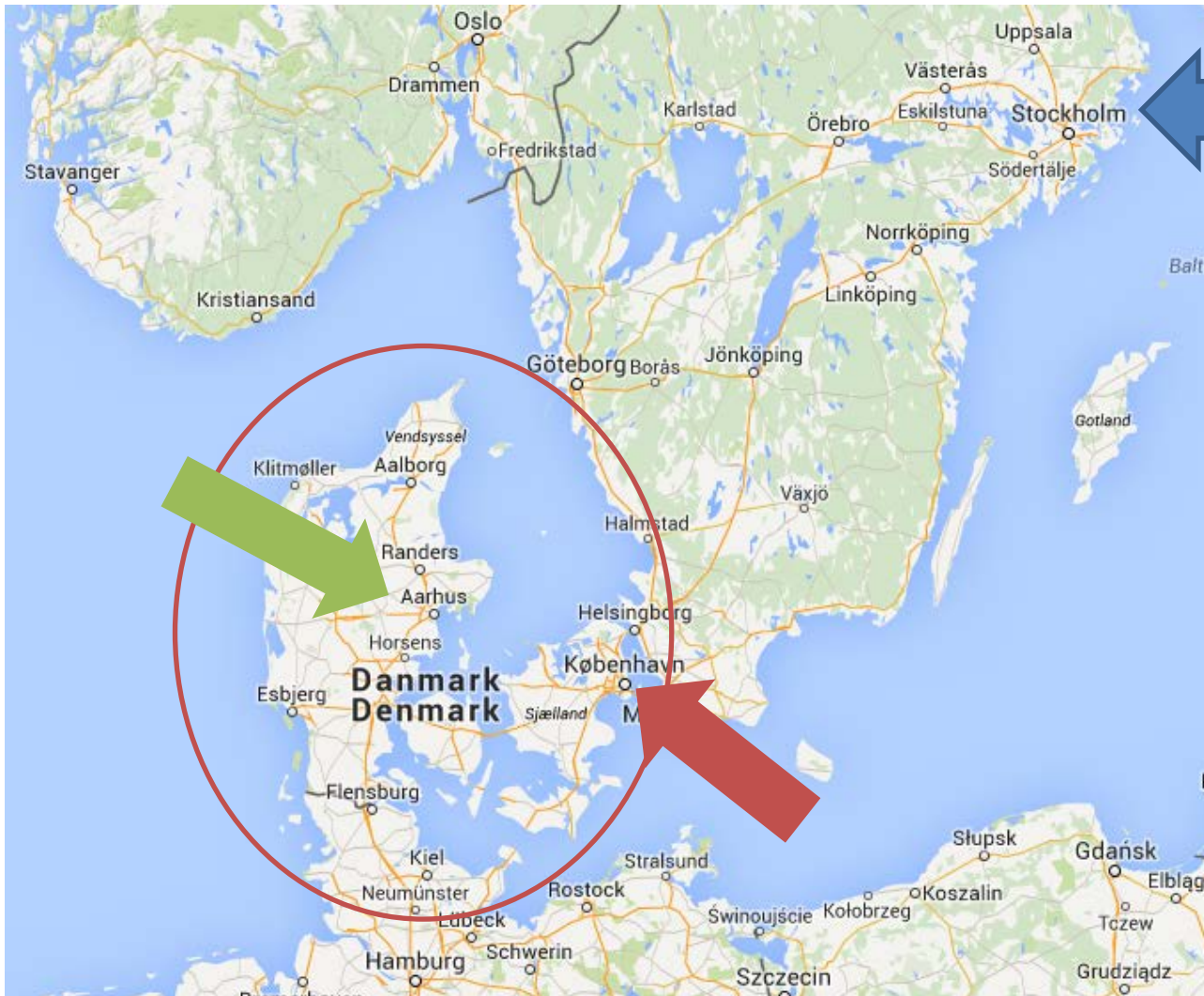# Chasing the news

report from 10 years of digital legal deposit in Denmark

Tonny Skovgård Jensen
Director, National Library Division
State and University Library
Denmark

Tonny Skovgård Jensen
tsj@statsbiblioteket.dk

STATS**BIBLIOTEKET**
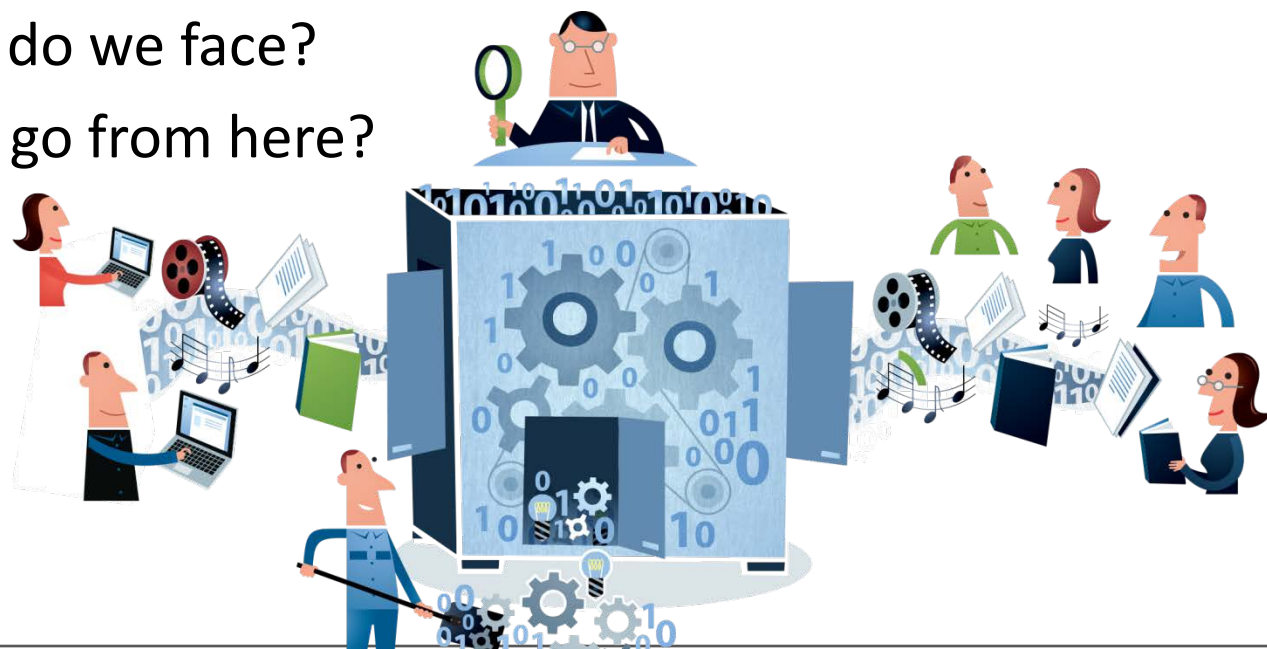
Where is it?

The State and University Library

- 3 main functions:
  - University libary
  - National servicies for public libraries
  - National library
- National Library #2
  - The Royal Library in Copenhagen is the oldest and biggest
- We specialize in media:
  - Newspapers
  - Radio
  - Television
  - The Danish part of the web (In cooperation with the Royal Library)
  - Recorded music
  - Historical audiorecordings
  - DVD / VHS

10 years of digital legal deposit

- Legal deposit act – 10 years anniversary
- Looking up - looking back – looking forward
- What did we learn?
- What did we collect?
- Which challenges do we face?
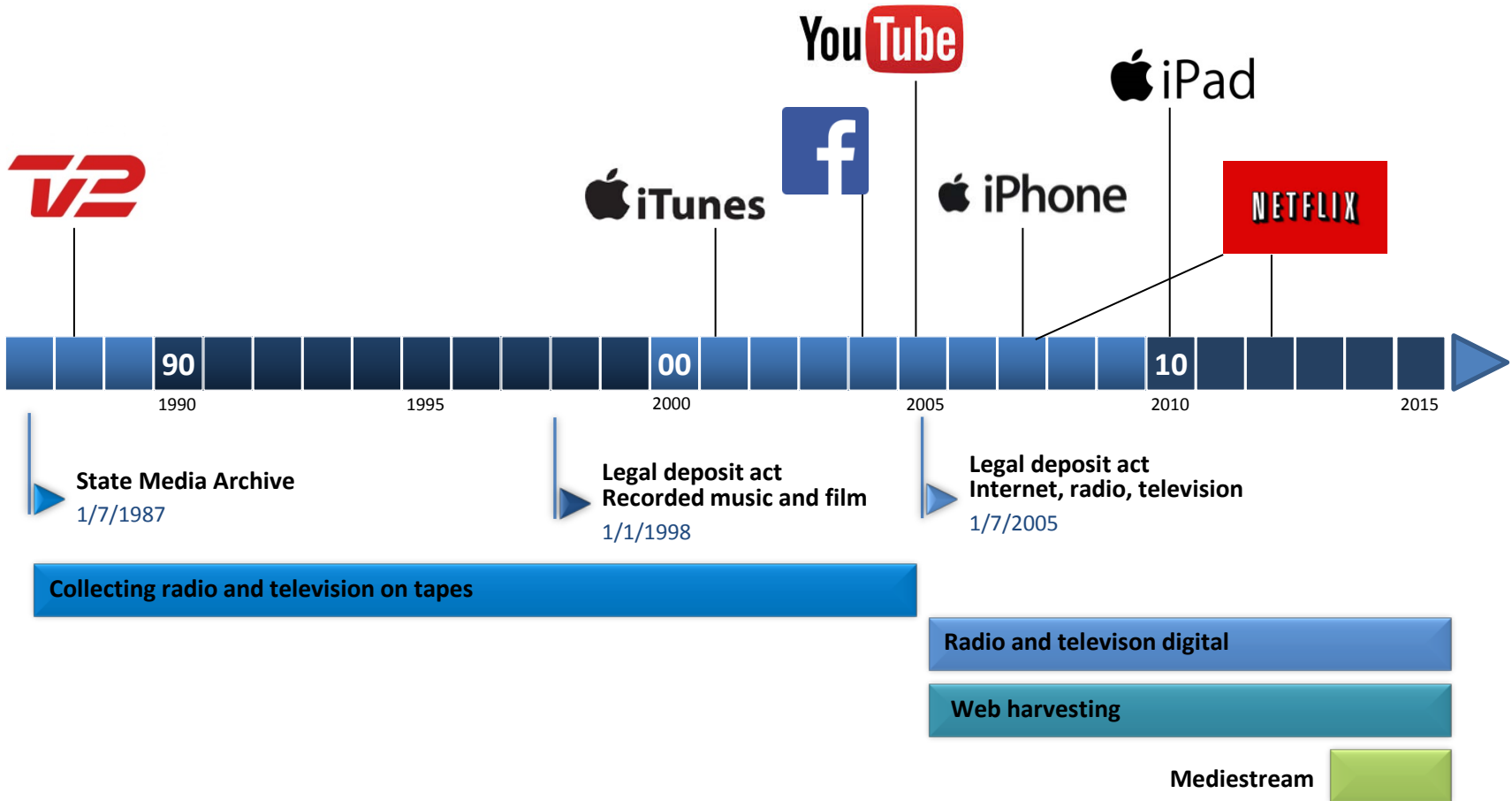- Where should we go from here?

Tonny Skovgård Jensen
tsj@statsbiblioteket.dk

STATS**BIBLIOTEKET**

Do you remember

2005

Tonny Skovgård Jensen
tsj@statsbiblioteket.dk

The world 2005

- Facebook was just about a year old, and had about 5 million users. Worldwide.
- YouTube was founded this year. Few people could imagine at that time the enormous amount of video that was soon to be distributed online.
- Twitter was founded the year after, in 2006
- You would have to wait two more years to see the first iPhone, one more year to hear of the first android smartphones, and 5 years to see the first Ipad.
- State of the art devices was still PDA's and also the first smartphones, like the Sony Ericsson P910, but slow connections and slow hardware made them quite limited for practical purposes.

Tonny Skovgård Jensen
tsj@statsbiblioteket.dk

STATS**BIBLIOTEKET**

History

State Media Archive
1/7/1987

Legal deposit act
Recorded music and film
1/1/1998

Legal deposit act
Internet, radio, television
1/7/2005

1990   1995   2000   2005   2010   2015

90 **00** **10**

Collecting radio and television on tapes

Radio and televison digital
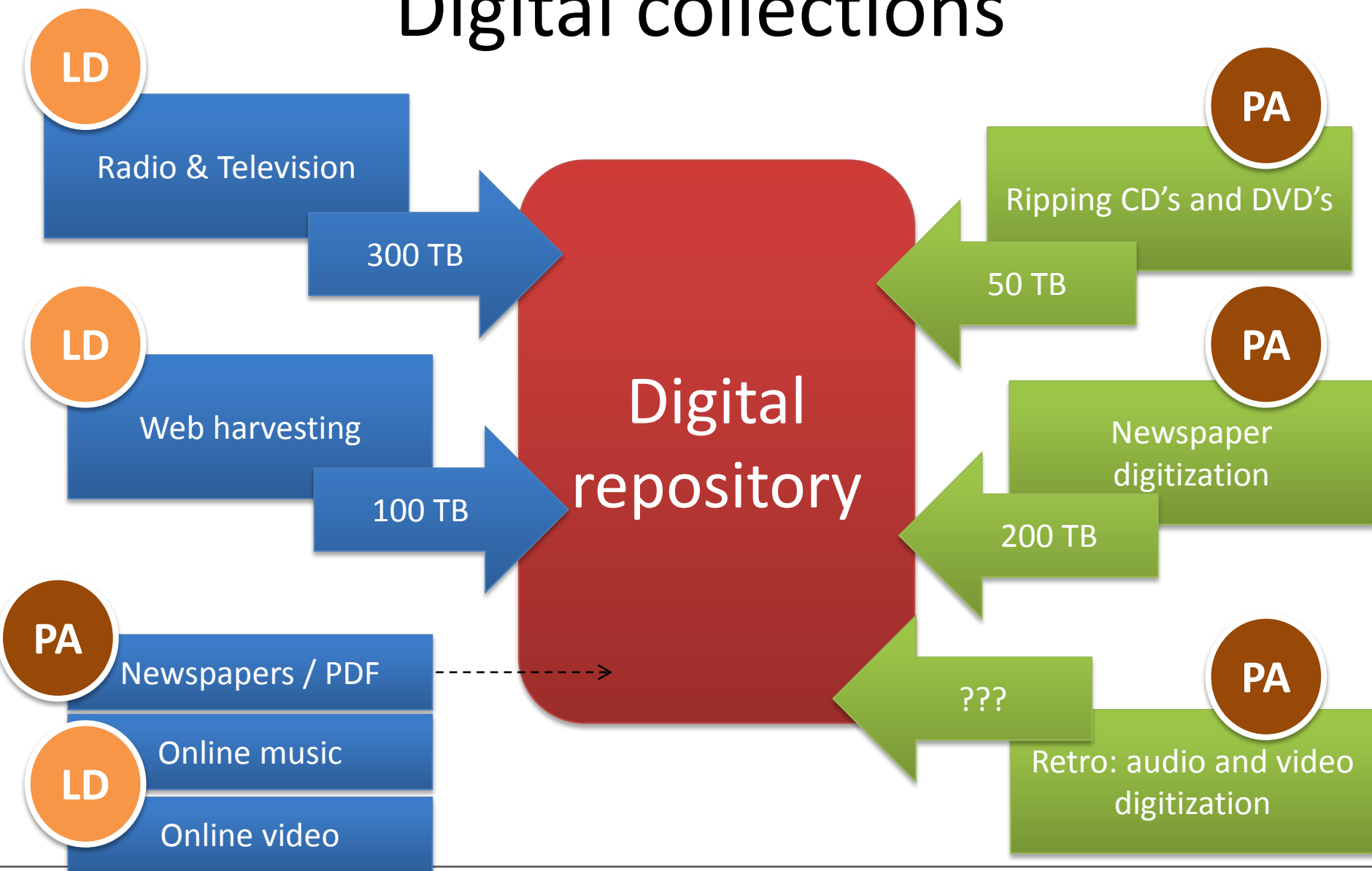
Web harvesting

Mediestream
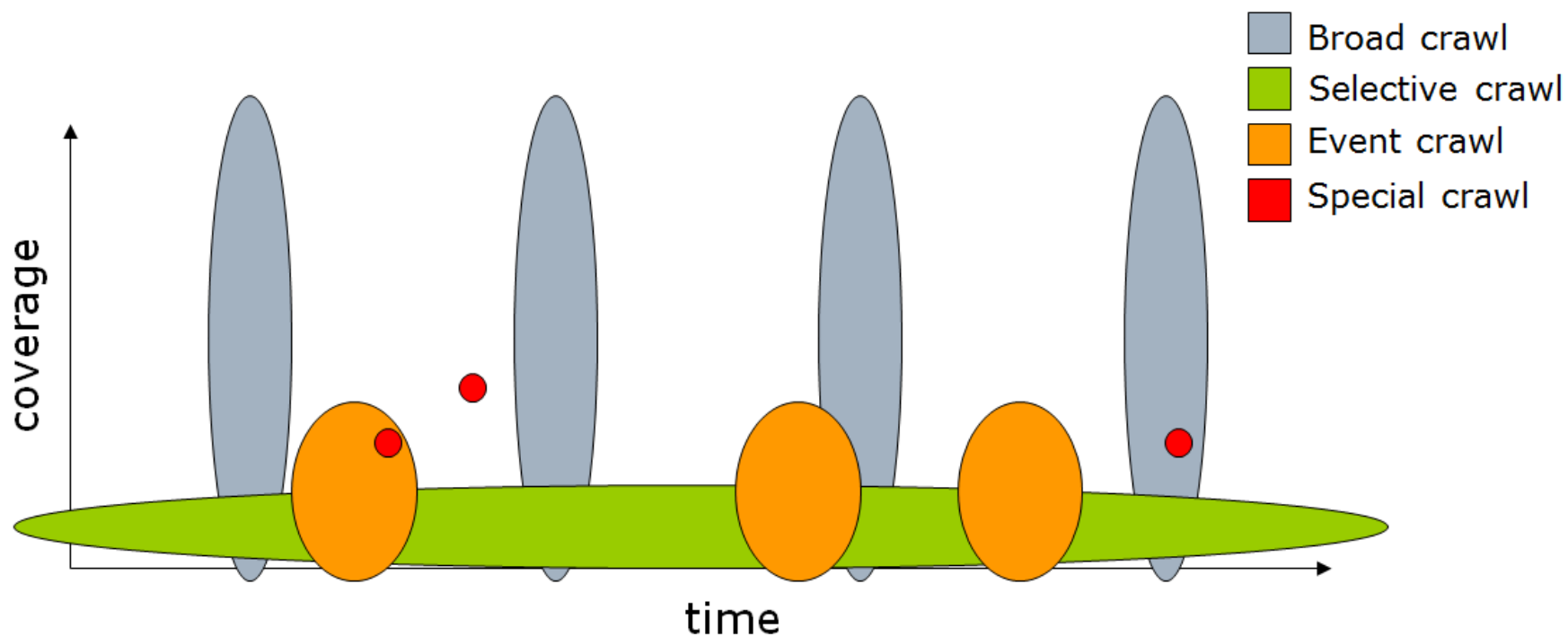
Digital collections

Digital legal deposit

- Publishers and producers do not deliver or upload the content. We have to collect or harvest it.

- News content is generally not handled as a seperate collection – it is mixed with other types of content.

Tonny Skovgård Jensen
tsj@statsbiblioteket.dk

Digital collections

LD — Radio & Television → 300 TB → Digital repository

LD — Web harvesting → 100 TB → Digital repository

PA — Newspapers / PDF ---→

LD — Online music

Online video

PA — Ripping CD's and DVD's ← 50 TB → Digital repository

PA — Newspaper digitization ← 200 TB → Digital repository

PA — Retro: audio and video digitization ← ??? → Digital repository

Tonny Skovgård Jensen
tsj@statsbiblioteket.dk

STATSBIBLIOTEKET

Harvesting the web

Legend:
- Broad crawl (grey)
- Selective crawl (green)
- Event crawl (orange)
- Special crawl (red)

Axes: coverage (vertical), time (horizontal)
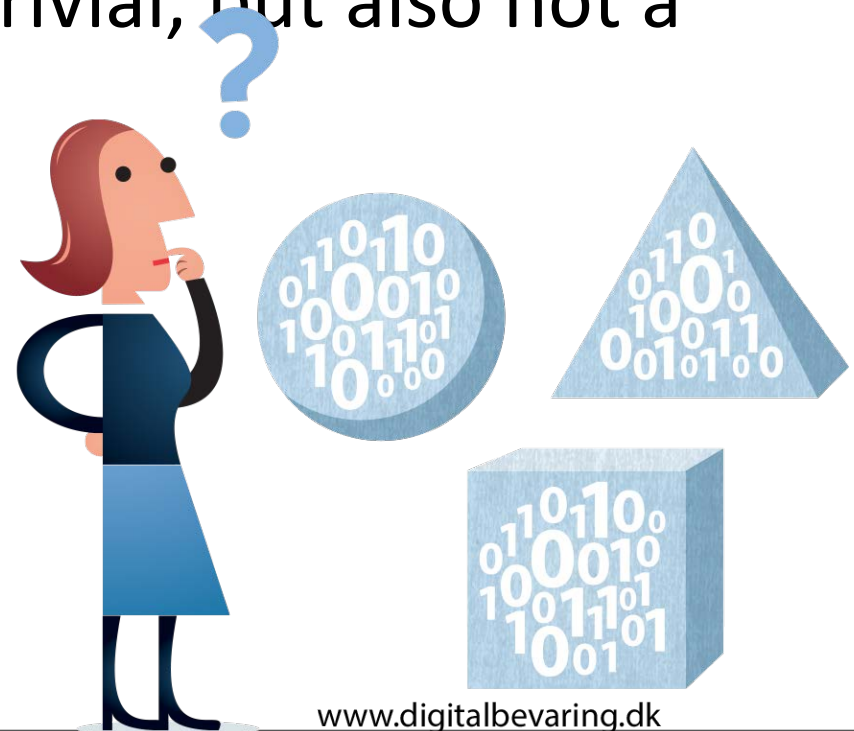
Tonny Skovgård Jensen
tsj@statsbiblioteket.dk

Harvesting "everything" … but …

- Very complex process
- Huge quantity – harvest / select / curate
- Where is the Danish content?
- A webpage is complex and dynamic
- Technology is developing fast – the harvester need to know how to capture the content
- Streaming
- Databases
- Paywalls
- Complex site structures
- Social media, mobile apps ….

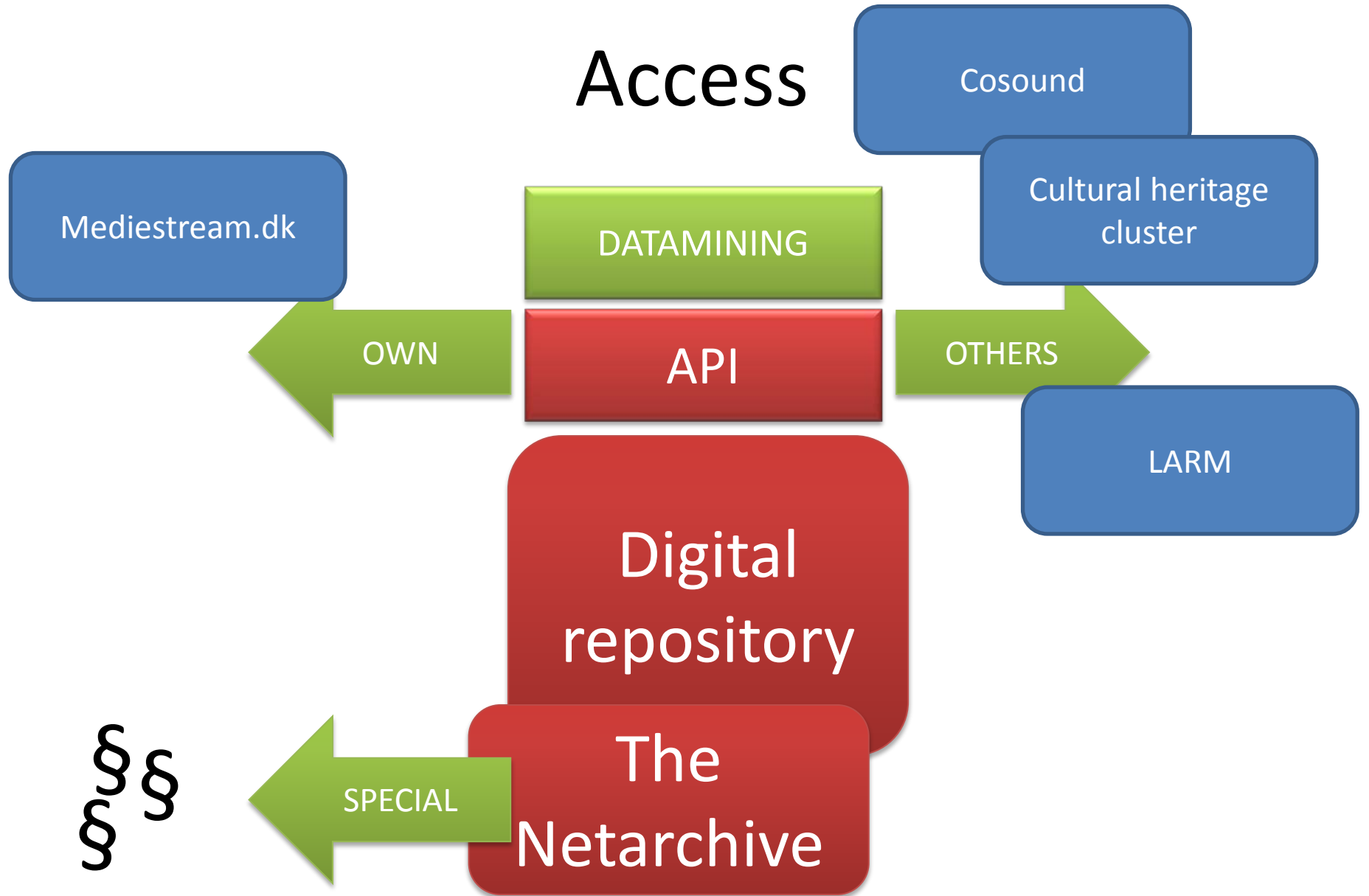Tonny Skovgård Jensen
tsj@statsbiblioteket.dk

Preservation

- ISO 16363 (Audit and certification of trustworthy digital repositories)

- Bit preservation – not trivial, but also not a complete jungle.

- Ready for conversion

www.digitalbevaring.dk

Tonny Skovgård Jensen
tsj@statsbiblioteket.dk

Access

Cosound

Cultural heritage cluster

Mediestream.dk

DATAMINING

OWN

API

OTHERS
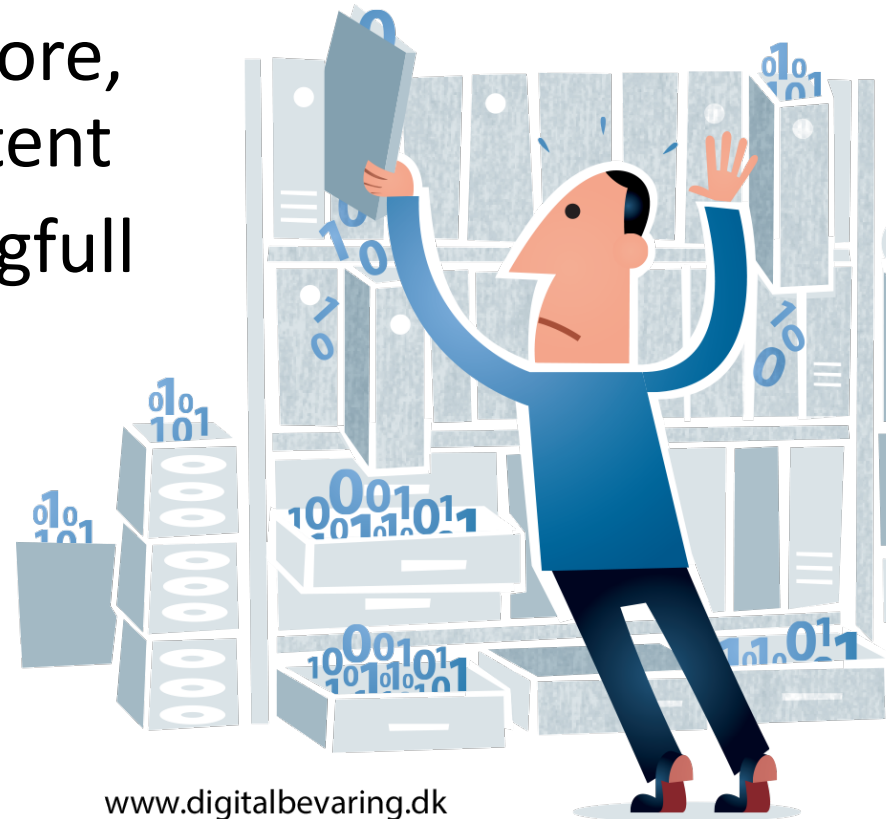
LARM

Digital repository

§§§

SPECIAL

The Netarchive

Access to the webarchive

- Everything in one arhive – personal data issues. New EU regulation is on the way.

- Very restrictive - access is only for academic researchers and students writing their thesis

    - Maybe not a problem in 2005, but it is now where a lot of important content is online only

- Combined legal and technical issue

- Access by URL (search is coming soon)

Summary – a new paradigm?

- Looking back – overwhelming amounts of digital content
- Looking forward – even more, and more online only content
- Can we select in a meaningfull way?
- Can we get it all?
- Can we still be in control?
- A case ...

www.digitalbevaring.dk

Tonny Skovgård Jensen
tsj@statsbiblioteket.dk

STATS**BIBLIOTEKET**

210810-41-20140714170211-00004-kb-prod-har-003.kb.dk.warc/175412702

Viewing version 1 of 6

17:41:19 jul 14, 2014

jun 29

jul 29

Help

MANDAG 14. JULI 2014

Jyllands-Posten Sport

VM-miniput overgår Tyskland og andre …

Fodbold - i samarbejde med
●TIPSBLADET.DK
SPIL-TIPS | LIVE | BLOG

| Forside | Premium | Indland | Politik | International | Erhverv | Sport | Kultur | Debat | Aarhus | Livsstil | Temaer ▾ | Mere ▾ |

| Fodbold ▾ | VM ▾ | Håndbold | Cykling | Tennis | Motorsport | Formel 1 ▾ | Golf | Boksning | Ishockey | Resultater | LIVE |

VM I FODBOLD 2014
GRUPPE: A | B | C | D | E | F | G | H | GRUPPESPIL | SLUTSPIL | #VMDK

MEST LÆSTE VM 2014

1. Danskere til tysk guld-fest: Det var vildt
2. VM-miniput overgår Tyskland og andre kæmper
3. Fodboldverden tilhører Tyskland
4. TV: Uroligheder i Argentina efter VM-skuffelse
5. Van Gaal frier til Robben: Kom bare med mig til United

VM SLUTSPIL 13.07.2014 kl. 23:38

Foto: Felipe Dana/AP

● ○ ○

1/8 FINALER ▾

- Kamp 49
- Kamp 50
- Kamp 51
- Kamp 52
- Kamp 53
- Kamp 54
- Kamp 55
- Kamp 56

Fodboldverden tilhører Tyskland

AF DAN PHILIPSEN

Nogen smuk VM-finale blev det aldrig, men et mesterværk til sidst gav tysk VM-guld.

Der er 15 lande på denne jord, der har flere mennesker at vælge fodboldspillere fra end Tyskland.

Men ingen har nogen så gode, som vores naboer mod syd.

A STØRRE | MINDRE

KAMPFAKTA

Tyskland-Argentina 1-0 (0-0)

Efter forlænget spilletid.

- Mål: 1-0 Mario Götze (113)

Fodboldverden vil i fire år frem tilhøre Tyskland. Den største ære i verdens bedste spil vundet på dramatisk ved at besejre Argentina med 1-0 i fodboldens vigtigste kamp.

FÅ DET FULDE OVERBLIK
Hele sagen

Questions?

TSJ@statsbiblioteket.dk

STATS**BIBLIOTEKET**

Credits

- Drawings: Jørgen Stamp, from www.digitalbevaring.dk
- Photos from the library: Thomas Søndergaard

Tonny Skovgård Jensen
tsj@statsbiblioteket.dk

STATS**BIBLIOTEKET**