

Assembling Collections from Publishers and In-house Digitisation: The Estonian Example.

Krista Kiisa, National Library of Estonia

(Version 22.05.2012)

Exactly 9 years ago at the IFLA newspaper Conference, in Berlin, in August 2003, I started my presentation with statement that Estonian old newspapers are more than ever in danger of becoming unreadable and damaged as there has never been such an active usage of them as during the last 15 years. At that time very high attention was paid to newspapers by readers due to the fact that during the Soviet-time there was no free access to pre-Soviet newspapers in libraries at all. They were all either destroyed or sent to “special storage”, meaning closed to the public. That was the main reason newspapers were the very first collection, National Library of Estonia started to digitise in 2002.

Nine years later, nothing much has changed. Our daily life in the library is still concentrated around newspapers. Digitisation of newspapers is on-going. At the moment 1,2 million images are accessible via the old newspapers’ database DEA (<http://dea.nlib.ee>). NLE is a partner in the Europeana Newspapers project (<http://www.europeana-newspapers.eu/>), that aims to aggregate and refine newspapers for TEL and Europeana. The database of Estonian articles, *Index Scriptorium Estoniae (ISE)* (<http://ise.elnet.ee>), contains articles from newspapers, magazines, journals, and serial publications from the 1990s onwards, receives more than 200 new articles each day, allowing the full-text to be accessed. And a huge amount of currently published periodicals, both newspapers and journals are deposited by publishers to the digital archive DIGAR (<http://digar.nlib.ee>) daily at the same time as these go to print.

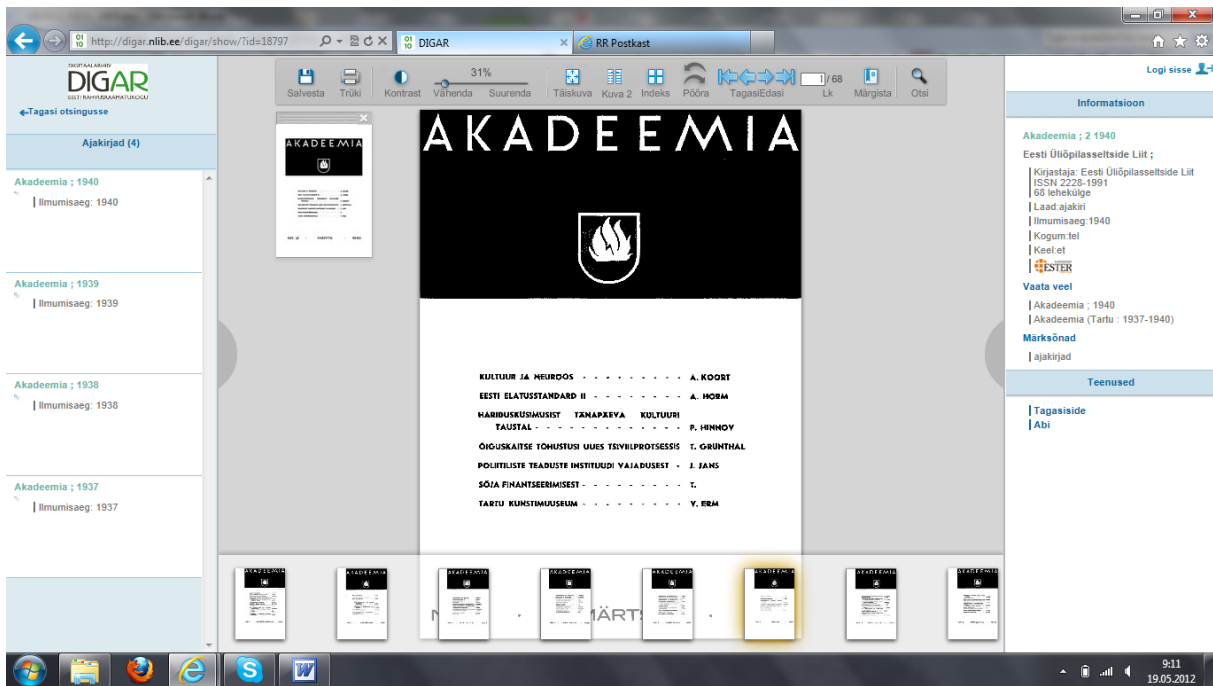
But what has changed in the last years is the attitude. Both parties, publishers and the library, have understood that in the situation of growing information flood they cannot manage without cooperation. The library is interested in providing better and quicker access to its collections, building e-services based on the information deposited by the publishers. By acquiring preprint files the library can reduce costs spent on microfilming, following the standards that original and archive files given to the archive by publishers are stored in file formats suitable for preservation. Publishers, at the same time, can reduce the cost for preservation. And benefit from library user interfaces, making links from their webpages to the collections in library’s digital archive site and getting much more visibility to their trademarks.

The current workflow of periodicals can be divided into two parts: publications digitised from the collections in libraries and publications that are deposited and submitted by publishers.

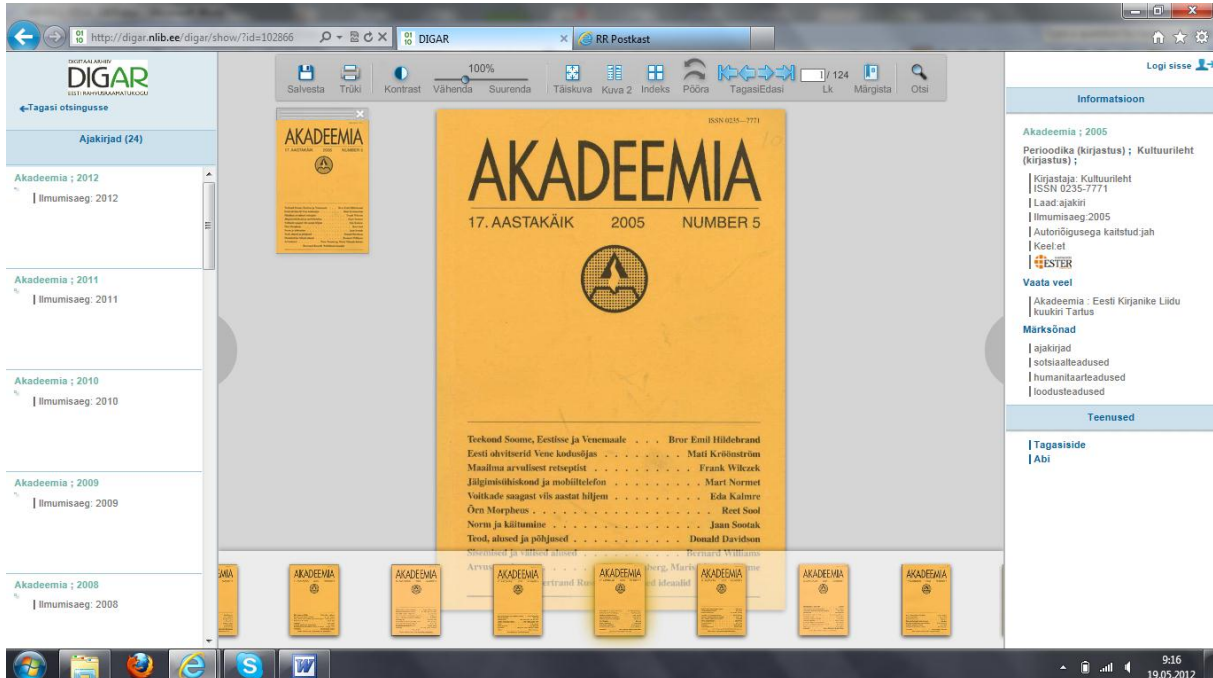
The inhouse work of digitising periodicals started with microfilming in 1993. It was a huge work, as no library in Estonia had a complete collection. The titles were gathered issue by issue to the National Library of Estonia. Microfilming took place in 1993-1998. Newspaper digitisation took place from microfilms and started in 2003. Followed by journals from 2006 which are digitised from paper original. The first selection was based on the responsibility of the National Library as the keeper of national heritage. Special attention was paid to the long-term preservation of periodicals that were physically in bad condition and hence in danger of

perishing. The starting point was out of copyright material, but as the red line was close enough for many publications under the digitisation, cooperation with right-holders and publishers came up as an issue quite soon. So agreements for microfilming and digitisation were signed with publishers. What concerns works published during the Soviet period, it has not been agreed whether the legal succession exists for the publishers of that period. It is unclear under which contract the authors worked for the corresponding publication. If these were employment contracts, the proprietary rights of the author should belong to the publisher; in case of other types of contract the permission of all authors should be asked. This would take a lot of time and work as every single issue would have to be dealt with separately. And even so it may not become clear. The biggest problem in digitisation projects is the identification of the authors, as well as finding out the authors' death date and successors. This problem especially arises with newspapers and periodicals as the number of authors, necessary to be cleared is huge. Usually these kind of materials have low commercial value but at the same time high cultural and historical value. The proposal for EU Orphan Works Directive does not seem to solve the problems with periodicals and embedded works. Before using the work, an institution needs to carry out diligent search. Diligent search for each article and photograph included in the newspaper is quite impossible. In Estonia there are no collective management organisations of writers, photographers or publishers that could help us finding the rights holders or represent them. One solution proposed to the problem would be creation of a authors' database. Described either on the basis of rights holders with details of their works or as a register of works currently identified as orphans in hope that rights owners will be encouraged to come forward to claim their rights. Efforts we are ready to take, clearing the copyright issues, is the confirmation that media still is the the tool that influences us most and we are ready to realise activities to gain unlimited free access to that material. Nothing has been decided yet in the National Library of Estonia, but we are holding the position, that if it's difficult to find the right holders, we rather make the works accessible free of charge in library digital archive and if the right owner comes out and asks to close it, we do that. Estonia is a land of free wifi. Why should we pay for using the cultural heritage in library?

But library is not the only one, making efforts for public access. From last years we have good examples, showing the interest of publishers, to give broader free access to their works. One example. We started the digitisation of journal „Akadeemia“, in the field of social science and humanities, covering years 1937-1940 at 2009.



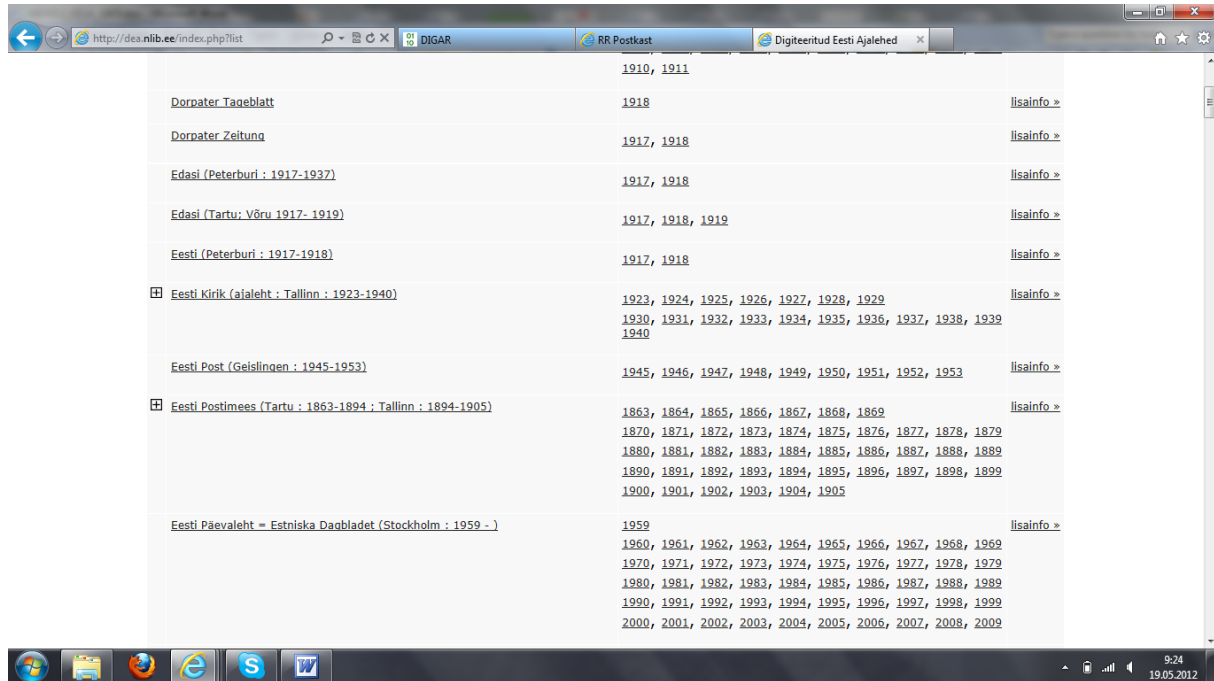
After going online with these years, the publisher, AS Kultuurileht, started to deposit electronically the issues, remaining in their editorial office workstations and we got access to the issues from 2009 - 2011. And then they decided to finance the digitisation of the missing years from e-collection of the journal. So we started the work for the years 1989 – 2009, free for users at the moment.



Now we already started work with the publishers second title called „Looming“ the very same way. It’s an example how publication will be made available to public and all of that absolutely free of charge for users.

But all these amounts of information need to keep and manage somehow.

It was obvious from the very beginning, that we need a coherent system for archiving. The first quick solution for newspapers was developing a digital image database, where a set of digital images is linked to the text database with needed bibliographical, administrative and technical descriptions. This can be still seen from the specialised newspaper portal at <http://dea.nlib.ee>



The digital archive DIGAR (<http://digar.nlib.ee>) was created 2004-2006, as a first attempt to preserve the growing volume of digital information. For that time being the main purpose was to archive the print-files of public sector institutions, collected for long-term preservation and wider access under the European Commission co-financed project reUSE. It was a good platform for getting acquainted with the publishers style of work, necessary to take into consideration in development of deposit module for the archive. In the first few years the work performed primarily to the archival processing of a quite narrow range of digital material.

In the situation where more and more information is published either only in digital form or in parallel digitally, the pressure increased for archiving information in different types and formats. More dedicated negotiations with the periodical publishers started. And now for some years already every morning servers of National Library of Estonia fulfil with daily newspaper and journal pre-print files, the publishers voluntarily deposit in our digital archive DIGAR. To do that, the agreement with National Library has to be signed beforehand, but as it is available electronically in digital archive's web page, registering as a depositor of the National e-Library can be done in the form of self-service, using the digital signature. Further authentication is carried out similarly to other state portals: by identifying persons based on their ID-cards, mobile id-s. Files can be downloaded via ftp or web upload with username and password given to the publisher.

There are no strict prescriptions written to the depositors for depositing purposes. We only expect them to keep to the agreed syntax of file names and structure of folders. The only thing we really ask them to avoid, is using the diacritics in file names. Objects from files and

metadata are generated using special scripts and are sent to the digital archive automatically. The objects in digital archive are linked to the library's OPAC, where more detailed metadata can be obtained. To avoid the duplicate digitisation and archiving, we have agreed with Estonian research libraries that use the same OPAC, that all digitised and e-publications must have a link in the electronic catalogue record. So there will be at least one place to find out if the item has been digitised in one of the Estonian libraries already and avoid duplicating the costs.

The screenshot shows the ESTER library OPAC interface. At the top, there is a search bar with the record ID 'b1071914' and a dropdown menu for 'Tallinna TÄIELIK kataloog'. Below the search bar, the main title is 'Akadeemia : Eesti Kirjanike Liidu kuukiri Tartus / SA Kultuurileht ; peatoimetaja Toomas Kiho'. The publication details include 'Tallinn : Kultuurileht, 1989-' and a URL 'http://tallinn.ester.ee/record=b1071914~S1*est'. A section titled 'LINK' contains the text 'TÄISTEKST RR arhiivis DIGAR AJAKIRJA SISUKORD'. Below this, there are three rows of metadata:

Asukoht	RR 3.k arhiivkogu
Viimati saabunud:	Aprill 2012 no.4
Kohaviit	eper
Asukoht	TLÜAR ajalehed ja eestikeelsed ajakirjad (kohalkasutus)
Leidumus	avariilil 2010,2011,2012
Viimati saabunud:	mai 2012 no.5
Asukoht	TLÜ Kunstide Instituudi raamatukogu
Leidumus	avariilil 2008,2009,2010
Viimati saabunud:	Detsember 2010 no.12

The browser address bar shows 'http://tallinn.ester.ee/search~S1*est?/b1071914/b1071914/1,1,1,B/856-b10719148FF=8d1,0,1,0' and the system tray shows the date '19.05.2012'.

Coming from publishers' voluntary deposits to library's activity of downloading the web publications it is important to mention that the amendment to the Legal Deposit Act entered into force on 01.06.2006 in Estonia and this places the National Library already under the obligation to collect and preserve web publications in addition to paper publications. That type of online publications are archived and made accessible in DIGAR as National Library's main principle is to enable the free access to it's collections, either in paper or in electronic form. Publishers who does not wish that his chargeable e-book is read for free on the Internet, have the right to set limitations for using the book. Otherwise the deposited electronic publications will be freely accessible via the Internet. If permission for public access via the Internet is denied, the deposited web publications will still be accessible via authorised workstations in the premises of National Library. Such workstations are only meant for searching and reading the material, copying, downloading or printing are not enabled. If the publisher does not wish to set so strict limitations, it is possible to provide access to the materials also in all reading rooms via library's inhouse computer network.

Creating access to digital information must be based on the rights and restrictions applied to publications. Users are divided into groups with different rights. Depositors and publishers have special rights to the original files sent to the archive. Employees of the National Library have rights and obligations related to the processing, converting and backing up of the files. Registered readers have more rights in accessing the materials acquired by the library. Users, reading free access materials, do not have to register in order to view the content of the

archive and browse the material with no usage restrictions.

The new role-based system for the administration of the e-library users rules out the duplication of data in the library's different information systems (DIGAR, ESTER, ISE, search portal, licensed databases, etc) and enables to administer all users from one place. Thus, different e-services can be used and it is possible to switch from one e-collection to another using one single authentication process. The new system enables publishers and other content producers to operatively apply usage restrictions on their publications and be sure that the restrictions are monitored by the library.

It is possible to search information from archived materials by either a simple or a complex search. There exists also a possibility to search from full texts or metadata. Searches can be made and results can be sorted by the types and collections of publications. Depending on the type of publication, search results can be viewed as thumbnails, by file data, metadata or as parts of objects.

In May 2012 the digital archive DIGAR provides access to 226 titles of journals. 81 of them are deposited by the publishers as pre-print files. The number of newspapers in digital form is 335 titles. 180 of them are deposited as pre-print objects by publishers. Plus there is a separate old newspaper image database with its 326 titles of newspapers from 1823-1940.

What next. The further development of the digital archive DIGAR will now primarily be focused on improving the functionality of the archive. It is necessary to develop a system for monitoring the correspondence of metadata to the digital objects required to ensure long-term preservation. The deposit copy will be applied to all publications in 2013. The volume of archived materials will rise enormously then. It makes necessary to pay more attention to checking the compliance of different file formats in the file handling software, develop different converting scripts, follow authenticity of the presentation of digital files, implement faster access solutions etc. The back-up capacity of the digital archive and its technical readiness to ensure the availability of information also require ongoing development. All this for better availability of our national heritage.