DIGITALISERING AF 32 MIO. AVISSIDER PÅ 3 ÅR

Aalborg 19-11-2014

Niels Bønding
nieb@statsbiblioteket.dk

OVERBLIK

- Kort om Statens Avissamling
- Avisdigitalisering – baggrund
- Avisdigitalisering – hvordan
- Tilgængelighed – hvordan får man adgang?
- Spørgsmål?

Niels Bønding
nieb@statsbiblioteket.dk

KORT OM
STATENS AVISSAMLING

Niels Bønding
nieb@statsbiblioteket.dk

STATENS AVISSAMLING

Statens Avissamling flytter ind. Demokratens fotograf var så heldig at være til stede, netop da ældre årgange af Demokraten blev båret ind! Manden i lys kittel er avisforvalter A. Jacobsen. Demokraten 24.3.1918.

STATENS AVISSAMLING

- Oprettet 1916, åbnet 1918
- Samling opbygges i medfør af Pligtafleveringsloven
- Modtager 20.000 dagblade og 16.000 ugeaviser om året
- Samlingen omfatter i alt
  - 94.000 avisbind
  - 28.000 bind med distriktsblade
  - 66.000 mikrofilmspoler
  - 100 mio. sider i alt
  - 23.000 hyldemeter

Niels Bønding
nieb@statsbiblioteket.dk

DET NYE MAGASIN FRA 2007

Niels Bønding
nieb@statsbiblioteket.dk

DET NYE MAGASIN, INDEFRA

Niels Bønding
nieb@statsbiblioteket.dk

DET NYE MAGASIN, INDEFRA

Niels Bønding
nieb@statsbiblioteket.dk

DIGITALISERINGSPROJEKTET - BAGGRUND

Niels Bønding
nieb@statsbiblioteket.dk

VESTINDISK PAKHUS

- KB har 32. mio. avissider liggende i Vestindisk Pakhus fra 1780, som ikke længere kan bruges til magasin.
- Avisdigitaliseringspro-jektet finansieres ved at kassere disse aviser frem for at flytte dem til et nyt magasin

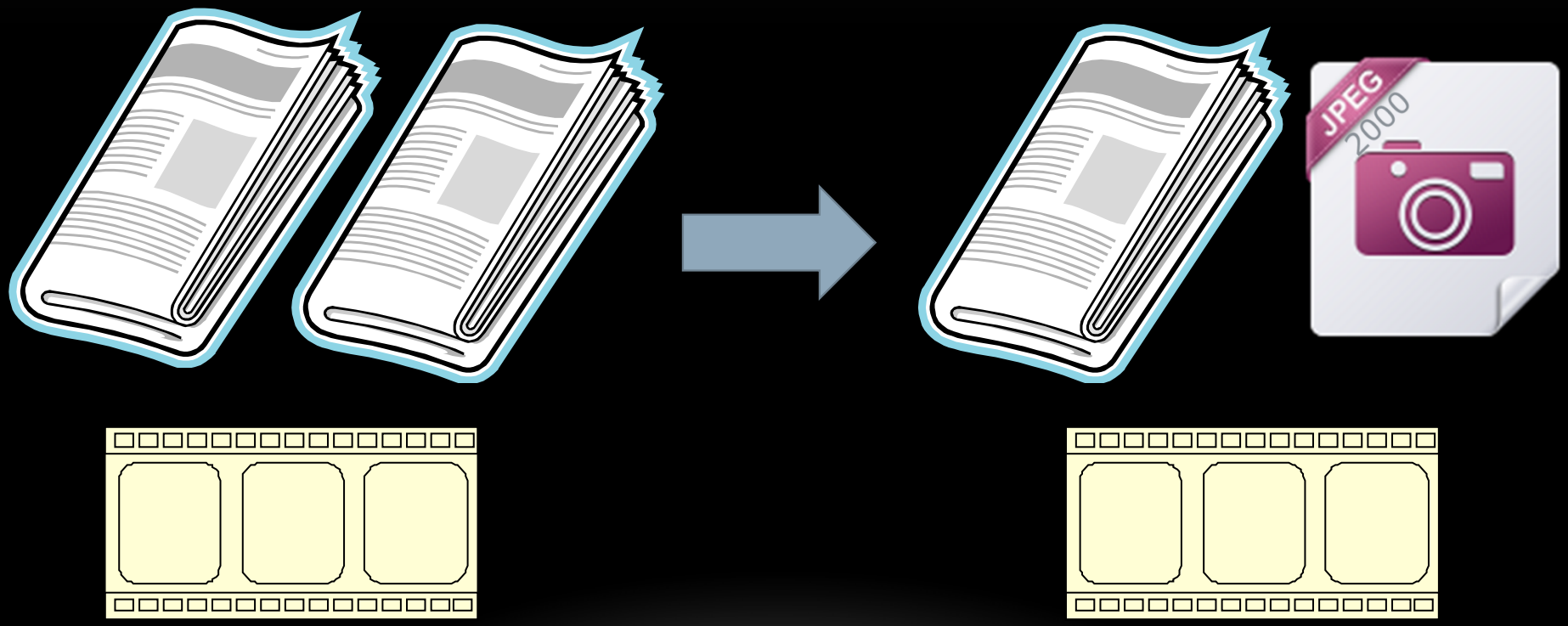
Niels Bønding
nieb@statsbiblioteket.dk

FINANSIERING

- Bevaringsformål – Finanslov 2012 – 16 mio. kr.
  - Digitalisering med henblik på kassation
  - Genererer billedfiler
- Tilgængeliggørelse – UMTS midler – 7 mio. kr.
  - Brugergrænseflade, søgning
  - OCR og segmentering
- Rigtig mange penge og arbejdstimer fra Statsbibliotekets løbende bevilling!

Niels Bønding
nieb@statsbiblioteket.dk

STRATEGI FOR BEVARING



Niels Bønding
nieb@statsbiblioteket.dk

TIDSPLAN FOR PROJEKTET

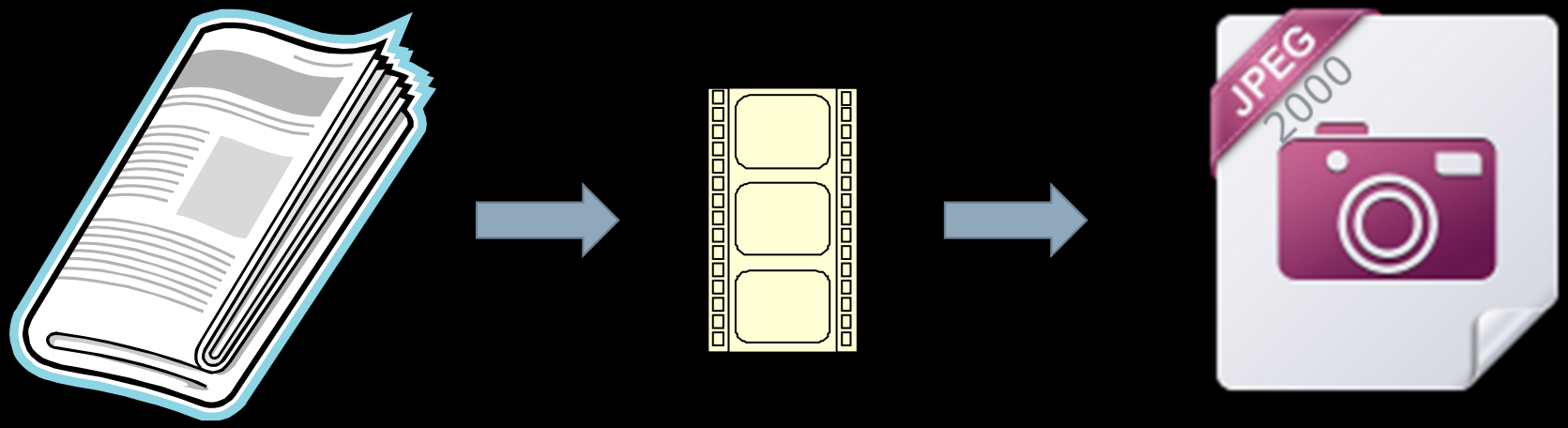
- 1. halvår 2014:
  - Pilotfase, produktionsmodning
- 2. halvår 2014:
  - Fuld produktion
  - 1.000.000 sider pr. måned
  - 50.000 sider pr. arbejdsdag
- 1. halvår 2017: Færdig med denne runde …

Niels Bønding
nieb@statsbiblioteket.dk

DIGITALISERING – HVORDAN?

Niels Bønding
nieb@statsbiblioteket.dk

DIGITALISERING FRA MIKROFILM SOM VI ALLEREDE HAR



Niels Bønding
nieb@statsbiblioteket.dk

MIKROFILMSPOLER

Niels Bønding
nieb@statsbiblioteket.dk

I KASSER

Niels Bønding
nieb@statsbiblioteket.dk

I REOLER

Niels Bønding
nieb@statsbiblioteket.dk

TAL

- 1 kasse indeholder op til 15 æsker
- 1 æske indeholder 1 spole med 1 film
- 1 film indeholder i gennemsnit 500 billeder svarende til 1.000 sider
- Vi skal håndtere 32.000 spoler, godt 2.100 kasser, og alle de metadata der hører til
- Det bliver til 32.000.000 avissider med tilhørende metadata
- Op mod 140.000.000 filer
- 800 TB kulturarv

Niels Bønding
nieb@statsbiblioteket.dk

HVILKE TITLER SKAL DIGITALISERES?

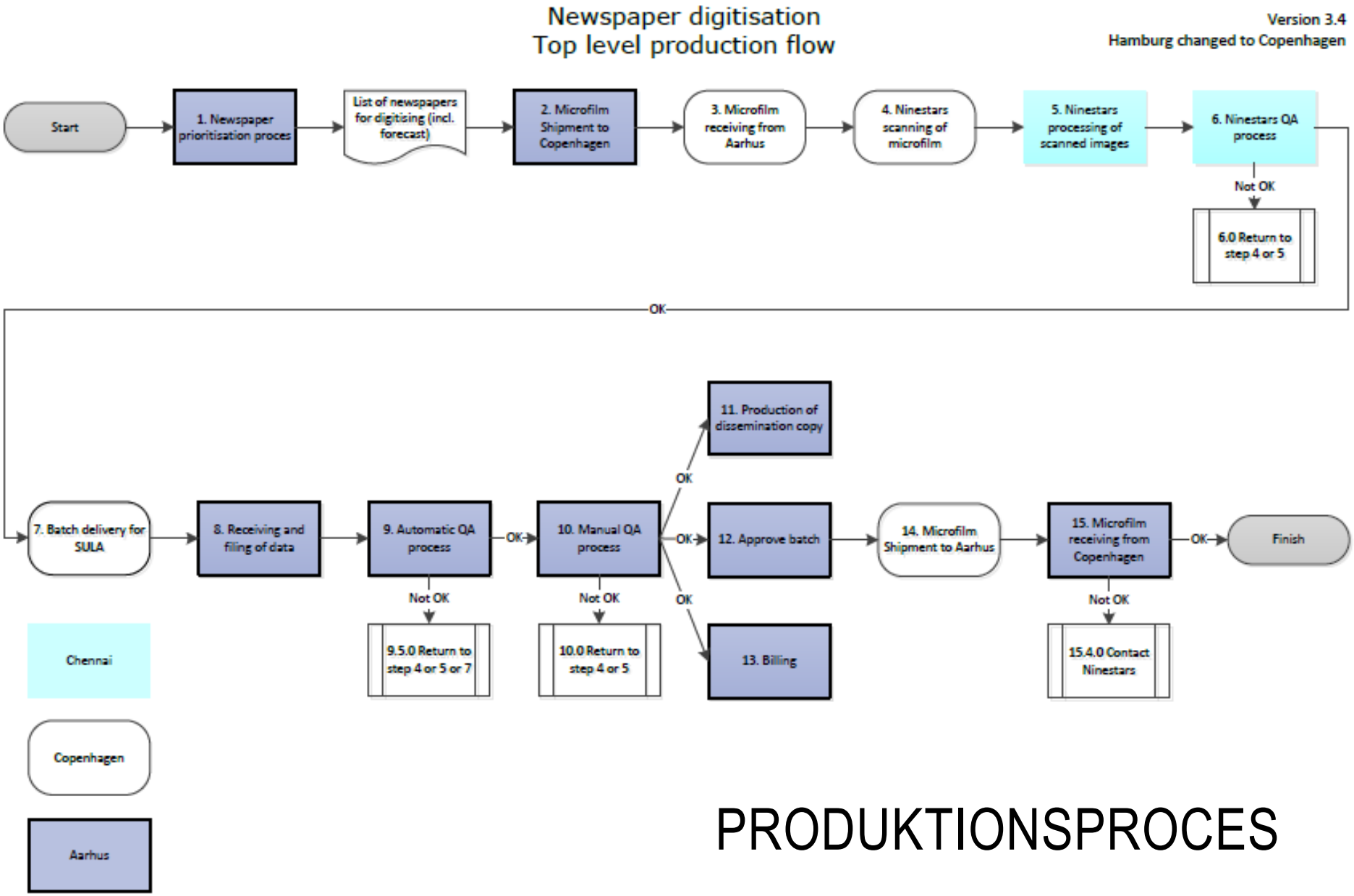
- Et stort puslespil, for at få mest mulig værdi for pengene
- De fleste nulevende dagblade
- Adresseavisen
- Aktuelt
- Land og folk
- Se listen på: http://blog.avisdigitalisering.dk/avistitler/

Niels Bønding
nieb@statsbiblioteket.dk

Ninestars

- Scanning i København hos Scanning.dk A/S
- Manuel efterbehandling i Chennai
- Hovedkvarter og IT-udvikling i Bangalore

Ninestars bygning i Chennai



PRODUKTIONSPROCES

Niels Bønding
nieb@statsbiblioteket.dk

1. PRIORITERINGSPROCES

- Hvad bestemmer rækkefølgen?
  - Kontrakter med de nulevende dagblade
  - Relevanskriterium

Niels Bønding
nieb@statsbiblioteket.dk

2/3. FILMENE SENDES TIL KØBENHAVN

Niels Bønding
nieb@statsbiblioteket.dk

4. SCANNING AF MIKROFILM

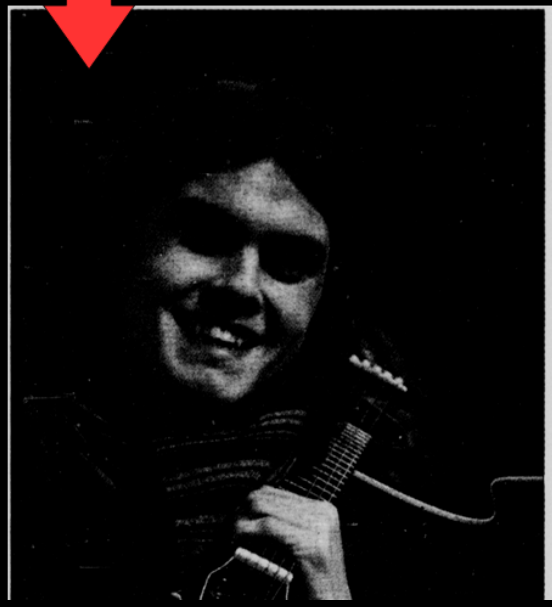
- 2 Mekel Mach 5 scannere placeret i Albertslund
- Vi har været med ved kalibreringen af scannerne. Vi og Ninestars har ved fælles hjælp opnået et rigtig godt resultat.
- Det rå output er TIFF scannet i 300 dpi. Den endelige fil er en JPEG2000 og kæden mellem dem er lossless
- Vi får samtidig målt emulsion density på mikrofilmene til brug for kassationsprocessen

Niels Bønding
nieb@statsbiblioteket.dk

BILLEDKVALITET – BEVARINGSKVALITET?


Følsom sanger fik masser af ros

Service guiden



udtrykke sig.
- På den måde kan jeg selv spille til, og er ikke afhængig af at skulle ud og finde en guitarist eller et band for at synge mine sange.

Følsomme sange

Som man kunne se ved hans audition, så er det de mere følsomme toner, d… Mathias når guit… frem.
- Jeg er mest til stille og rolig musik. Det er det, der tiltaler mig mest, men jeg har ikke nogen forbilleder som sådan siger han og tilføjer, at sangeren og sangskriveren Tim Christensen da er meget cool. Mathias skriver endnu ikke sange selv, men det er noget som han gerne vil kunne engang i fremtiden.

udtrykke sig.
- På den måde kan jeg selv spille til, og er ikke afhængig af at skulle ud og finde en guitarist eller et band for at synge mine sange.

Følsomme sange

Som man kunne se ved hans audition, så er det de mere følsomme toner, der tiltaler Mathias når guitaren findes frem.
- Jeg er mest til stille og rolig musik. Det er det, der tiltaler mig mest, men jeg har ikke nogen forbilleder som sådan siger han og tilføjer, at sangeren og sangskriveren Tim Christensen da er meget cool. Mathias skriver endnu ikke sange selv, men det er noget som han gerne vil kunne engang i fremtiden.

Niels Bønding
nieb@statsbiblioteket.dk

BILLEDKVALITET – BEVARINGSKVALITET!


BØRS & FINANS

Sløje pensionsafkast

Ny dansk diabetes-medicin på vej

Offentlige vagthunde affives


en af
inden
ersyge,
rrence

Zea-
ledes
forsk-
s, som
middel
abetes
forsøg

dheds-
givet
grønt
e fase
ealand
egne-
er skal
gemid-

Arkivfoto: LENE ESTHAVE PEDERSEN

Samtidig regner v
at lægemidlet vil vise
have en mæthedsføle
re tid. Det vil sige, a
vil spise mindre og d
vægttab, hvilket er g
betes-patienter,« sige
ness.

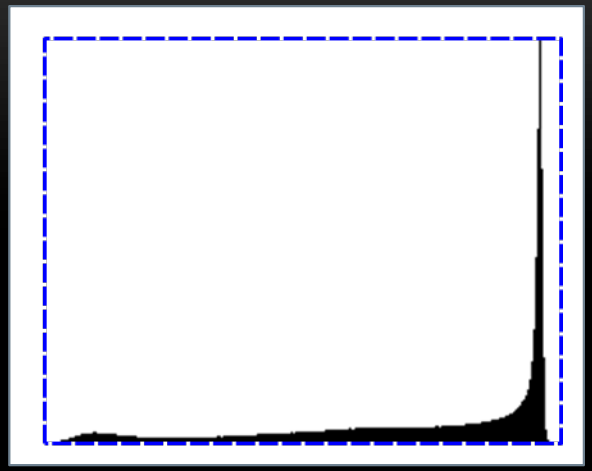
Flere af de antidia
midler, som er på
dag, har netop den
sukkersygepatienter
at tage på i vægt.

Zealand Phar
samarbejder om diab
tet med irske Elan C
specialist i at

Gigantisk marked

udvikle forskellige
gemidler kan indtage
videre bliver lægemi
ret i patienterne.

Markedet for type



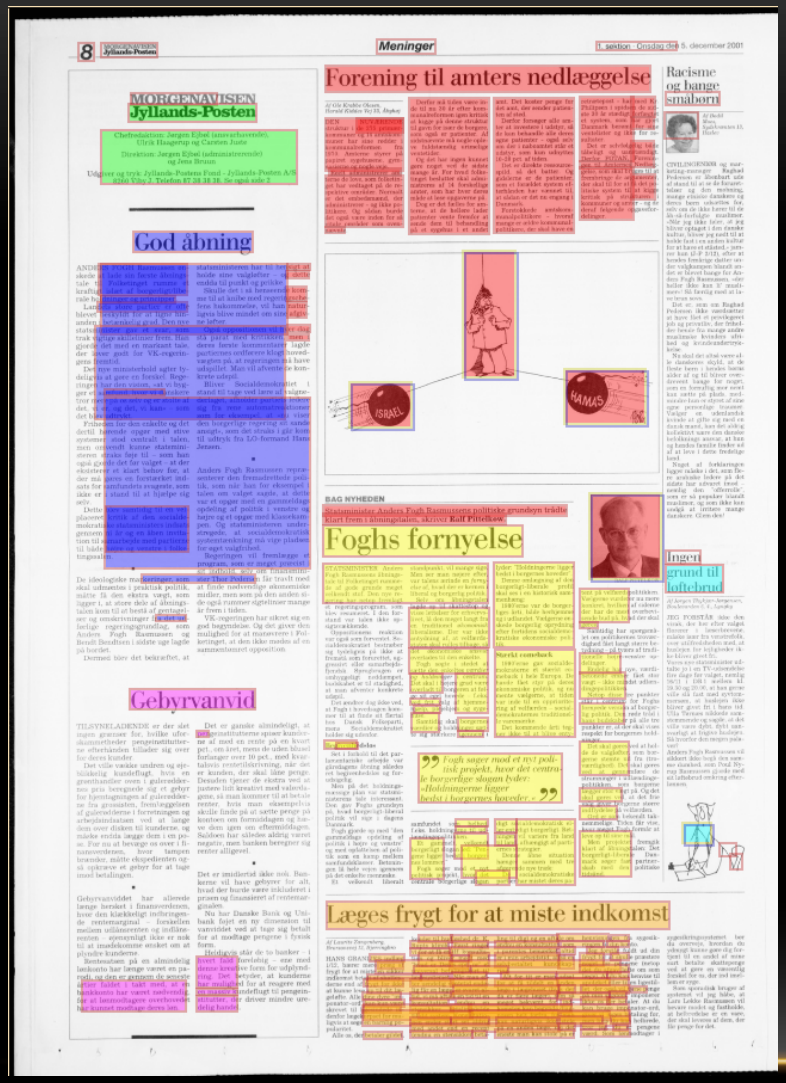
Niels Bønding
nieb@statsbiblioteket.dk

5. EFTERBEHANDLING I CHENNAI

- Split, crop og deskew
- Metadata
  - Tekniske metadata om den enkelte scannede side
    - Fx størrelse, opløsning, dato og operatør
  - Metadata om mikrofilmen
    - Titler, datoer, film og filmning
  - Udgave og udgivelse og den enkelte side
    - Morgen, aften, weekend, ekstra eller særudgave
    - Placering af siden i udgaven og på mikrofilmen
- OCR og segmentering

Niels Bønding
nieb@statsbiblioteket.dk

OCR OG SEGMENTERING - FORMIDLING



Forening til amters nedlæggelse

Racisme og bange småborn

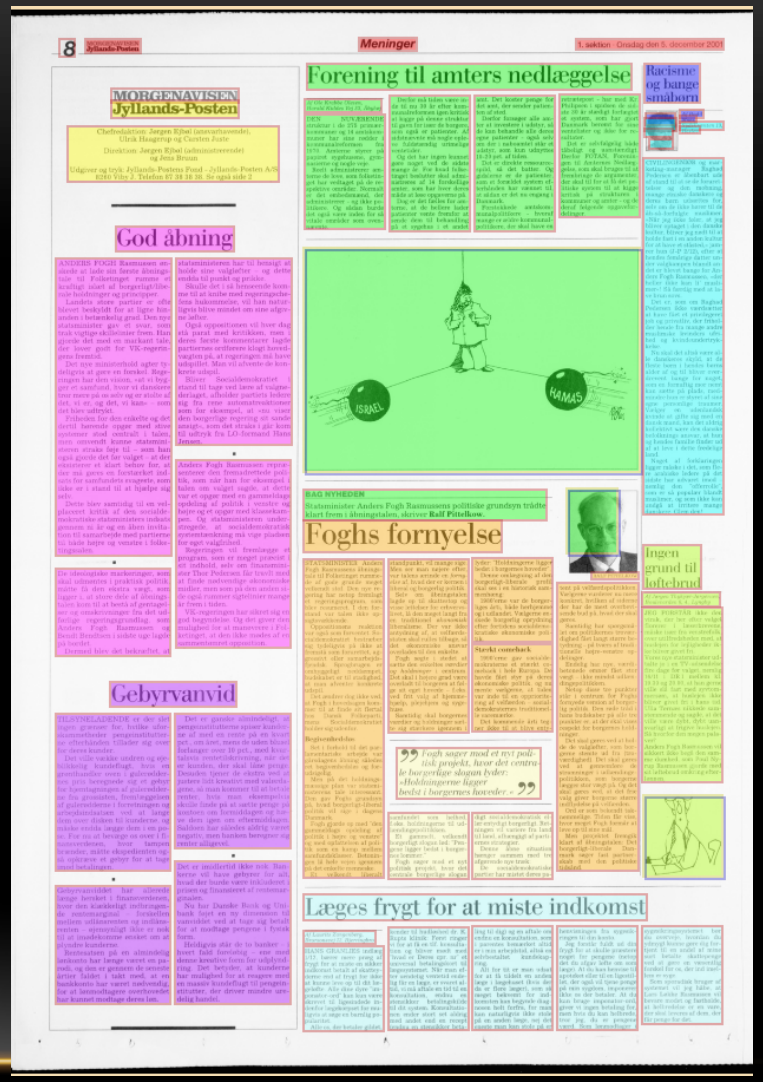
God åbning

Gebyrvanvid

Foghs fornyelse

Ingen grund til loftebrad

Læges frygt for at miste indkomst



Forening til amters nedlæggelse

Racisme og bange småborn

God åbning

Gebyrvanvid

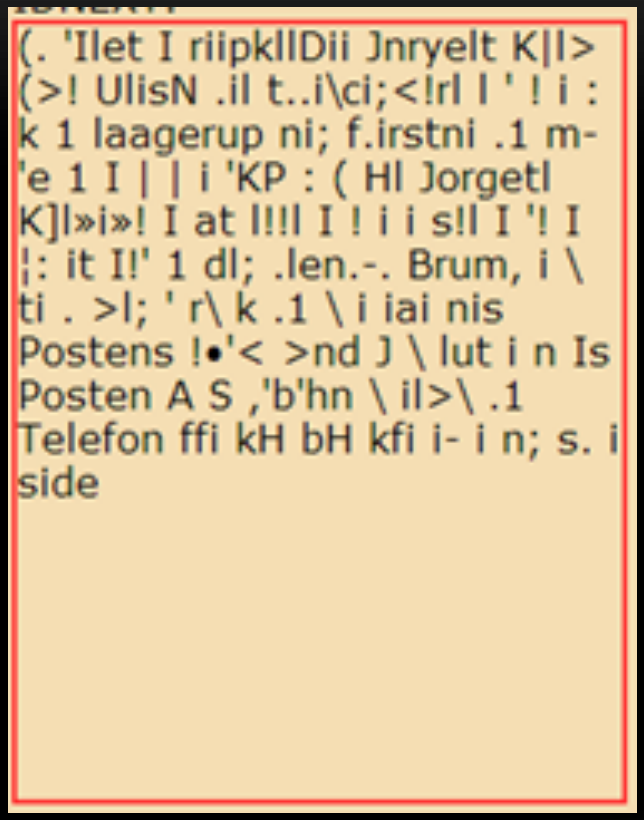
Foghs fornyelse

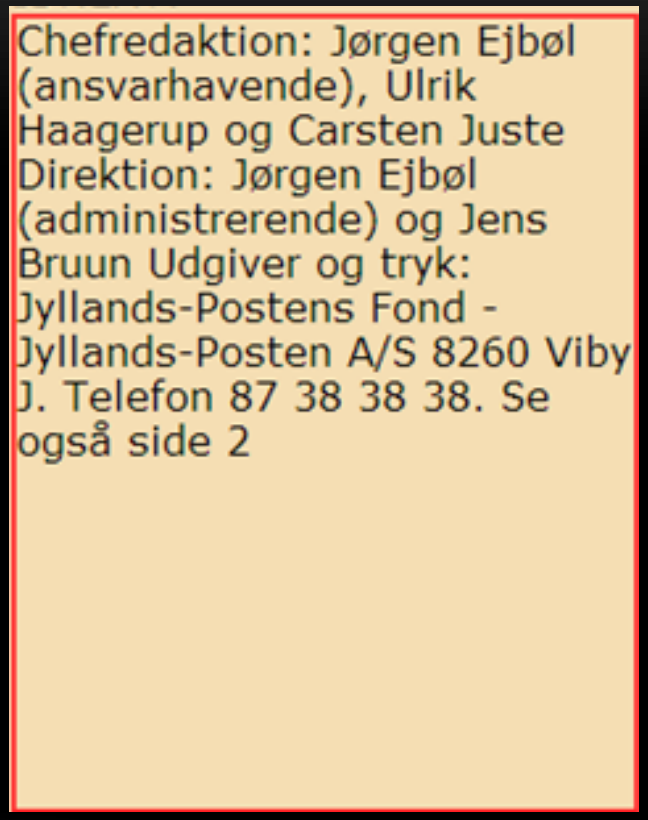
Ingen grund til loftebrad

Læges frygt for at miste indkomst

Niels Bønding
nieb@statsbiblioteket.dk

OCR OG SEGMENTERING - FORMIDLING





6/7/8/9/10 MODTAGELSE AF FILER OG KVALITETSKONTROL

- Ninestars sender de færdige batches til os, men inden da har de allerede kørt filerne igennem vores automatiske kvalitetskontrol.
- Vi modtager filerne, og kører dem igennem endnu en gang.
- Derefter foretager vi en manuel stikprøvekontrol og kigger på de sider, den automatiske kontrol fandt bemærkelsesværdige.

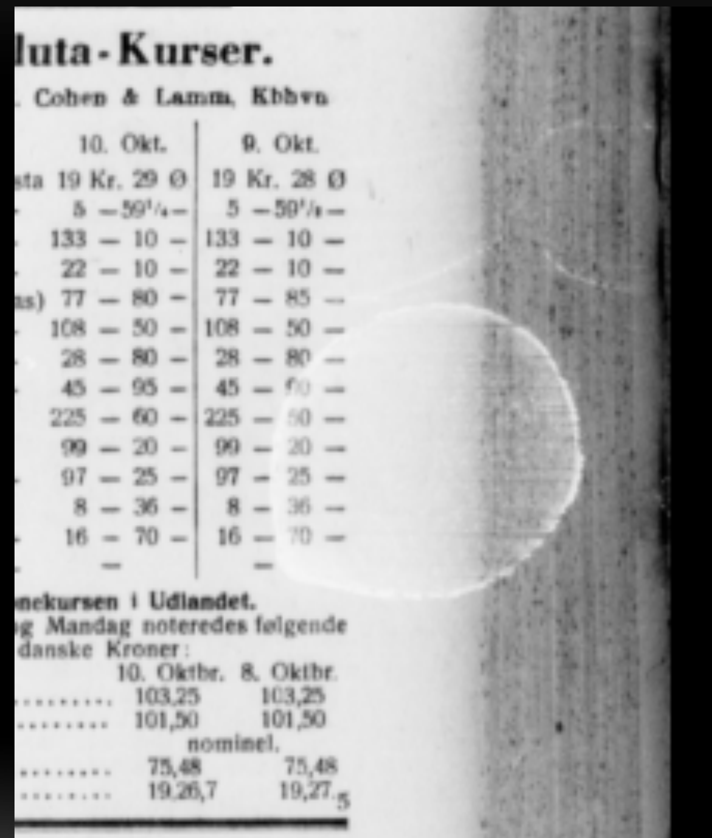
Niels Bønding
nieb@statsbiblioteket.dk

MANUEL KVALITETSKONTROL

- 50.000 sider hver arbejdsdag
- ISO standard DS/ISO 2859
- Billedkvalitet – notere evt. fejlkilde
- Kontrollere metadata
- OCR og segmentering

Niels Bønding
nieb@statsbiblioteket.dk

FOR KVALITETEN BLIVER JO ALDRIG BEDRE END ORIGINALMATERIALET



Niels Bønding
nieb@statsbiblioteket.dk

HVILKEN KVALITET TIL HVILKET FORMÅL?

- Vi skal **bevare** de digitaliserede aviser i den bedst mulige kvalitet for fremtiden
- Vi skal **tilgængeliggøre** de digitaliserede aviser i den bedst mulige kvalitet for nutiden
- Vi skal kunne **afgøre** hvorvidt bevaringskopi og mikrofilm er tilstrækkelig gode til, at vi kan vælge at **kassere** den papirkopi, der ligger i Vestindisk Pakhus. For det er jo der, pengene kommer fra.

Niels Bønding
nieb@statsbiblioteket.dk

NÅR DE DIGITALE FILER ER GODKENDT

- Fremstiller vi digitale formidlingskopier
- Modtager vi mikrofilmene retur fra København
- Begynder vi at arbejde med kassationsprocessen for en given titel
- Og vi tilføjer supplerende metadata

Niels Bønding
nieb@statsbiblioteket.dk

SUPPLERENDE METADATA

- Sammenkædning af titler over tid
- Geografiske data
- Kobling til De Danske Aviser
  - http://www.statsbiblioteket.dk/nationalbibliotek/adgang-til-samlingerne/aviser/de-danske-aviser

avisID	titel	fra	til	DDA-id	udgivelsessted
avisID	title	from	to	DDA-id	place of publication
aktuelt	Socialisten	1871-07-21	1874-05-09	1-137	København
aktuelt	Social-Demokraten	1874-05-10	1959-03-31	1-137	København
aktuelt	Aktuelt	1959-04-01	1978-01-16	1-137	København
aktuelt	Lands-avisen Aktuelt	1978-01-17	1982-12-31	1-137	København
aktuelt	Aktuelt	1983-01-01	1987-05-02	1-137	København
aktuelt	Det Fri Aktuelt	1987-05-02	1997-08-07	1-137	København
aktuelt	Aktuelt	1997-08-08	2001-04-06	1-137	København

Niels Bønding
nieb@statsbiblioteket.dk

TILGÆNGELIGGØRELSE

Niels Bønding
nieb@statsbiblioteket.dk

TILGÆNGELIGGØRELSE

- Copyright, lovgivning, økonomi sætter rammerne.
- På Statsbiblioteket, Det Kongelige Bibliotek, Det Danske Filminstitut: adgang til alt via Mediestream/aviser.
- Online på Mediestream.dk: fri adgang for alle til aviser udenfor copyright.
- Adgang via nogle dagblades hjemmesider på varierende vilkår.
- Anden adgang måske senere?

ADGANG TIL AVISERNE

På SB, KB, DFI

Adgang til ALT

Online for alle

Adgang til gamle aviser

Avisernes websites

Adgang til enkelte avisers arkiv

Niels Bønding
nieb@statsbiblioteket.dk

ADGANG TIL AVISERNE

Periode	Adgang på SB og KB	Snippets/ thumbnails	Fjernadgang alle	Academic License
1950+	Ja	Afhænger af copydan-aftale		
1940	Ja			
1930	Ja			
1920	Ja			
1910	Ja			
1900	Ja	Ja	Ja	Ja
1890	Ja	Ja	Ja	Ja
1880	Ja	Ja	Ja	Ja
1870	Ja	Ja	Ja	Ja
1860	Ja	Ja	Ja	Ja
ældre	Ja	Ja	Ja	Ja

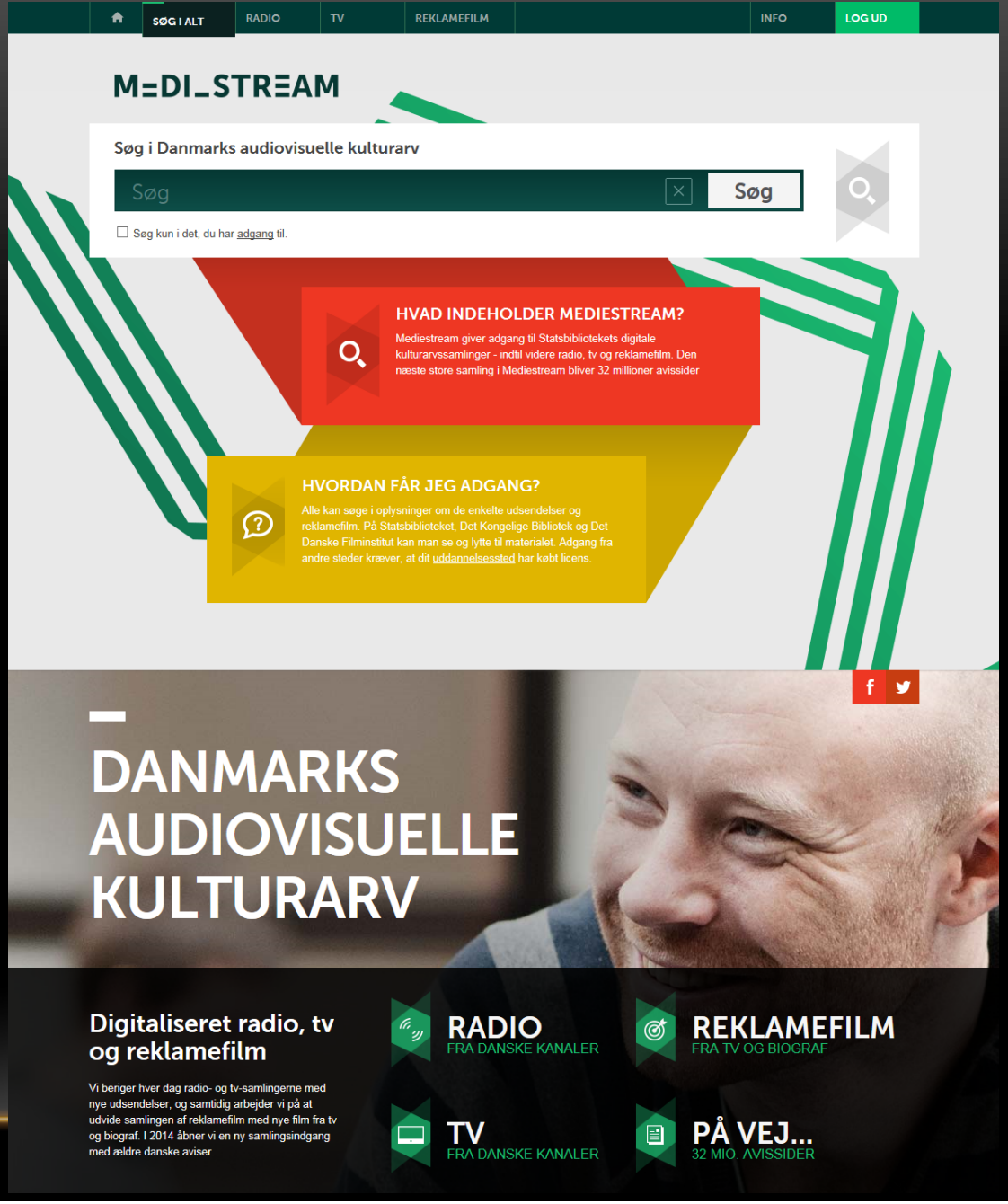
Niels Bønding
nieb@statsbiblioteket.dk

SØGEMULIGHEDER

- Titel, dato og sidetal
- Tekstgenkendelse / OCR
  - Frakturskrift – lavere genkendelsesgrad
- Segmentering – identifikation af ”artikler” – automatisk

Niels Bønding
nieb@statsbiblioteket.dk

mediestream/aviser åbner foråret 2015



Niels Bønding
nieb@statsbiblioteket.dk

STATSBIBLIOTEKET

De

til Forsendelse med Posten

allene privilegerede

Kiøbenhavnske

Maanedlige Stats-Tidender

Første Stykke

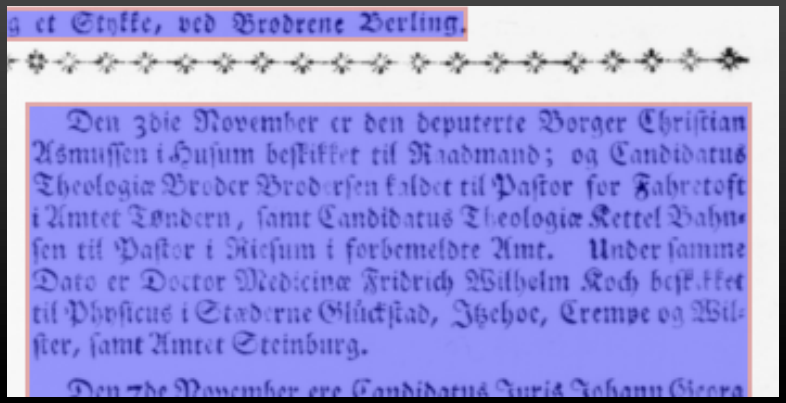
for

Januarii Maaned 1787.

Af disse Tidender udgives Maanedlig et Stykke, ved Brødrene Berling.

Den 22de Martii 1786 har Hs. Kongel. Majestæt allernaadigst behaget at medeele Amtmand Olav Stephensen i Island, for hans Godhedighed mod Trængende i de sammesteds sidst indtrufne haarde Aaringer, samt andre af ham tilforn ved adskillige Leiligheder udøvede patriotiske Handlinger, den store Medaille pro Meritis, i Guld, og tilliige be-

Den 3die November er den deputerte Borger Christian Asmussen i Husum beskikket til Raadmand; og Candidatus Theologiæ Broder Brodersen kaldet til Pastor for Fahretoft i Amtet Tøndern, samt Candidatus Theologiæ Kettel Bahnsen til Pastor i Riesum i forbemeldte Amt. Under samme Dato er Doctor Medicinæ Fridrich Wilhelm Koch beskikket



et Stykke, ved Brodrene Berling.

Den 3die November er den deputerte Borger Christian Asmussen i Husum beskikket til Raadmand; og Candidatus Theologiæ Broder Brodersen kaldet til Pastor for Fahretoft i Amtet Tøndern, samt Candidatus Theologiæ Kettel Bahnsen til Pastor i Riesum i forbemeldte Amt. Under samme Dato er Doctor Medicinæ Fridrich Wilhelm Koch beskikket til Physicus i Staderne Glückstad, Itzehoe, Crempe og Wilster, samt Amtet Steinburg.

Den 7de November ere Candidatus Juris Johann Cicero

Den zdie November er den dcputerte Borger Christian Asmnssen i Husum bestikket til Raadmand; og Candidatus Theologis Brodér Brod rs-m kaldet til Pastor for Fahretost i Amtet Tsndcrn , samt Candidatus Tl eologi« Kettel Bahn« sen til Pastor i Ricsum i forbemeidte Amt. Under samme Dato er Doctor Mediciner Fridrich Wilhelm Koch bcfk.kk« til Physicus i Staderne Gluckstad, Itzehoe, Crempe ogWil- fter, samt Amtet Stcinbnrg. Den 7de

Den 3die November er den deputerte Borger Christian Asmussen i Husum beskikket til Raadmand; og Candidatus Theologiæ Broder Brodersen kaldet til Pastor for Fahretoft i Amtet Tøndern, samt Candidatus Theologiæ Kettel Bahnsen til Pastor i Riesum i forbemeldte Amt. Under samme dato er Doctor Medicinæ Fridrich Wilhelm Koch beskikket til Physicus i Staderne Gluckstad, Itzehoe, Crempe og Wilster, samt Amtet Steinburg.

STATUS MEDIO NOVEMBER 2014

- Er digitaliseret:
  - Bornholms Tidende
  - Jyllands-Posten
  - Aarhus Stiftstidende

  } i alt ca. 2.500.000 sider
- Er i proces:
  - Horsens Folkeblad
  - Adresseavisen

FØLG MED OG LÆS MERE PÅ avisdigitalisering.dk

Niels Bønding
nieb@statsbiblioteket.dk

HVAD SÅ NÅR VI ER FÆRDIGE MED DE 32.000.000 SIDER?

- Vi vil gerne lave crowdsourcing til forbedring af OCR, segmentering og metadata.
- Og der er stadig 30 mio. sider på mikrofilm og 40 mio. på papir at tage fat på.

Niels Bønding
nieb@statsbiblioteket.dk

SPØRGSMÅL?

Niels Bønding
nieb@statsbiblioteket.dk