

All together now – Digitization and use of Swedish-language newspapers

Tuula Pääkkönen

MSc., Information Systems Specialist

National Library of Finland / Research Library / Digital Humanism group

Digitization and deployment of Swedish-language newspapers

- Project to digitize all Swedish-language newspapers in Finland
- Years 2021-2023
- About 6 million pages made available



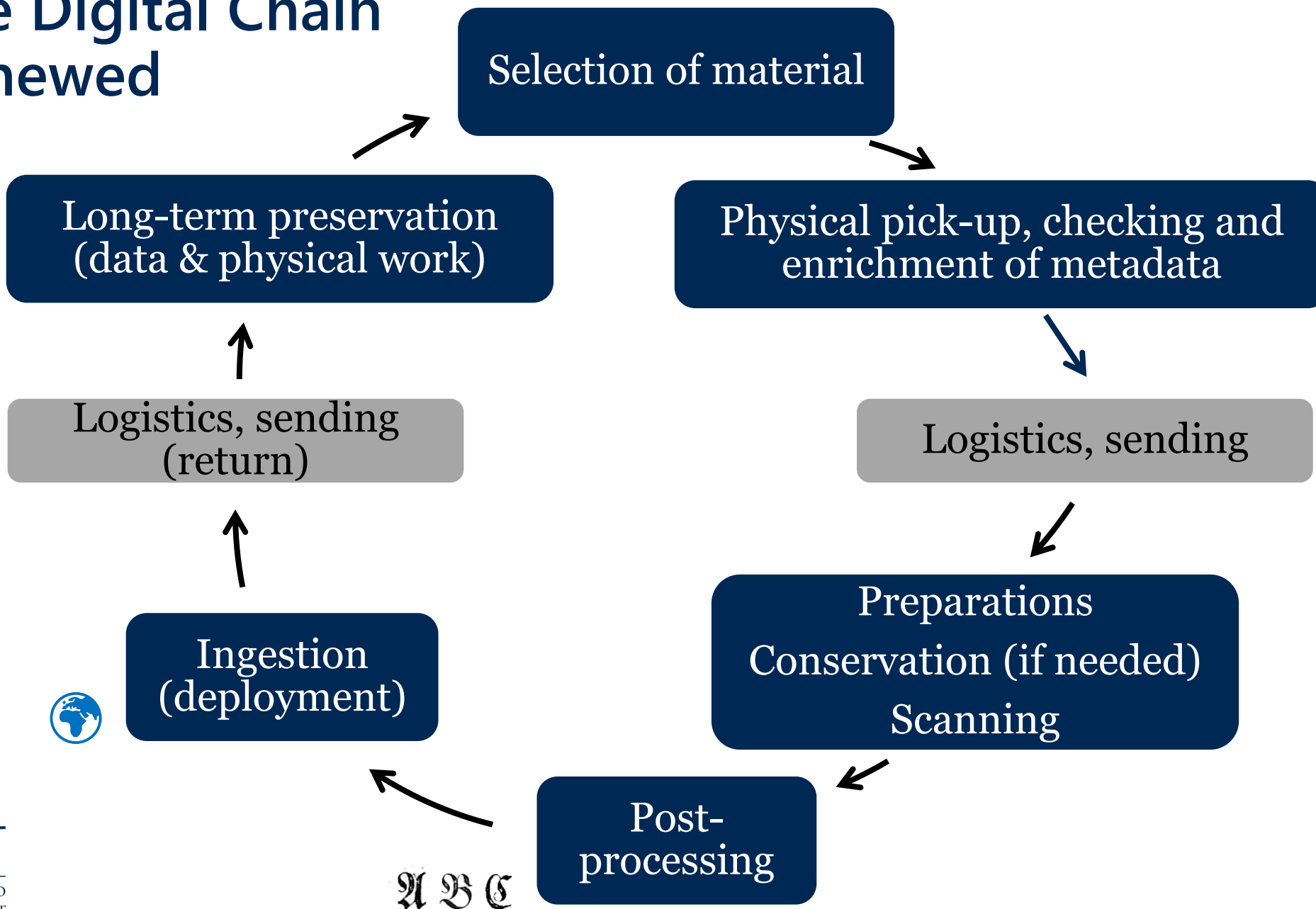
KONSTSAMFUNDET



STIFTELSEN
Tre Smeder



The Digital Chain Renewed



It all begins at selection

- List of titles, where language was Swedish
 - Or combined Swedish & Finnish
- Based on National Bibliography
 - Checking of metadata
 - Identifiers important
 - ISSN or National Bibliography Number
- Digitization order based on title name
 - A->Ö
 - (except if there was something complex in the title)

Titel	fk / ISSN
Aftonposten	fk10006
Annonsbladet för Kimito Bygden	0782-7105
Annonsbladet för Kimito, Dragsfjärd, Västanfjärd och Hitis	fk11010
Annonsbladet för Kimitoön	2343-5380
Arbetarbladet	0789-8274
Arbetarbladet	fk10000
Arbetarbladet-Svenska Demokraten	fk10001
Björneborgs Tidning	1457-4292
Borgåbladet	0358-6294
Dagens Tidning	2242-1009
Fiskarbladet	fk16621
Folktidningen	fk10057
Folktidningen Ny Tid	0357-1548
Folkviljan	fk10058
Hangö	1458-0764

Together in Collaboration

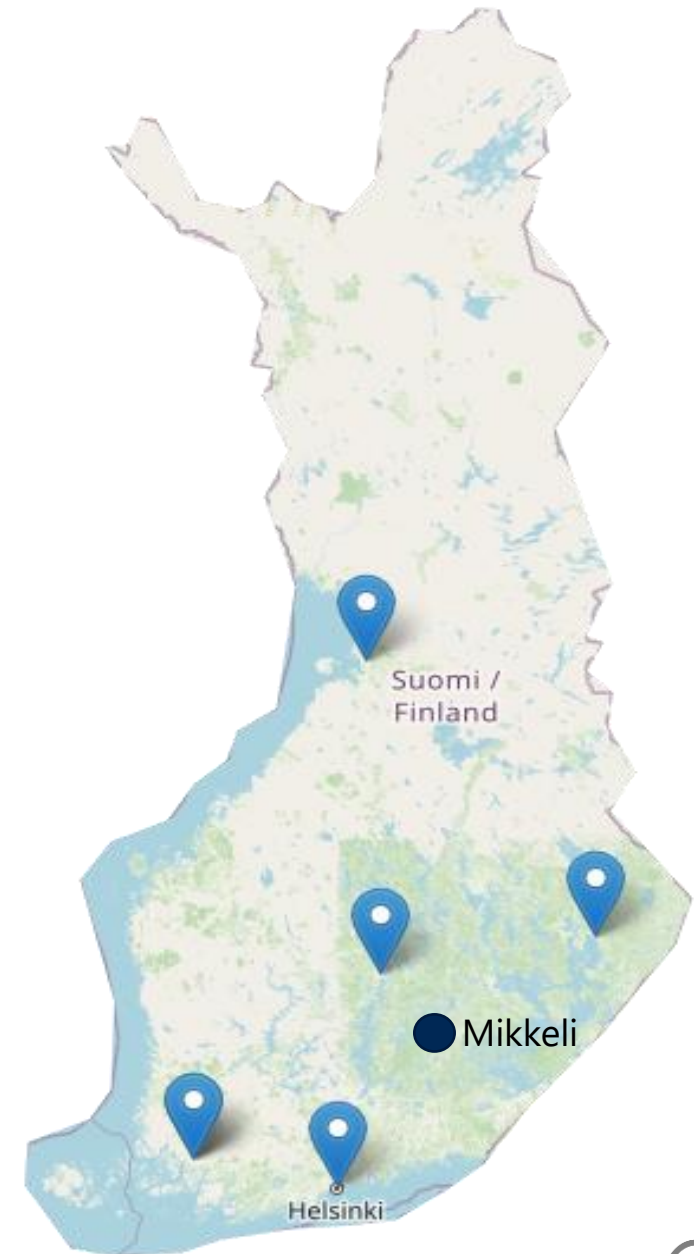
- Contract work and collaboration
- Materials opened with National Library, Svenska litteratursällskapet i Finland, Ålands landskapsarkivet, Pressarkivföreningen & Kopiosto
 - Usage within premises of archive locations
 - Generic public use
- University of Helsinki project model in use
 - Reporting, follow-up throughout



- Svenska kulturfonden
- Svenska litteratursällskapet i Finland
- Föreningen Konstsamfundet
- Stiftelsen Tre Smeder
- Stiftelsen Brita Maria Renlunds minne
- Stiftelsen för Åbo Akademi
- Svenska folkskolans vänner
- Harry Schaumans Stiftelse

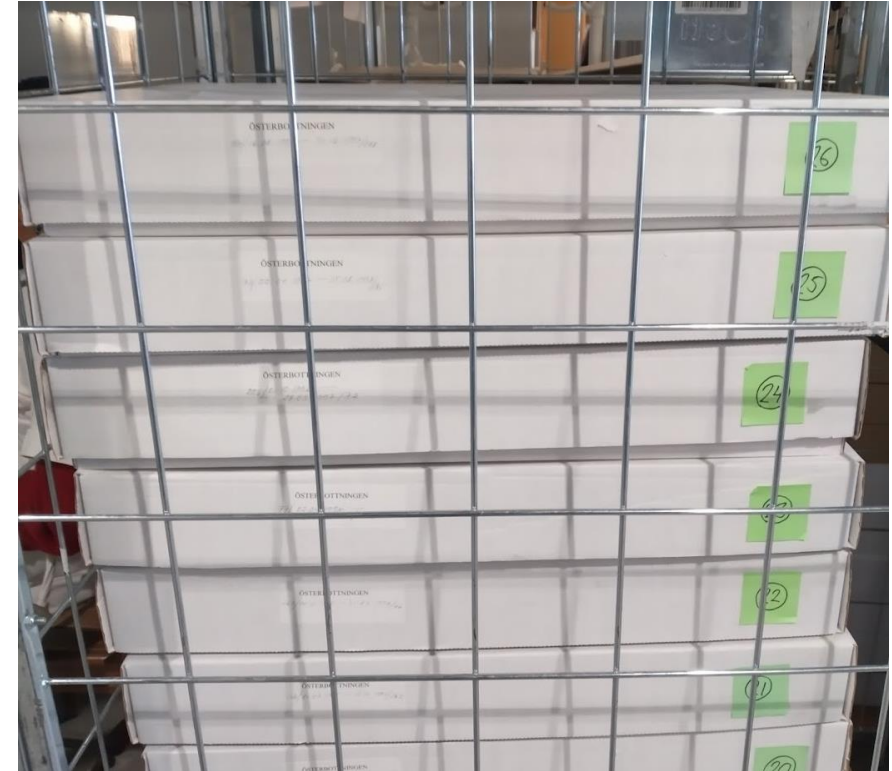
Physical and Digital space

- Material arrives to Mikkeli for digitization
- Works are tracked throughout the process
 - Preparation
 - Scanning
 - Repacking and returning
 - Quality control
- Defining of digitization order <-> equipment & personnel availability



Local logistics

- Moving material from storage to processing
 - Unbinding bound volumes
 - Page cutting
 - Repacking after scanning
-
- Huge volumes in this digitization effort!





Two Supag scanners and digitizer at work

Post-processing

Several background servers that run automatically 24/7 for:

- Layout analysis (text paragraphs, illustrations)
- Text recognition
 - Several languages
- Importing to [Digi.kansalliskirjasto.fi](https://digi.kansalliskirjasto.fi)
 - Database
 - Search index



Long-term preservation & research

- Material delivered to CSC (IT centre for science) for long-term preservation, 500+ years
 - digitalpreservation.fi
- Currently project ongoing called FIN-Clariah, which enables use of materials within CSC high-computing environment
 - Copyright-free material
 - Collaboration with other cultural heritage organizations
 - Both metadata and contents
 - Additional data package and metadata development for data packages in planning
 - <https://www.kielipankki.fi/organization/fin-clariah/>

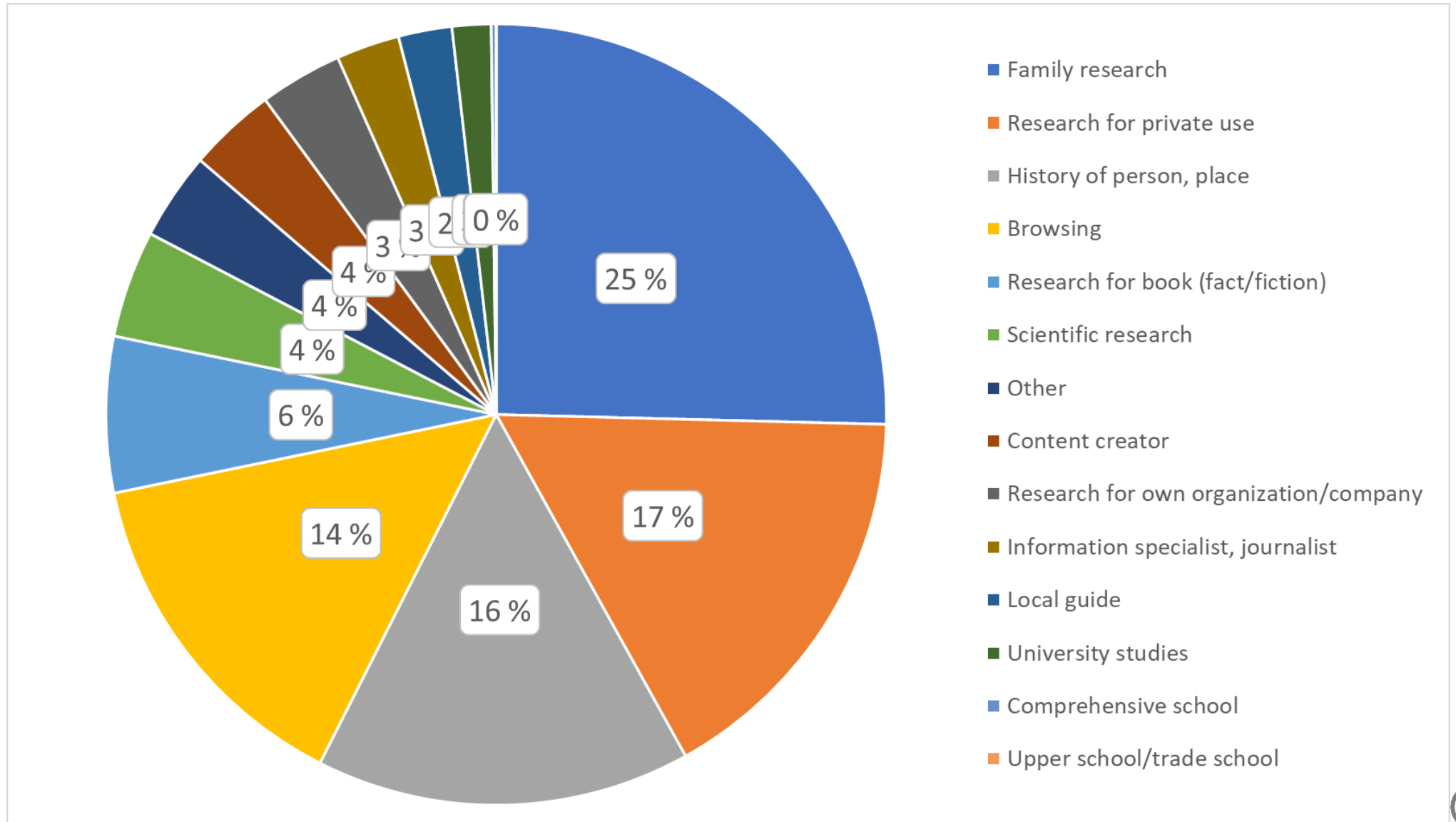


Dall-E, AI generated image
Prompt: "is anybody there future"



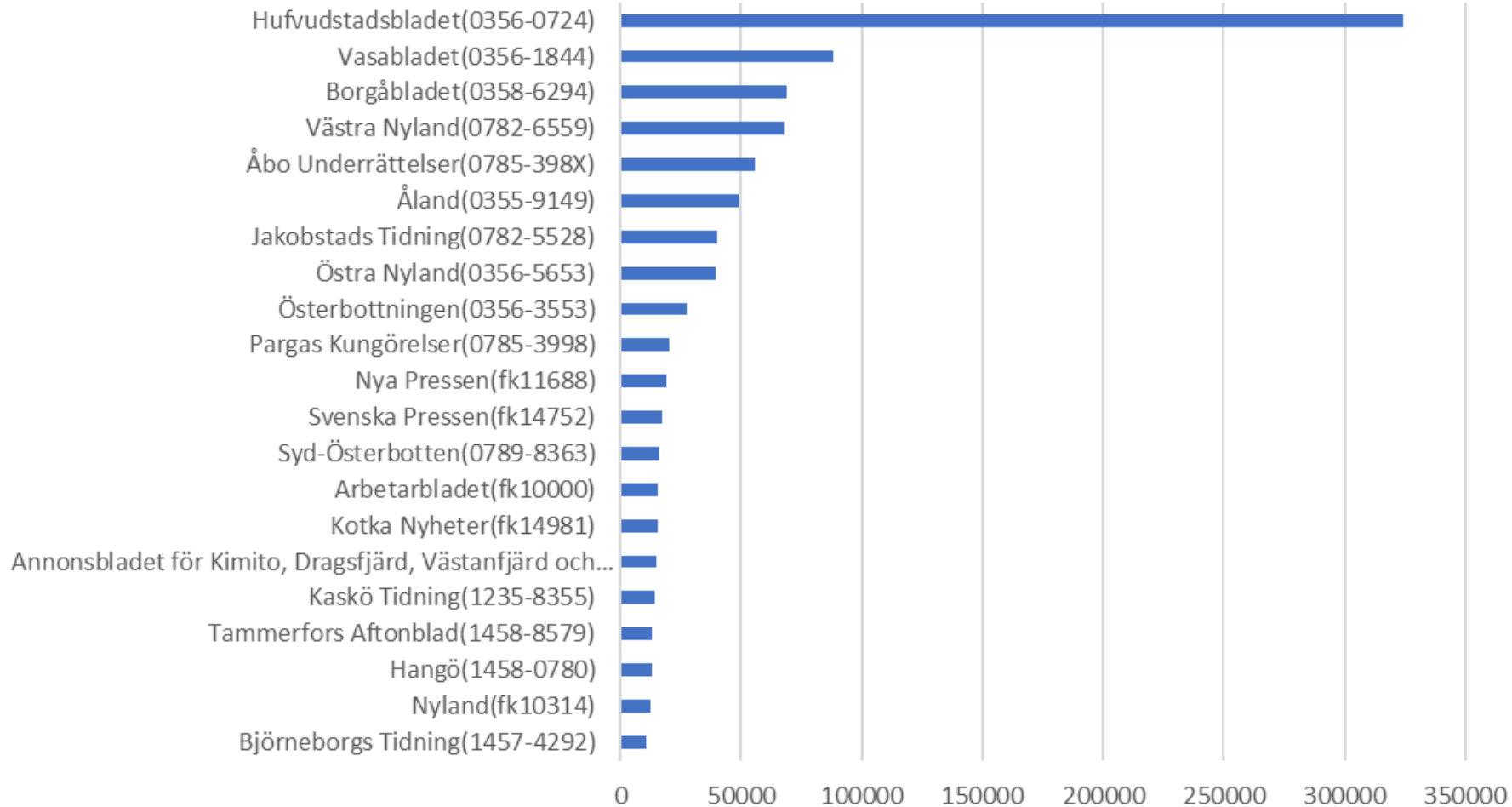
Use in digi.nationallibrary.fi

User survey 02/2022, n=202, usage purpose

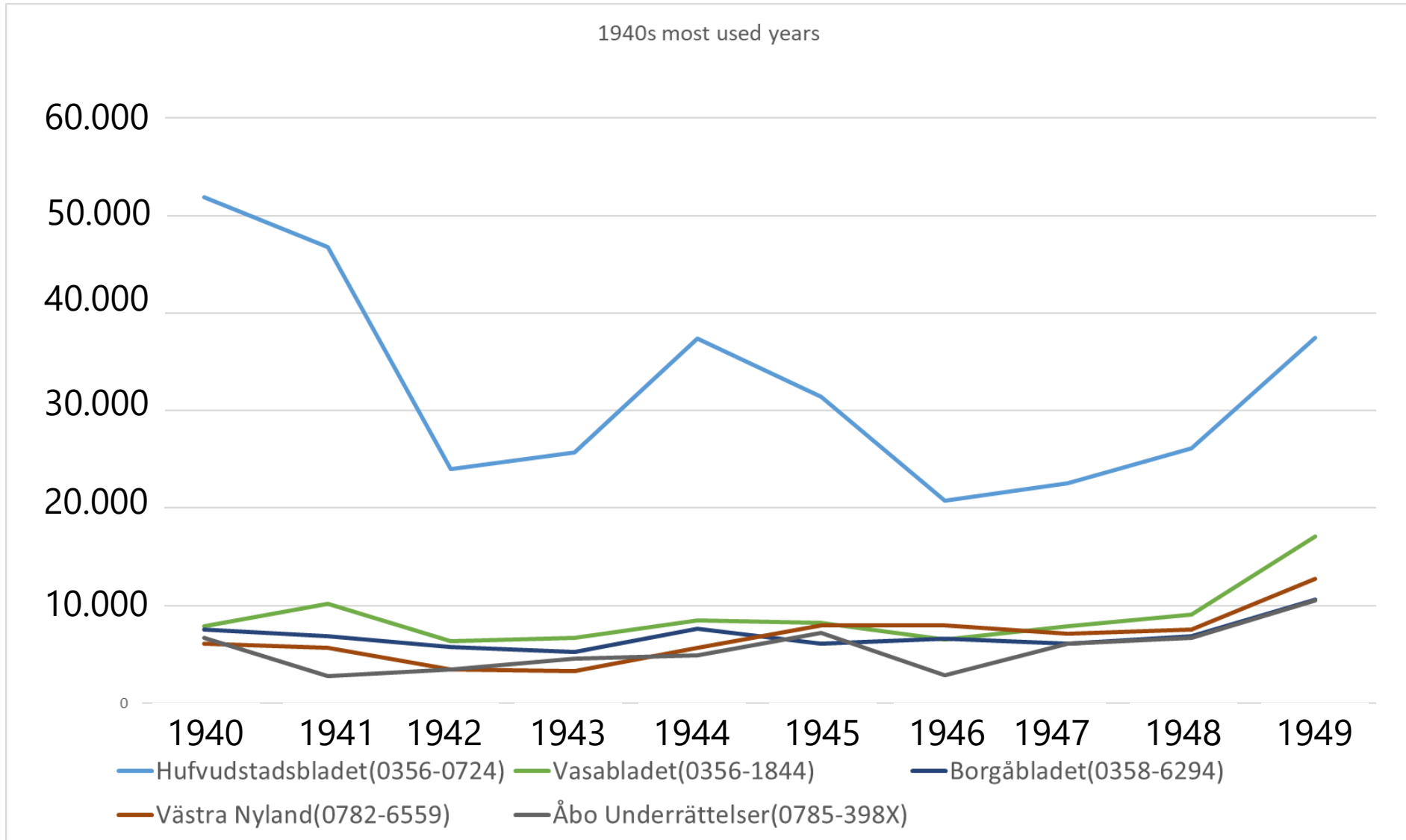


Over 10.000 page views, years 1940-1949

Swedish-language newspapers, usage 1940-1949 issues



Top-5 newspapers 1940 & 1944 most read (years 2020-2023)



Free for research, for many fields of research

9 949 pages with search word

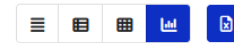


9849 sidor 200 delar



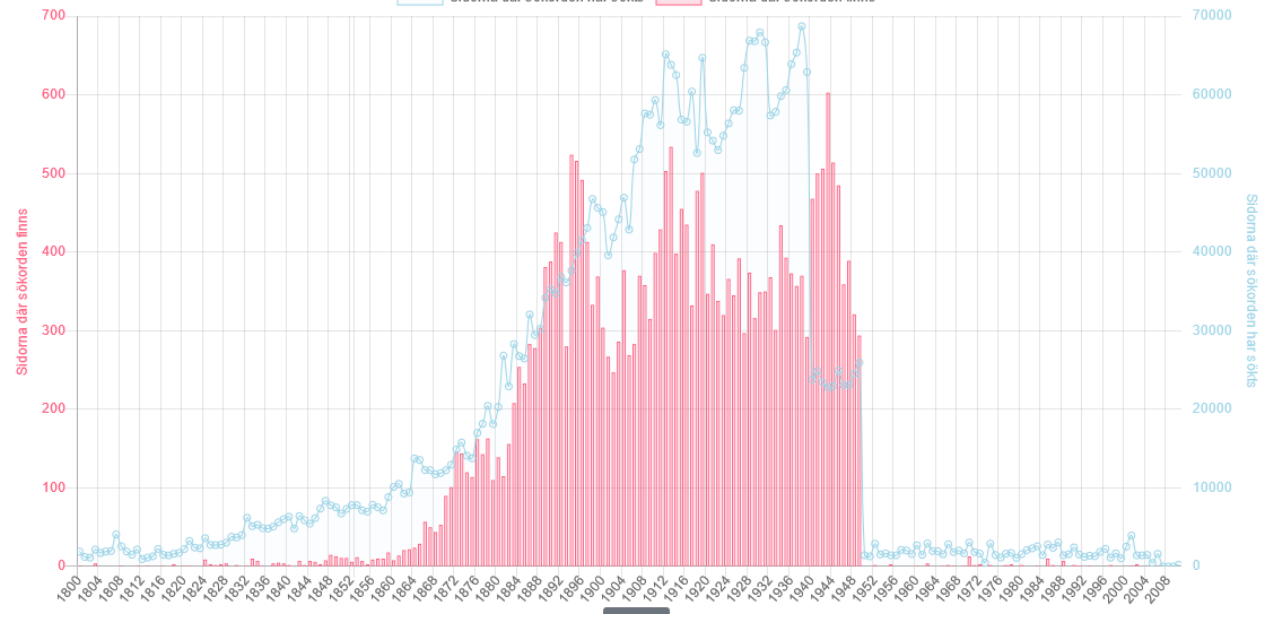
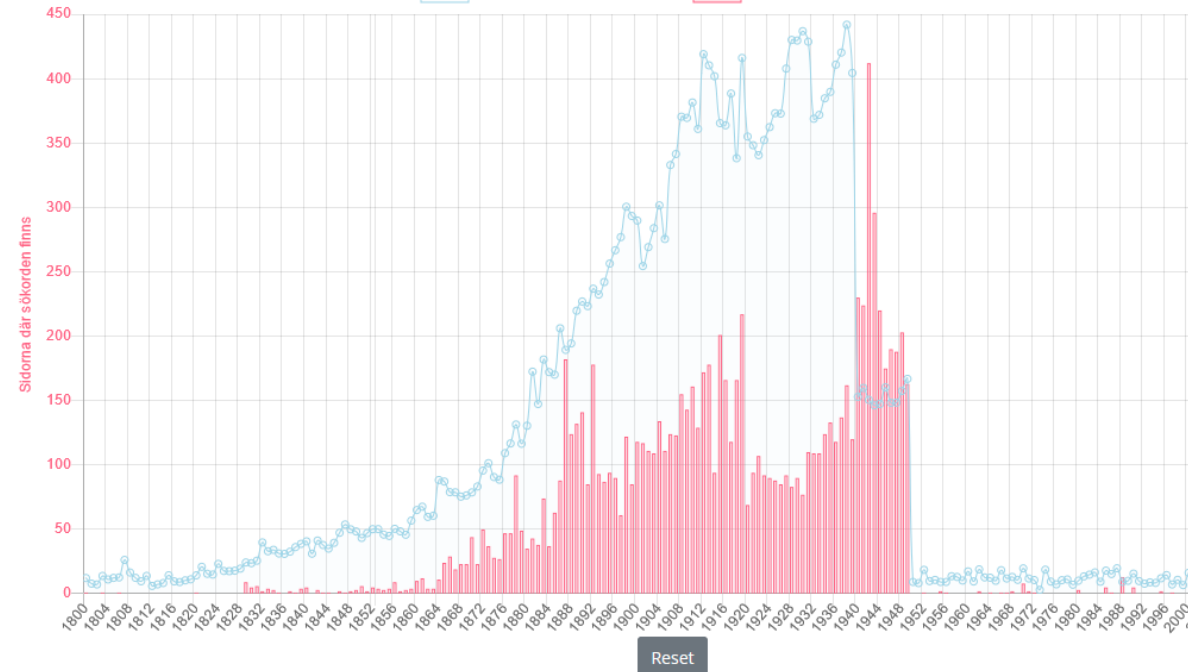
27491 sidor 396 delar

27 491 pages w search



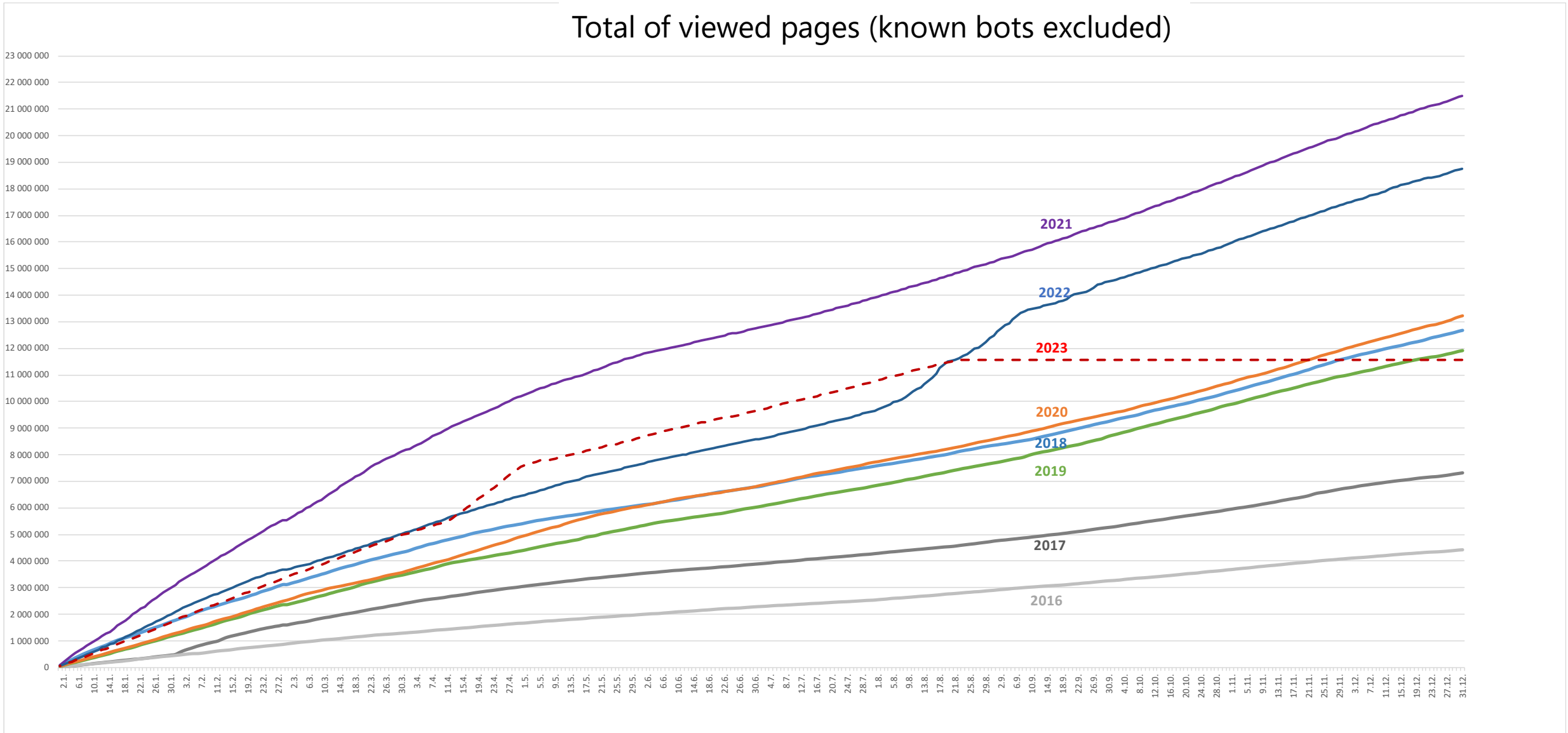
Sidorna där sökorden har sökts Sidorna där sökorden finns

Sidorna där sökorden har sökts Sidorna där sökorden finns



Thank you for interest!

Total of viewed pages (known bots excluded)



Summary

- Good material for use by public and researchers
 - First years promising
 - Solid asset for future
- Future-proof : good quality images basis for future processing and research needs
- All together : large, consistent collection utilized and available
- Long-term preservation
 - Long , consistent issue runs now available



www.kansalliskirjasto.fi

All material available at: Digi.kansalliskirjasto.fi