884, 1885, 1886, 1887, 1888, 1889, 1890, 1891	Isainfo >
Oma Maa (Viljandi : 1922-1940)	1922, 1923, 1924, 1925, 1926, 1927, 1928, 1929 1930, 1931, 1932, 1933, 1934, 1935, 1936, 1937, 1938, 1939 1940	Isainfo >
Õpetajate Leht (Tallinn : 1930-1940)	1930, 1931, 1932, 1933, 1934, 1935, 1936, 1937, 1938, 1939 1940	Isainfo >
Otepää Teataja (1932-1944)	1932, 1933, 1934, 1935, 1936, 1937, 1938, 1939 1940, 1941, 1942, 1943, 1944	Isainfo >
Paide Teataja (Paide : 1900-1912)	1900, 1901, 1902, 1903, 1904, 1905, 1906, 1907, 1908, 1909 1910, 1911, 1912	Isainfo >
Pealinna Teataja (Peterburi : 1910)	1910	Isainfo >
Pealinna Teataja (Petrograd : 1914-1917)	1914, 1915, 1916, 1917	Isainfo >
Pealinna Teataja (Tallinn : 1933-1939)	1933, 1934, 1935, 1936, 1937, 1938, 1939	Isainfo >
Pemause Zetung (Pärnu : 1869-1918)	1897, 1898, 1899 1900, 1901, 1902, 1903, 1904, 1905, 1906, 1907, 1908, 1909 1910, 1911, 1912, 1913, 1914, 1915, 1918	Isainfo >
Perno Postimees ehk Näddalileht (Pärnu : 1857-1885)	1857, 1858, 1859 1860, 1861, 1862, 1863, 1864, 1865, 1866, 1867, 1868, 1869 1870, 1871, 1872, 1873, 1874, 1875, 1876, 1877, 1878, 1879 1880, 1881, 1882, 1883, 1884, 1885	Isainfo >
Peterburi Teataja (Peterburi : 1908-1914)	1908, 1909, 1911, 1912, 1913, 1914	Isainfo >
Petseri Teataja (Võru : 1921-1922)	1921, 1922	Isainfo >
Petseri Uudised : Petseri eestlaste häälkandja (Petseri : 1932-1937)	1932, 1933, 1934, 1935, 1936, 1937	Isainfo >
Põhja Kodu (Narva : 1919-1940)	1919 1920, 1921, 1922, 1923, 1924, 1925, 1926, 1927, 1928, 1929 1930, 1931, 1932, 1933, 1934, 1935, 1936, 1937, 1938, 1939 1940, 1941, 1942, 1943, 1944	Isainfo >
Põhja-Eesti (Narva : 1929-1940)	1929 1930, 1931, 1932, 1933, 1934, 1935, 1936, 1937, 1938, 1939 1940	Isainfo >
Põhjanael (Narva : 1906-1907)	1906, 1907	Isainfo >
Põllumees (Tartu : 1895 - 1912)	1895, 1896, 1897, 1898, 1899 1900, 1901, 1902, 1903, 1904, 1905, 1906, 1907, 1908, 1909 1910, 1911, 1912	Isainfo >

Europeana Newspapers project

- OCR and OLR (article segmentation) for scanned images – 600 000 images.
- Will be ingested to Europeana and NLE's coming newspapers portal
- hopes on crowdsourcing to compensate the poor quality, gained from microfilm scanned images



Until 2014

- <http://dea.nlib.ee>
- <http://digar.nlib.ee/digar/lihtotsing?m=s&l=ajaleht&q=ajalehed>



despite being “old news”, newspapers continue to
be at the forefront of digital library development

...

and sometimes driving them



THANK YOU!

Krista Kiisa

Krista.Kiisa@nlib.ee

