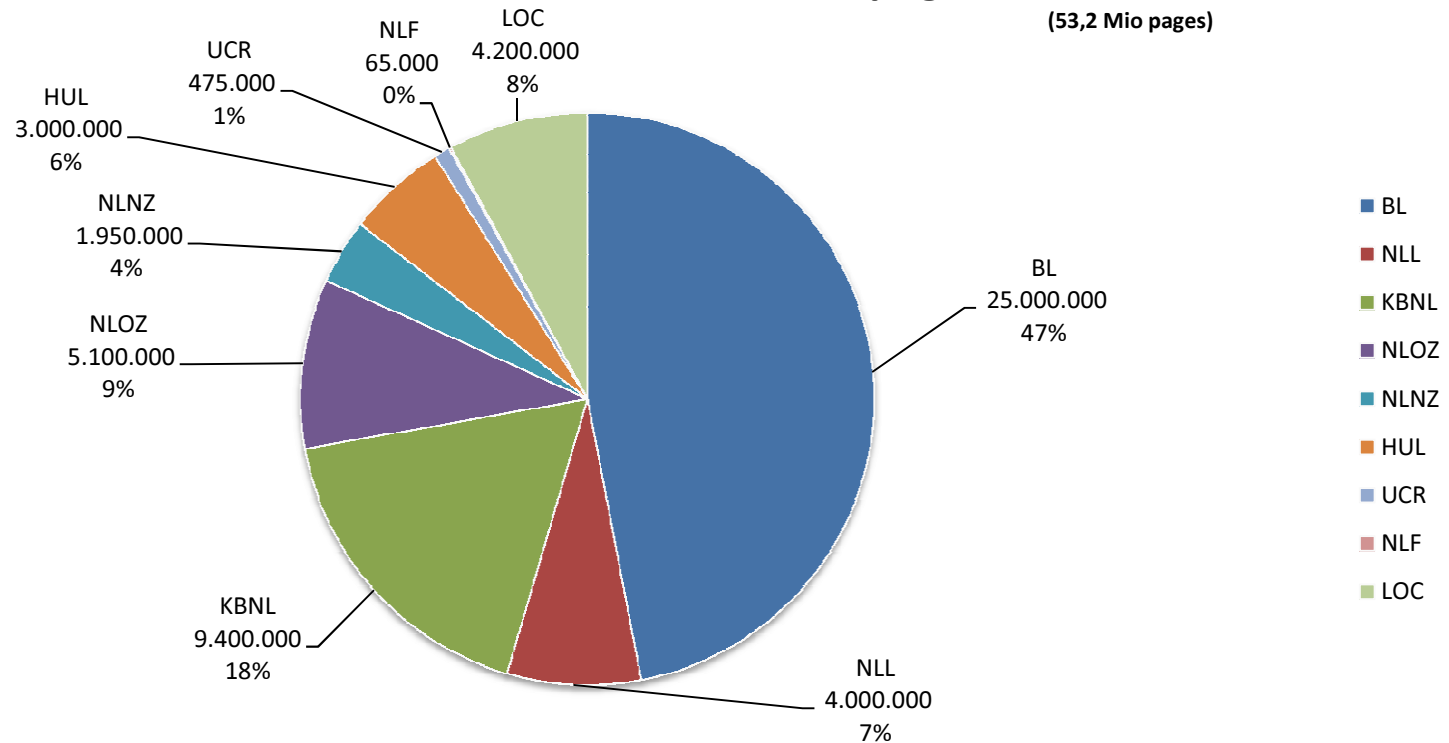
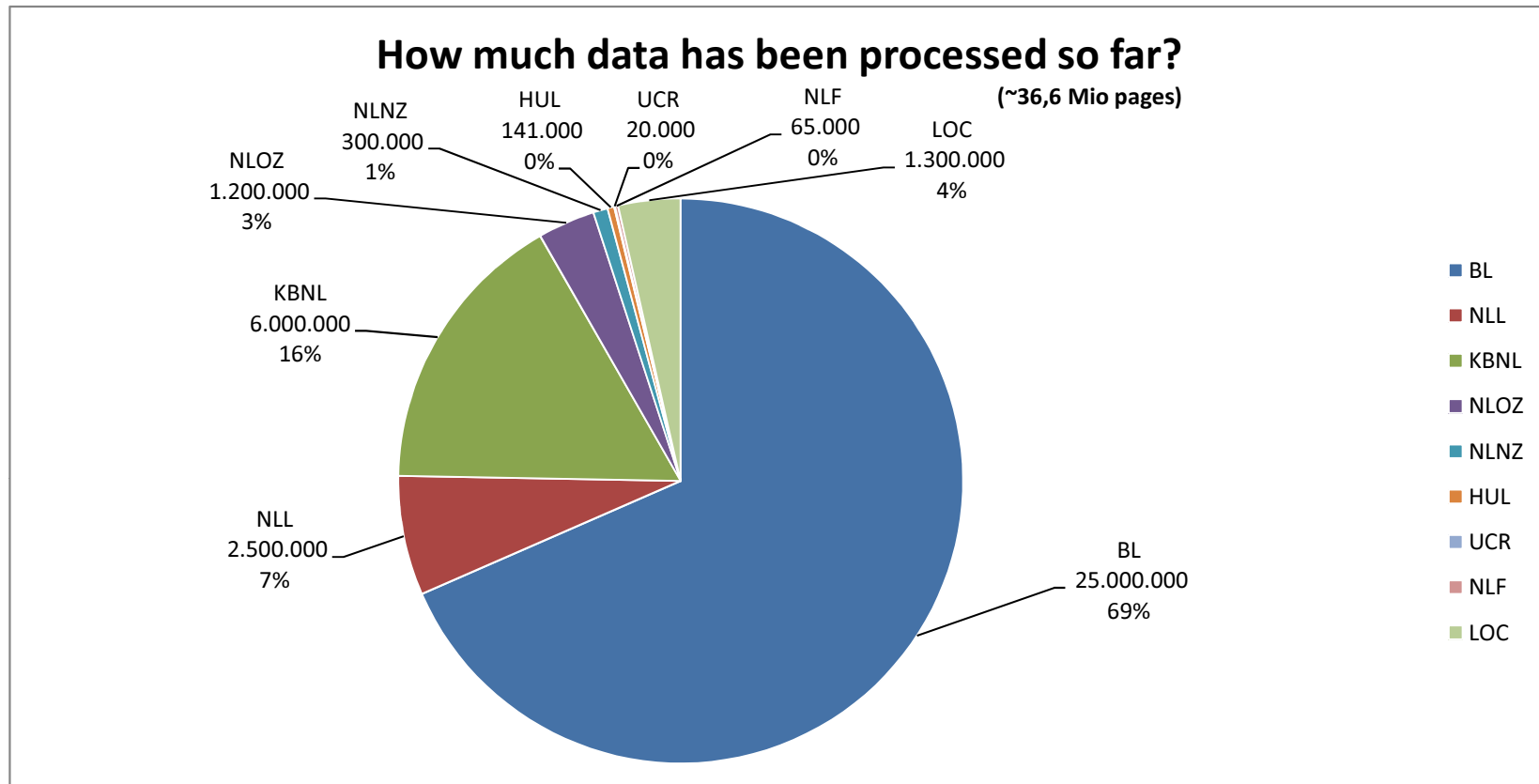


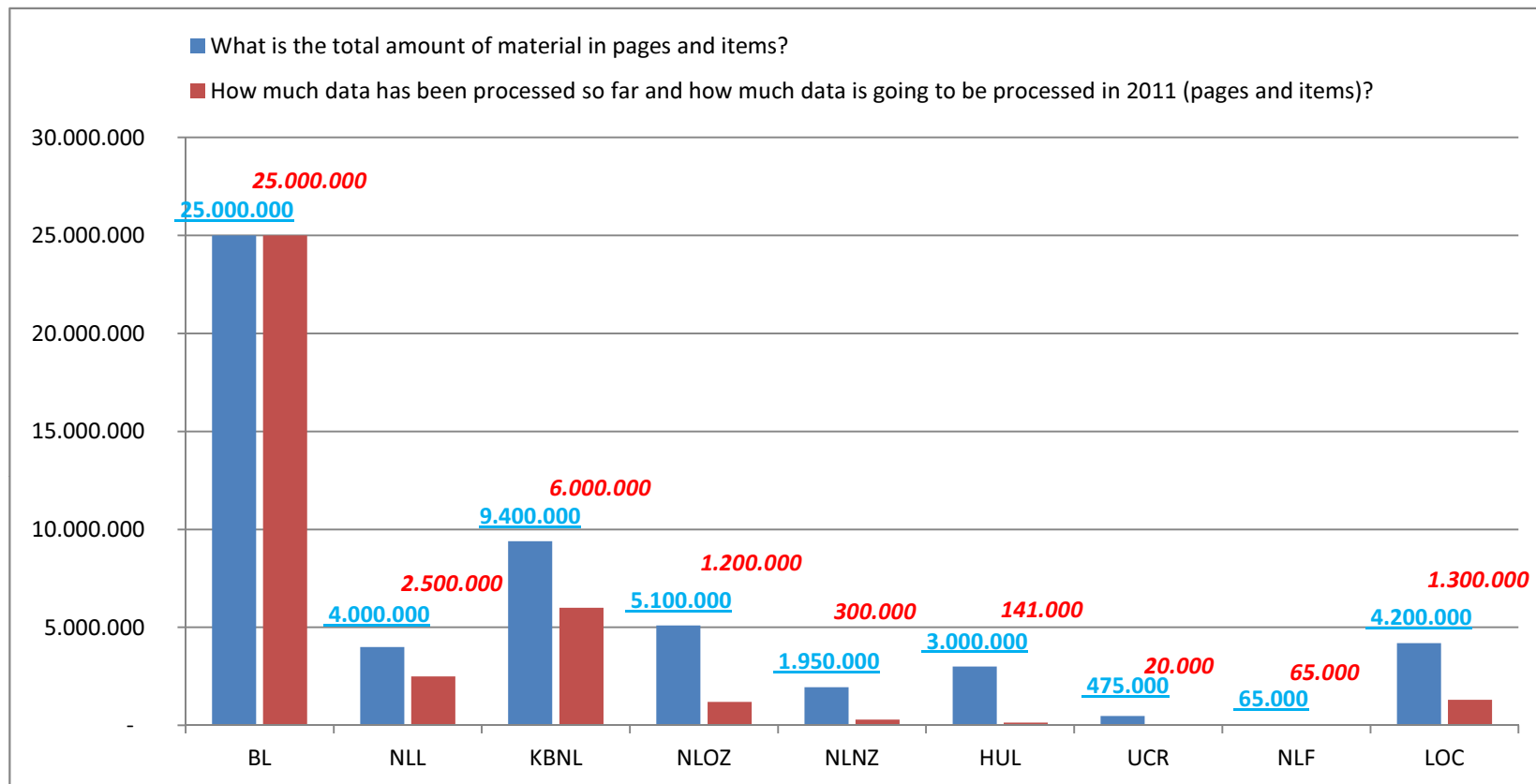
| Question  | BL                         | NLL   | KBNL  | NLOZ                                       | NLNZ                              | HUL   | UCR  | NLF  | LOC   | Sum        |  |
|---|----------------------------|---|---|--|-----------------------------------|---|--|--|---|------------|--|
| Organization  | British Library            | National Library of Latvia  | National Library of the Netherlands                             | National Library of Australia              | National Library of New Zealand   | Harvard University Library  | University of California (Riverside) Library | National Library of Finland  | Library of Congress   |            |  |
| What is the name of the digitization initiative or project?   | Microsoft Live Book Search | ERDF co-financed Project “Development of digital library – 2nd round” | Databank Digitale Dagbladen (Databank Digital Daily Newspapers) | Australian Newspaper Digitisation program  | Papers Past                       | a combination of short and long-term projects and ongoing/ programatic digitization | California Digital Newspaper Collection      | Ephemera project   | National Digital Newspaper Program  |            |  |
| What is the total amount of material in pages and items?  | 25.000.000                 | 4.000.000   | 9.400.000   | 5.100.000                                  | 1.950.000                         | 3.000.000   | 475.000                                      | 65.000   | 4.200.000   | 53.190.000 | newspaper pages and 600 newspaper titles |
| When did the initiative or project start?   | 01.04.2007                 | 01.06.2009  | 01.01.2007  | 01.01.2007                                 | 01.01.2001                        | 01.01.2005  | 01.01.2005                                   | 01.06.2008   | 01.01.2004  |            |  |
| How much data has been processed so far and how much data is going to be processed in 2011 (pages and items)?               | 25.000.000                 | 2.500.000   | 6.000.000   | 1.200.000                                  | 300.000                           | 141.000   | 20.000                                       | 65.000   | 1.300.000   | 36.526.000 | plus 4500 items                          |
| What material is digitized (books, pamphlets, newspapers, maps, fotos, others, source type: originals, microfilm, digital)? | Books                      | Books, newspapers, magazines, scientific journals, maps               | newspapers  | newspapers, journals                       | newspapers                        | books, pamphlets, serials   | newspapers                                   | pricelists, leaflets, Nat welfare during Wartime/regulations and rules, Sports/rules, programmes, schedules, annual reports, Travel advertisements | Newspapers (microfilm)  |            |  |
| Was the scanning performed in house or outsourced?  | outsourced                 | outsourced  | outsourced  | outsourced                                 | outsourced                        | in house  | outsourced                                   | in house   | outsourced  |            |  |
| Was the digitization performed in house or outsourced?  | outsourced                 | outsourced  | outsourced  | outsourced                                 | outsourced                        | in house  | outsourced                                   | in house   | outsourced  |            |  |
| What software tools are used in this project?   | docWORKS                   | docWORKS  | docWORKS  | inhouse                                    | docWORKS                          | docWORKS, adobe photoshop   | docWorks, Veridian                           | docWORKS, Remote Item Tracking, ScanClient, Sharepoint   | docWORKS, in-house QA and other vendor tools  |            |  |
| What is the amount of data produced in the initiative/project (in TB)?  | 100                        | 50  | 500   | 76   | 9                                 | 40  | 20   | 1,5  | 500   | 1297       |  |
| What are the file formats the data is stored (text, meta data and image files)?   | METS/ALTO, JP2, PDF        | METS/ALTO, JPG, JP2, PDF  | ALTO, MIX, JP2, PDF, MPEG21                                     | METS/ALTO, TIFF, JPEG                      | METS/ALTO, TIFF, PDF              | METS, JP2, TXT  | METS/ALTO, TIFF, JP2, PDF                    | METS/ALTO, JPG, JP2, PDF   | METS/ALTO (MODS, PREMIS, MIX), TIFF, JP2, PDF   |            |  |
| Is the output data model based on a registered METS profile - if so, which one?   | No                         | METS 1.4+   | yes   | LOC  | n/a                               | Harvard profile   | LOC  | no   | LOC   |            | NDNP article-level                       |
| How is the data presented to the public? (PDF, Website, ...)  | PDF, Amazon free download  | website (2012)  | website   | website                                    | Website                           | website   | website                                      | website/PDF  | http://chroniclingamerica.loc.gov/  |            |  |
| Can you provide a link to a sample accessible in the web? If yes, please enter the link.                                    | no                         | no  | http://kranten.kb.nl/   | http://nla.gov.au/nla.news-article65199570 | http://paperspast.natlib.govt.nz/ | http://nrs.harvard.edu/urn-3:FHCL:2448900   | http://cdnc.ucr.edu                          | http://www.doria.fi  | Issue:<br>http://chroniclingamerica.loc.gov/lccn/sn85032923/1911-10-25/ed-1/seq-1/<br>Sample batch data:<br>http://chroniclingamerica.loc.gov/data/pst/batch_pst_carnegie_ver01/data/ |            | not yet, it's coming                     |
| Can you provide a sample of the preservation copy via ftp? If yes, please enter the URL.                                    | no                         | no  | no  | no   | no                                | no  | upon request                                 | upon request   | no  |            |  |

## What is the total amount of material in pages and items?

(53,2 Mio pages)







## What is the amount of produced data (in TB)?

(Total actually 1297 TB)

