Data enrichment for the future

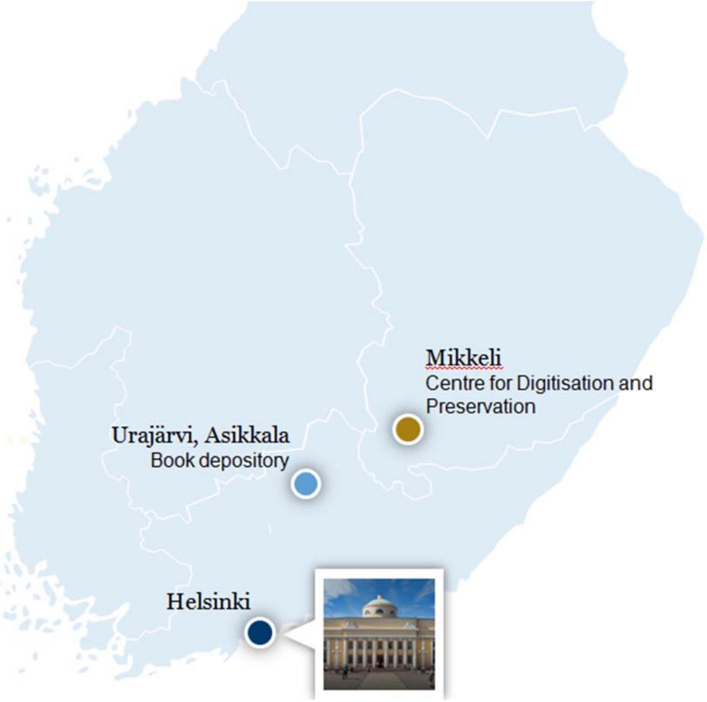
Presenter	The National Library of Finland / Centre for Preservation and Digitisation
	Jukka Kervinen (jukka.kervinen@helsinki.fi)
Date	10.5.2012
Event	Conference "Semantic technologies in libraries: from text to structure - from words to meaning"

Overview

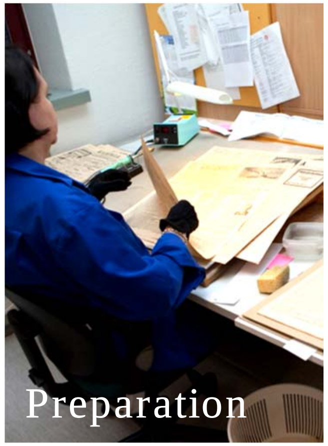
- Background
- Digitization
- Before semantics

National Library of Finland
Centre for Digitisation and Preservation

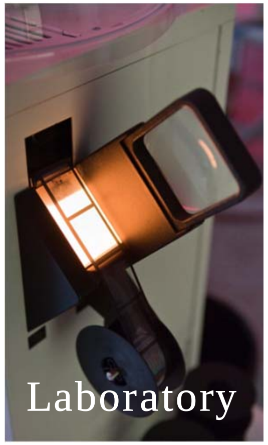
- Established in 1990
- Digitisation started in 1998
- Over 50 employees
- Emphasis on newspapers
- Yearly average (past three years):
  - Microfilming: 1,3 million exposures
  - Digitisation: 1,3 million pages
  - Audio digitisation 1 300 cassettes
  - Conservation: 12 000 units



Preservation


Preparation

Microfilming

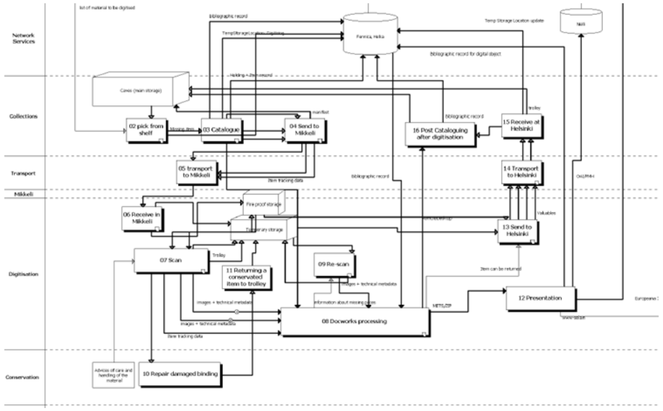

Laboratory

QA

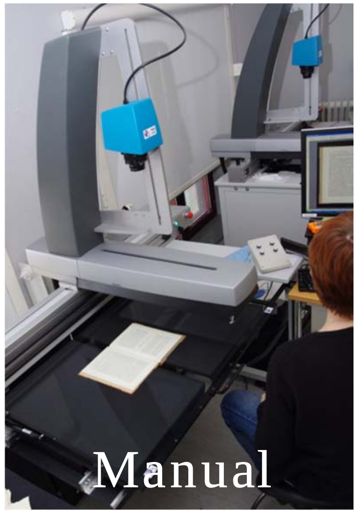
Computer Output Microfilm

Conservation

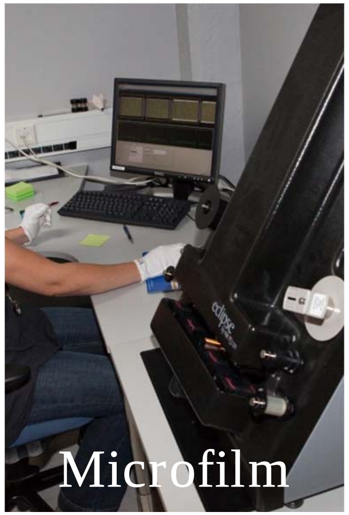
Planning and development



Digitisation


Manual

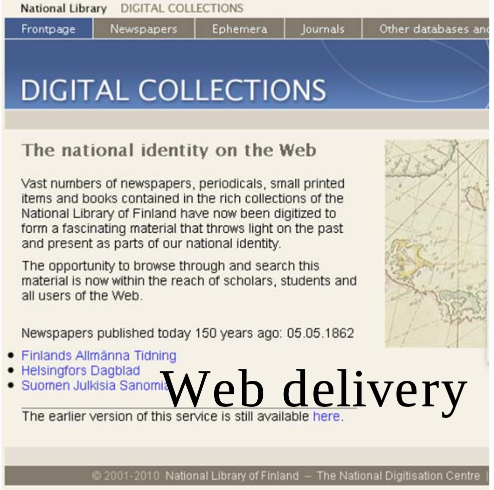
Automatic


Microfilm

Photo

Audio

Postprocessing


Web delivery

National Library DIGITAL COLLECTIONS

Frontpage | Newspapers | Ephemera | Journals | Other databases an

DIGITAL COLLECTIONS

The national identity on the Web

Vast numbers of newspapers, periodicals, small printed items and books contained in the rich collections of the National Library of Finland have now been digitized to form a fascinating material that throws light on the past and present as parts of our national identity.

The opportunity to browse through and search this material is now within the reach of scholars, students and all users of the Web.

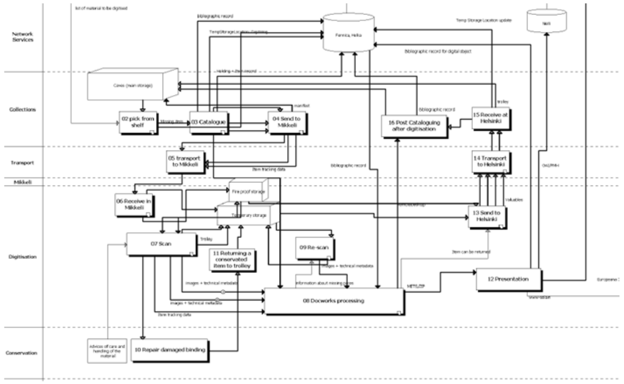
Newspapers published today 150 years ago: 05.05.1862

- Finlands Allmänna Tidning
- Helsingfors Dagblad
- Suomen Julkisia Sanomia

The earlier version of this service is still available here.

© 2001-2010 National Library of Finland – The National Digitisation Centre

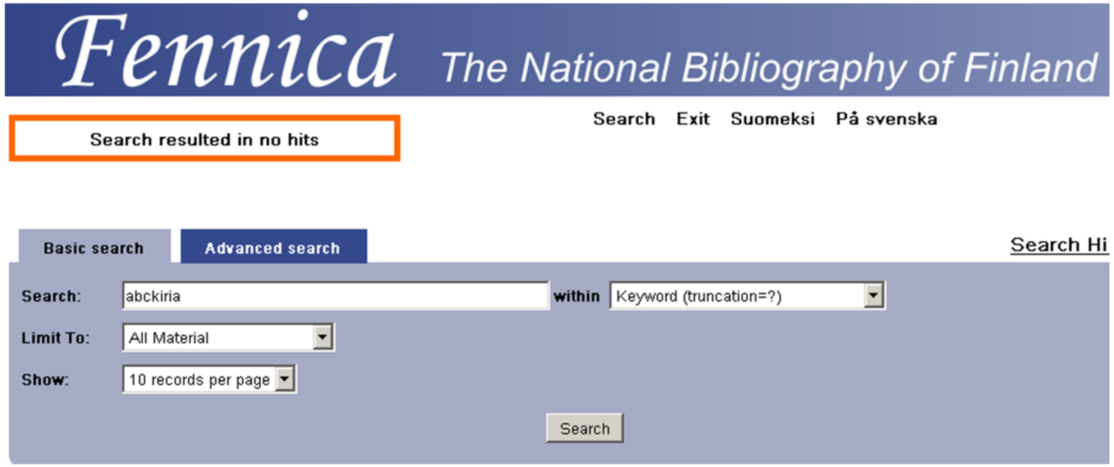
Planning and development



Road to semantics in digitised content

Uncataloged printed item

Undiscoverable



Fennica The National Bibliography of Finland

Search Exit Suomeksi På svenska

Search resulted in no hits

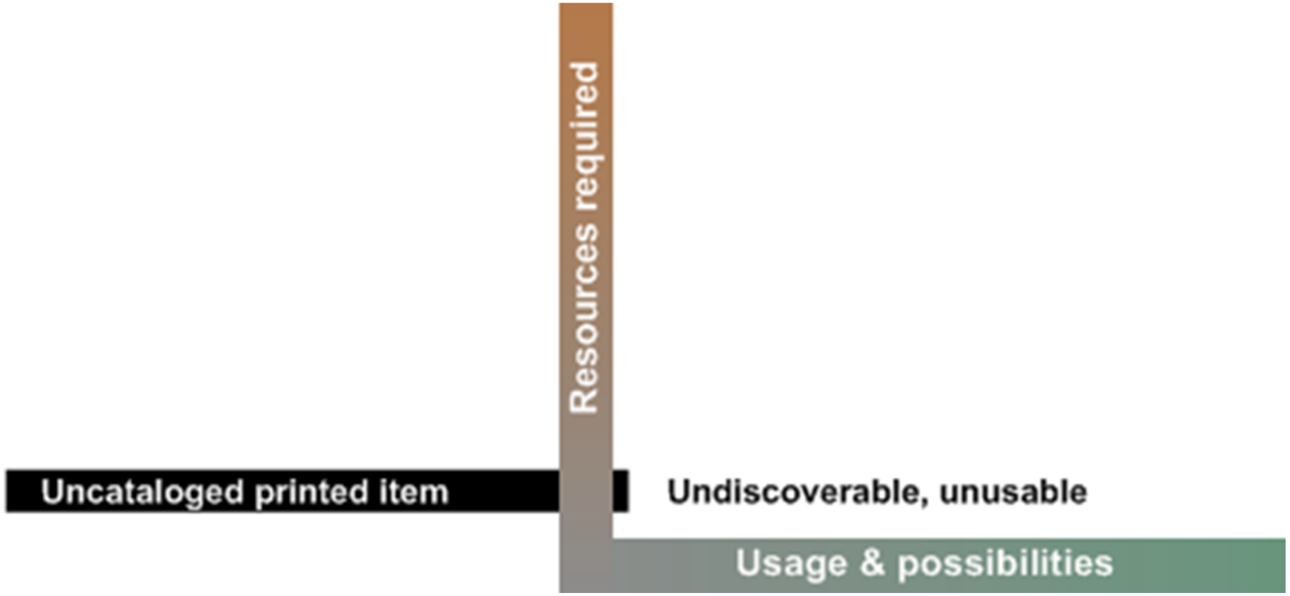
Basic search | Advanced search | Search Hi

Search: abckiria within Keyword (truncation=?)

Limit To: All Material

Show: 10 records per page

Search



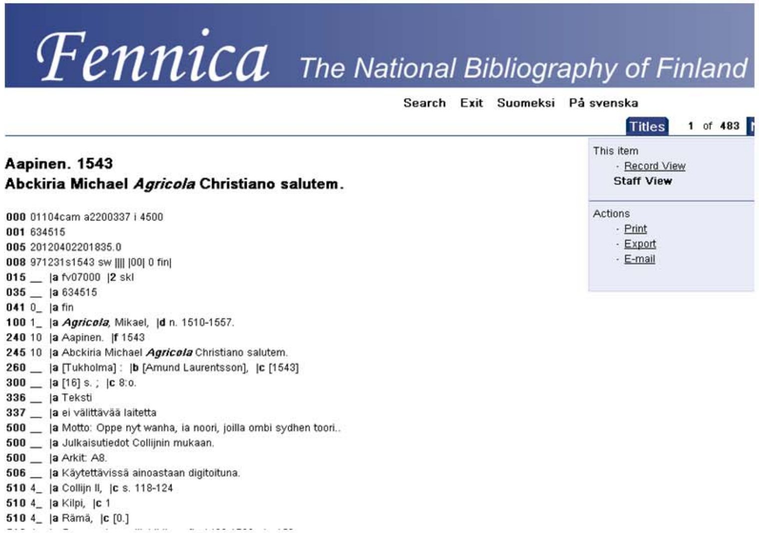
Resources required

Uncataloged printed item

Undiscoverable, unusable

Usage & possibilities

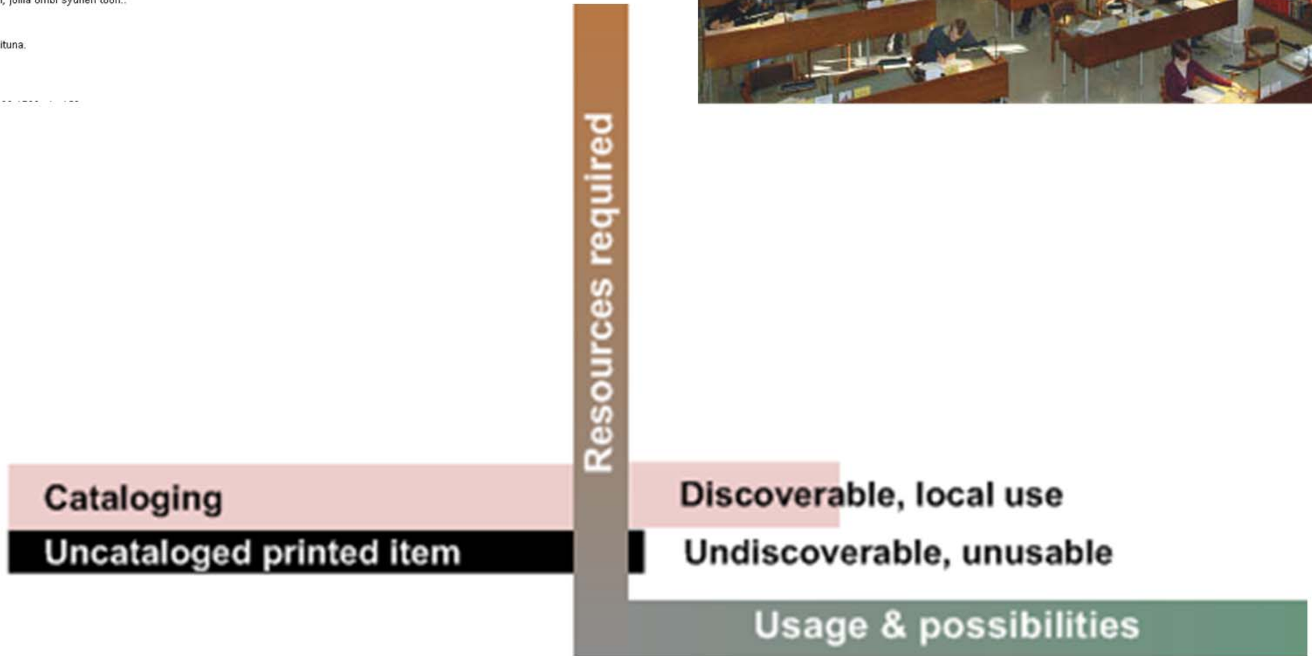
Cataloging, discoverable



Aapinen. 1543
Abckiria Michael *Agricola* Christiano salutem.

```
000 01104cam a2200337 i 4500
001 634515
005 20120402201835.0
008 971231s1543 sw |||| |00| 0 fin|
015 __  |a fv07000  |2 skl
035 __  |a 634515
041 0_  |a fin
100 1_  |a Agricola, Mikael,  |d n. 1510-1557.
240 10  |a Aapinen.  |f 1543
245 10  |a Abckiria Michael Agricola Christiano salutem.
260 __  |a [Tukholma] :  |b [Amund Laurentsson],  |c [1543]
300 __  |a [16] s. ;  |c 8:o.
336 __  |a Teksti
337 __  |a ei välittävää laitetta
500 __  |a Motto: Oppe nyt wanha, ia noori, joilla ombi sydhen toori..
500 __  |a Julkaisutiedot Collijnin mukaan.
500 __  |a Arkit: A8.
506 __  |a Käytettävissä ainoastaan digitoituna.
510 4_  |a Collijn II,  |c s. 118-124
510 4_  |a Kilpi,  |c 1
510 4_  |a Rämä,  |c [0.]
```

Local use



Resources required

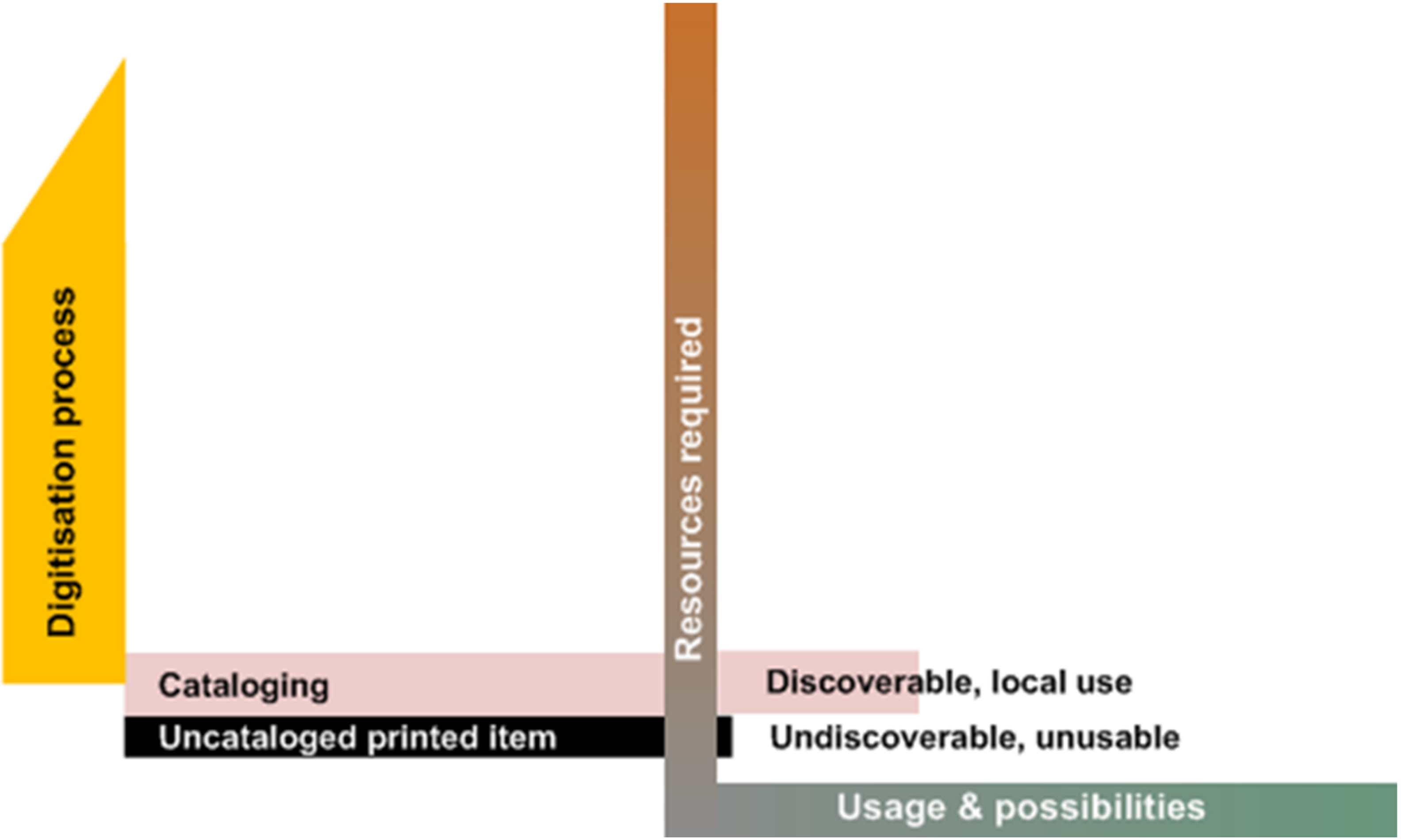
Cataloging

Uncataloged printed item

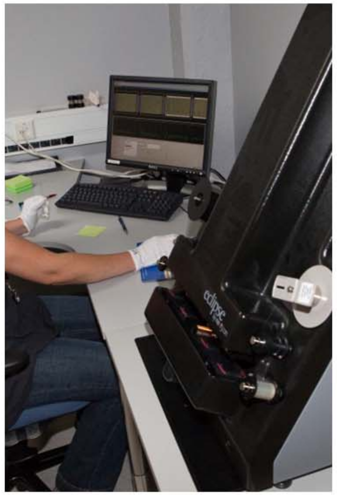
Discoverable, local use

Undiscoverable, unusable

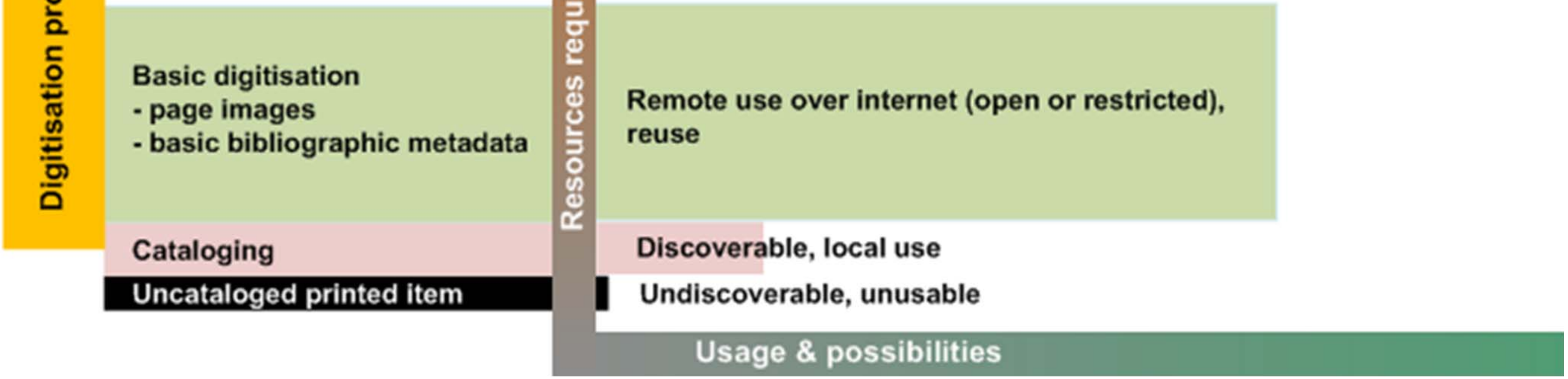
Usage & possibilities

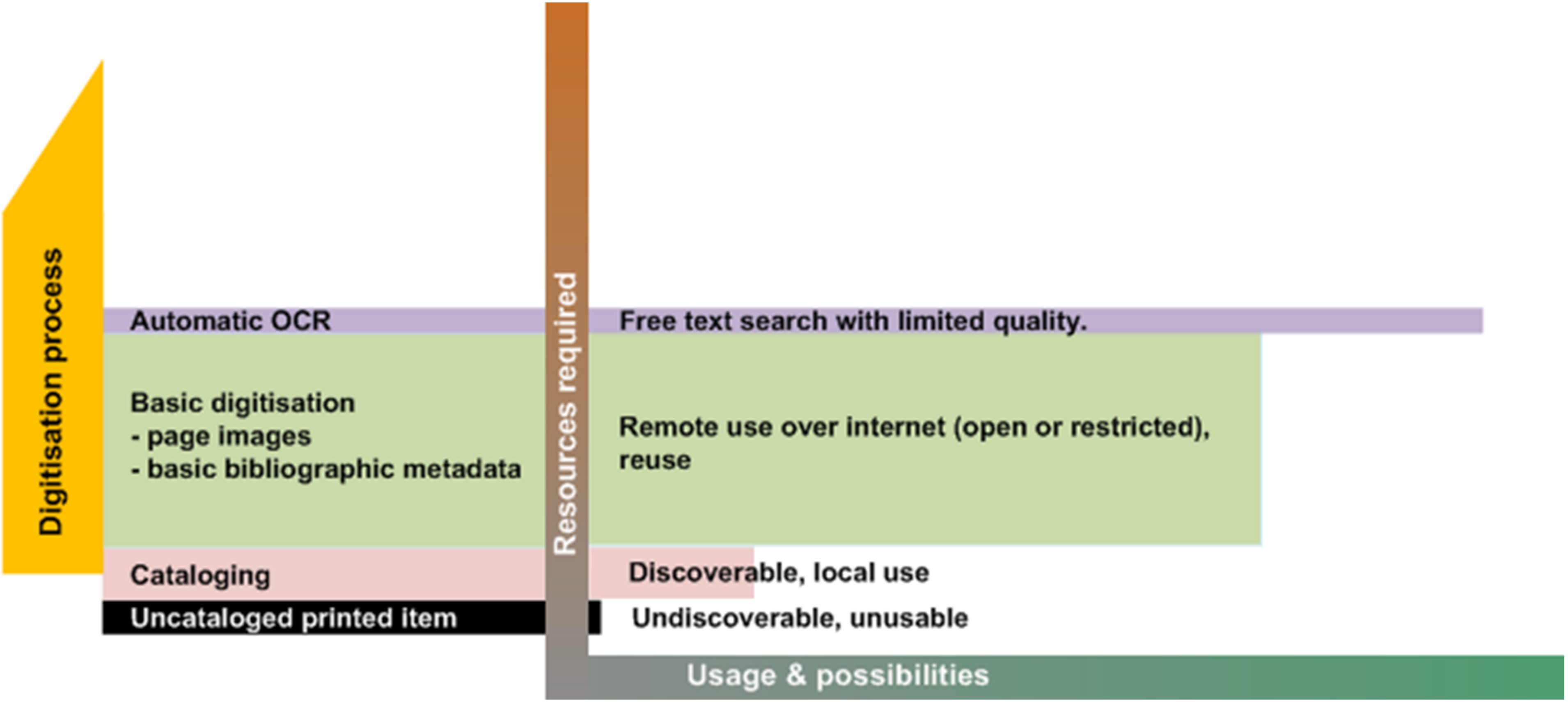

Digitisation process
Resources required
Cataloging
Discoverable, local use
Uncataloged printed item
Undiscoverable, unusable
Usage & possibilities

Basic digitisation



Huvudstadsbladet
No. 154
1897-04-10
8 pages





Digitisation process

Resources required

Automatic OCR

Free text search with limited quality.

Basic digitisation
- page images
- basic bibliographic metadata

Remote use over internet (open or restricted), reuse

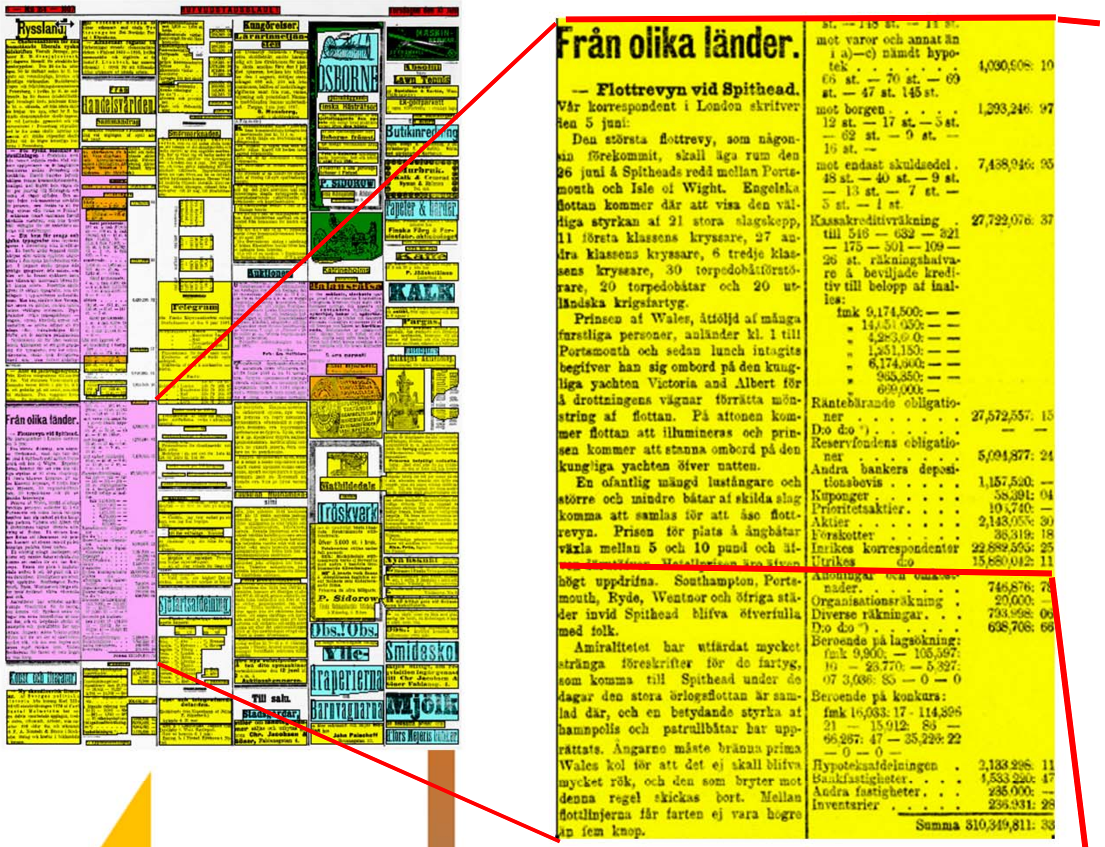
Cataloging

Discoverable, local use

Uncataloged printed item

Undiscoverable, unusable

Usage & possibilities



Från olika länder.

— Flottrevyn vid Spithead. Vår korrespondent i London skrifver den 5 juni:

Den största flottrevy, som någonsin förekommit, skall äga rum den 26 juni å Spitheads redd mellan Portsmouth och Isle of Wight. Engelska flottan kommer där att visa den väldiga styrkan af 21 stora slagskepp, 11 första klassens kryssare, 27 andra klassens kryssare, 6 tredje klassens kryssare, 30 torpedobåtförstörare, 20 torpedobåtar och 20 utländska krigsfartyg.

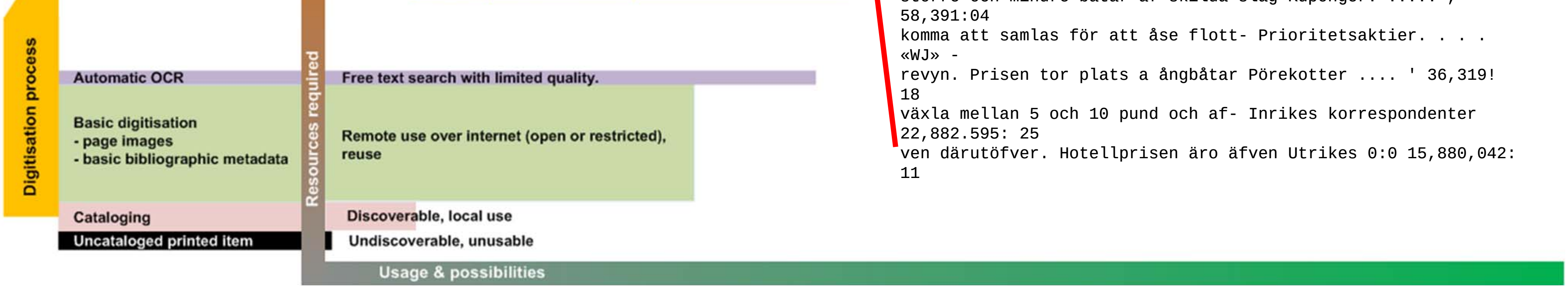
Prinsen af Wales, åtföljd af många furstliga personer, anländer kl. 1 till Portsmouth och sedan lunch intagits begifver han sig ombord på den kungliga yachten Victoria and Albert för å drottningens vägnar förrätta mönstring af flottan. På aftonen kommer flottan att illumineras och prinsen kommer att stanna ombord på den kungliga yachten öfver natten.

En ofantlig mängd lustångare och större och mindre båtar af skilda slag komma att samlas för att åse flottrevyn. Prisen för plats å ångbåtar växla mellan 5 och 10 pund och äfven därutöfver. Hotellprisen äro äfven högt uppdrifna. Southampton, Portsmouth, Ryde, Wentnor och öfriga städer invid Spithead blifva öfverfulla med folk.

Amiralitetet har utfärdat mycket stränga föreskrifter för de fartyg, som komma till Spithead under de dagar den stora örlogsflottan är samlad där, och en betydande styrka af hamnpolis och patrullbåtar har upprättats. Ångarne måste bränna prima Wales kol för att det ej skall blifva mycket rök, och den som bryter mot denna regel skickas bort. Mellan flottlinjerna får farten ej vara högre än fem knop.

st. — 118 st. — 11 st. mot varor och annat än i a)—c) nämdt hypotek	4,030,908: 10
66 st. — 70 st. — 69 st. — 47 st. 145 st. mot borgen	1,293,246: 97
12 st. — 17 st. — 3 st. — 62 st. — 9 st. — 16 st. — mot endast skuldsedel .	7,438,946: 95
48 st. — 40 st. — 9 st. — 13 st. — 7 st. — 3 st. — 1 st. Kassakreditivräkning .	27,722,076: 37
till 516 — 632 — 321 — 175 — 301 — 109 — 26 st. räkningshafvare å beviljade kreditiv till belopp af inalles: fmk 9,174,500: — — „ 14,451,050: — — „ 4,283,000: — — „ 1,251,150: — — „ 6,174,600: — — „ 965,850: — — „ 699,000: —	
Räntebärande obligationer	27,572,557: 15
D:o d:o *)	— —
Reservfondens obligationer	5,094,877: 24
Andra bankers depositionsbevis	1,157,520: —
Kuponger	58,391: 04
Prioritetsaktier	103,740: —
Aktier	2,143,055: 30
Förskotter	36,319: 18
Inrikes korrespondenter	22,882,595: 25
Utrikes d:o	15,880,042: 11
Aflöningar och omkostnader	746,876: 78
Organisationsräkning .	20,000: —
Diverse räkningar . . .	793,998: 06
D:o d:o *)	638,706: 66
Beroende på lagsökning: fmk 9,900: — 105,597: 10 — 23,770: — 5,327: 07 3,036: 85 — 0 — 0	
Beroende på konkurs: fmk 16,033: 17 - 114,398 21 — 15,912: 86 — 66,267: 47 — 35,228: 22 — 0 — 0 —	
Hypoteksafdelningen .	2,133,298: 11
Bankfastigheter	1,533,220: 47
Andra fastigheter . . .	235,000: —
Inventarier	236,931: 28
Summa	310,349,811: 33

```
****\u25a0 ..... 17,903,378; il
290 st — 63 "st. —163
Från olika länder. st. — 118 st. 11 st
>S*Ö|1  lÖ?1n£i&* mot varor och annat än
li all UlEl\a lailUöl . i a)-c) nänidt hypo-
tek 4,030,S08: 10
_. ,. \u25a0»«_,« _ 66 st. — 70 st. — 69
— Flottrevyn via Spitaead. 8t. _ 47 st. 145 st
Vår korrespondent i London skriiver mot borgen .....
1,203,246. 97
den 5 juni: 12 st. — 17 st — 3 st
Den största flottrevy, som någon- -~ ® st. — 9 at. -
sin förekommit, skall äga ram den ?~Z1 , ,j a , ~ron(, A-
-« . . o o, ... , j, „ ,\u25a0»„_*„. mot endast skuldsedel.
,438.910: 9a
26 juni k Spitheads redd mellan Ports- .g gL-~ 40 st — 9 st
mouth och Isle of Wight. Engelska _ 13 st _ 7 st. —
flottan kommer där att visa den väl' 5 st — 1 st.
diga styrkan af 21 stora slagskepp, Eassakreditivräkning .
27,722,076: 37
11 första klassens kryssare, 27 an- — — 321
dra klassens kryssare, 6 tredje kläs- 26 A rakningshafva.
sens kryssare, 30 torpedobåttorstö- re & beviljade kredi-
rare, 20 torpedobåtar och 20 ut- tiv till belopp af inal-
lindska krigsfartyg. ies:
Prinsen af Wales, åtföljd af många 14151 050: -
furstliga personer, anländer kl. 1 till " 4*283$ 0:
Portsmouth och sedan lunch intagits „ 1,251,150:
begitver han sig ombord på den kung- « alls ««v """ —
liga yachten Victoria and Albert för » oqo: — ~"
å drottningens vägnar förrätta mön- Käntebärande' obligaöo-
string af flottan. På aftonen kom- ner 27,572,557: 15
mer flottan att illumineras och prin- D:o d:o>*) . .... —
sen kommer att stanna ombord på den obligat». u
kungliga yachten öfver natten. Andra "bankers" deposi-
En ofantlig mängd lustångare och tionsbevis 1,157,520: —
större och mindre båtar af skilda slag Kuponger. ..... ,
58,391:04
komma att samlas för att åse flott- Prioritetsaktier. . . .
«WJ» -
revyn. Prisen tor plats a ångbåtar Pörekotter .... ' 36,319!
18
växla mellan 5 och 10 pund och af- Inrikes korrespondenter
22,882.595: 25
ven därutöfver. Hotellprisen äro äfven Utrikes 0:0 15,880,042:
11
```



Digitisation process

Resources required

Automatic OCR

Free text search with limited quality.

Basic digitisation
- page images
- basic bibliographic metadata

Remote use over internet (open or restricted), reuse

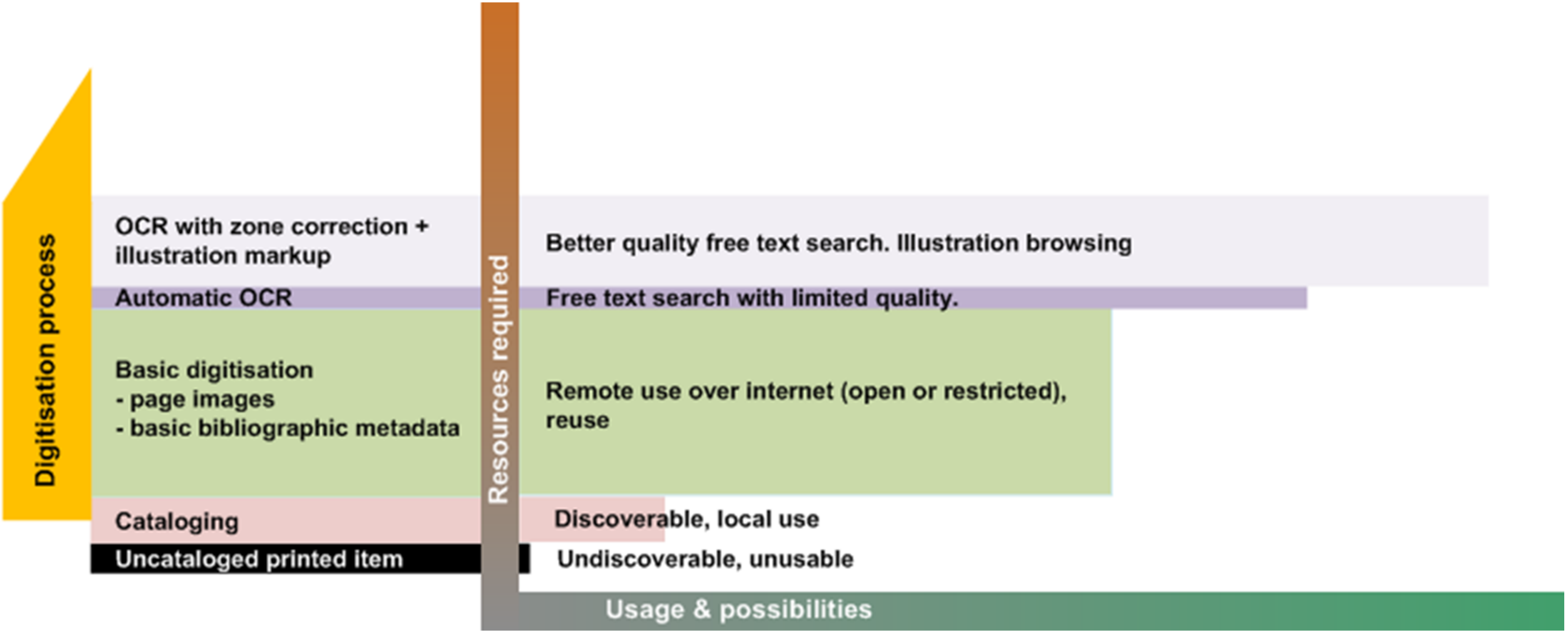
Cataloging

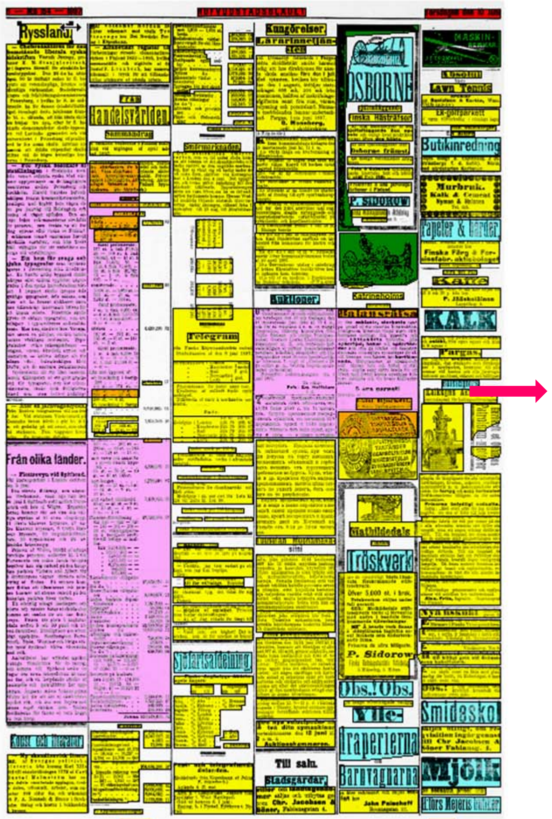
Discoverable, local use

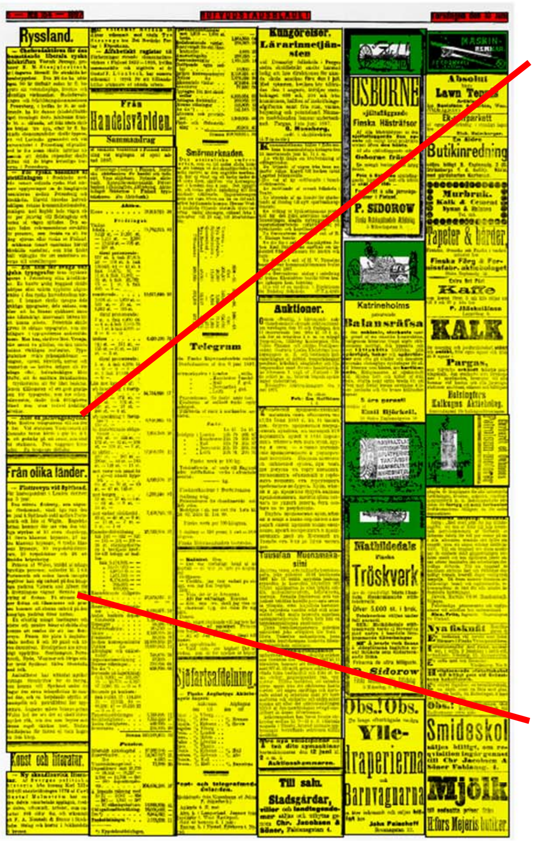
Uncataloged printed item

Undiscoverable, unusable

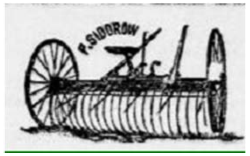
Usage & possibilities

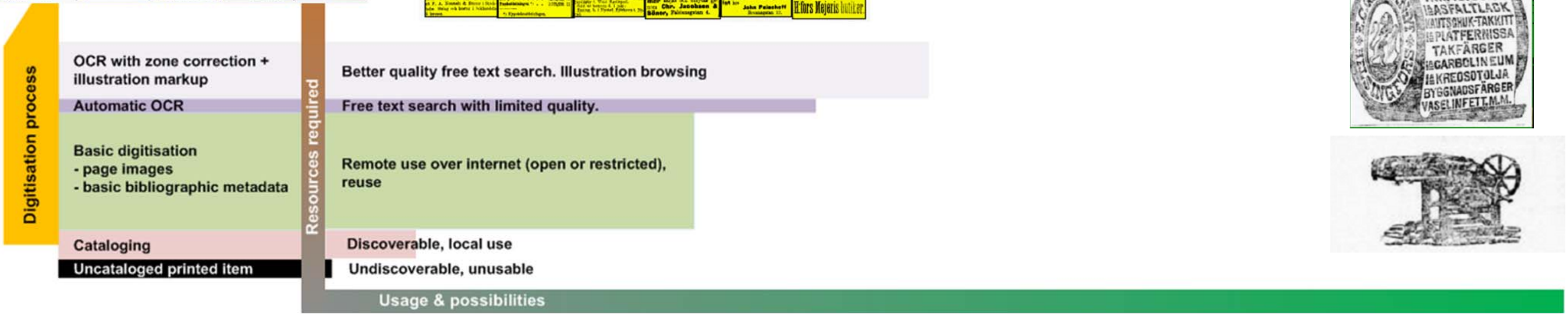

Digitisation process
Resources required
OCR with zone correction + illustration markup
Better quality free text search. Illustration browsing
Automatic OCR
Free text search with limited quality.
Basic digitisation
- page images
- basic bibliographic metadata
Remote use over internet (open or restricted), reuse
Cataloging
Discoverable, local use
Uncataloged printed item
Undiscoverable, unusable
Usage & possibilities





Från olika länder.
— Flottrevyn vid Spithead.
Vår korrespondent i London skritver den 5 juni:
Den största flottrevy, som någonsin förekommit, skall äga ram den 26 juni å Spitheads redd mellan Portsmouth och Isle of Wight. Engelska flottan kommer där att visa den väl' diga styrkan af 21 stora slagskepp, 11 första klassens kryssare, 27 andra klassens kryssare, 6 tredje klassens kryssare, 30 torpedohåitorstörare, 20 torpedobåtar och 20 utländska krigsfartyg. I
Prinsen af Wales, åtföljd af många furstliga personer, anländer kl. 1 till Portsmouth och sedan lunch intagits begifver han sig ombord på den kungliga yachten Victoria and Albert för å drottningens vagnar förrätta mönstring af flottan. På aftonen kommer flottan att illumineras och prinsen kommer att stanna ombord på den kungliga yachten öfver natten.
En ofantlig mängd lustångare och större och mindre båtar af skilda slag komma att samlas för att åse flottrevyn. Prisen för plats å ångbåtar växla mellan 5 och 10 pund och äl-





Digitisation process

Resources required

OCR with zone correction + illustration markup

Better quality free text search. Illustration browsing

Automatic OCR

Free text search with limited quality.

Basic digitisation
- page images
- basic bibliographic metadata

Remote use over internet (open or restricted), reuse

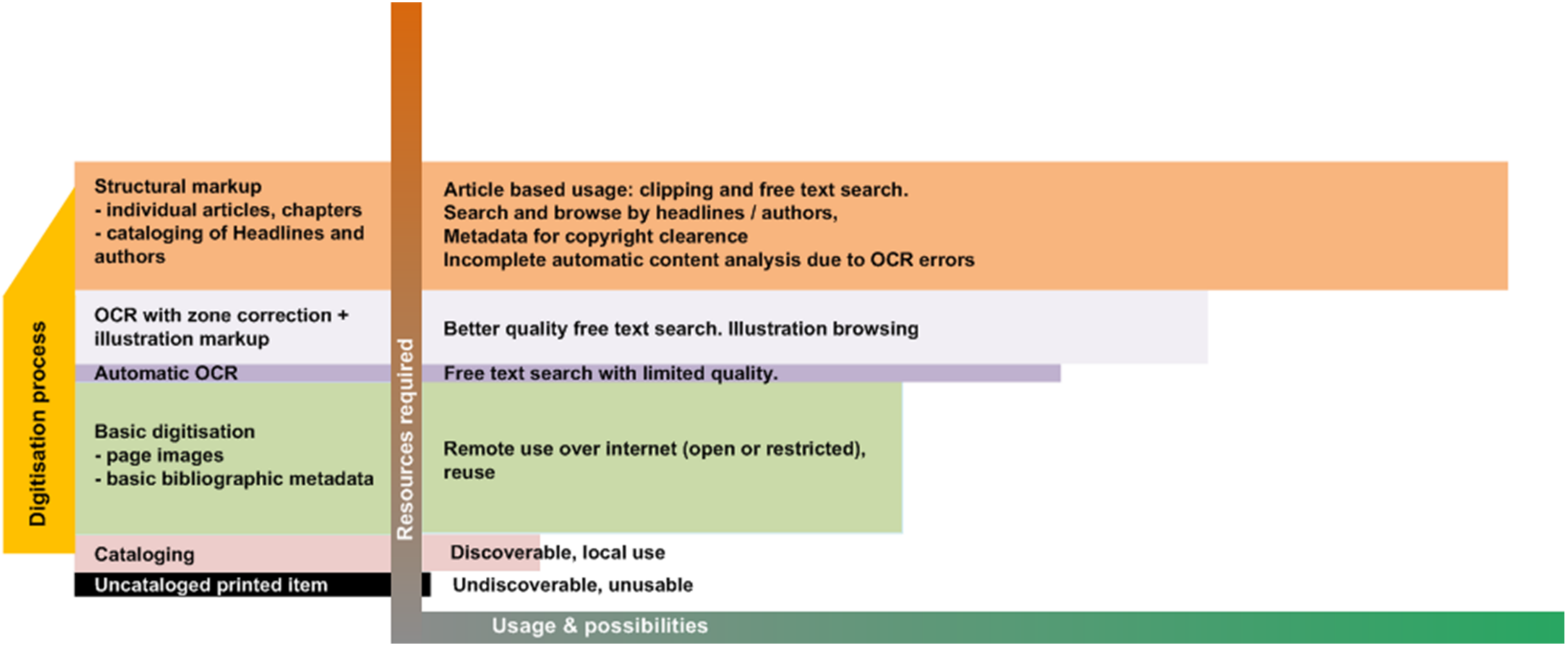
Cataloging

Discoverable, local use

Uncataloged printed item

Undiscoverable, unusable

Usage & possibilities



Digitisation process

Resources required

Usage & possibilities

Structural markup
- individual articles, chapters
- cataloging of Headlines and authors

Article based usage: clipping and free text search.
Search and browse by headlines / authors,
Metadata for copyright clearence
Incomplete automatic content analysis due to OCR errors

OCR with zone correction + illustration markup

Better quality free text search. Illustration browsing

Automatic OCR

Free text search with limited quality.

Basic digitisation
- page images
- basic bibliographic metadata

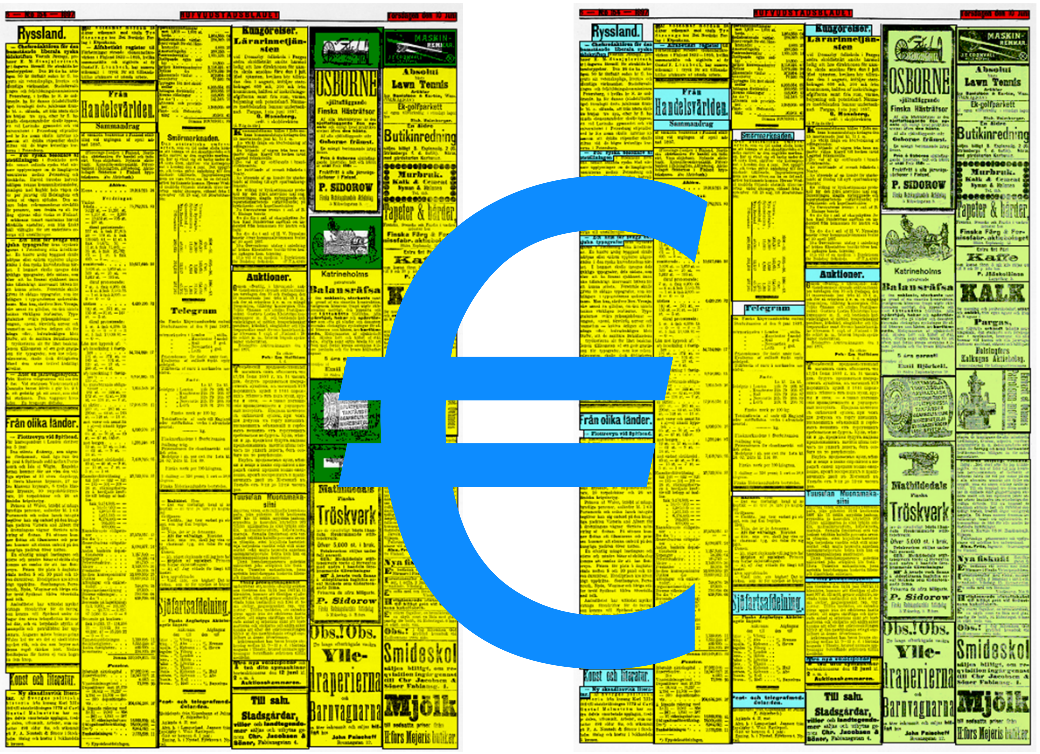
Remote use over internet (open or restricted), reuse

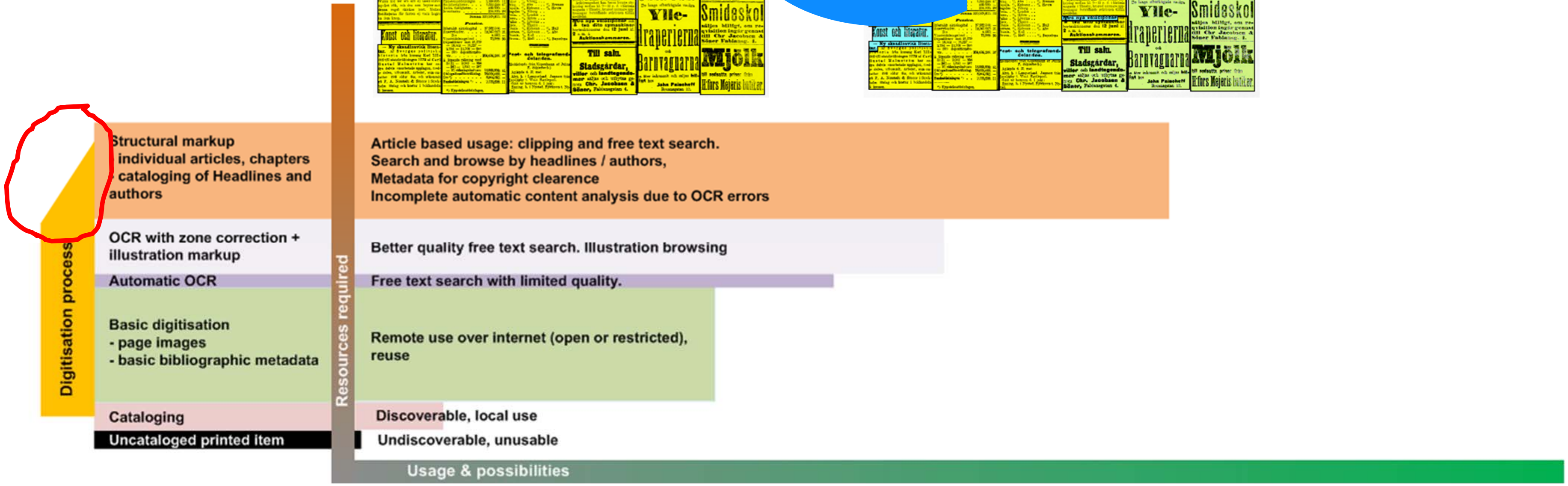
Cataloging

Discoverable, local use

Uncataloged printed item

Undiscoverable, unusable





Resources required

Digitisation process

Structural markup
- individual articles, chapters
- cataloging of Headlines and authors

Article based usage: clipping and free text search.
Search and browse by headlines / authors,
Metadata for copyright clearence
Incomplete automatic content analysis due to OCR errors

OCR with zone correction + illustration markup

Better quality free text search. Illustration browsing

Automatic OCR

Free text search with limited quality.

Basic digitisation
- page images
- basic bibliographic metadata

Remote use over internet (open or restricted), reuse

Cataloging

Discoverable, local use

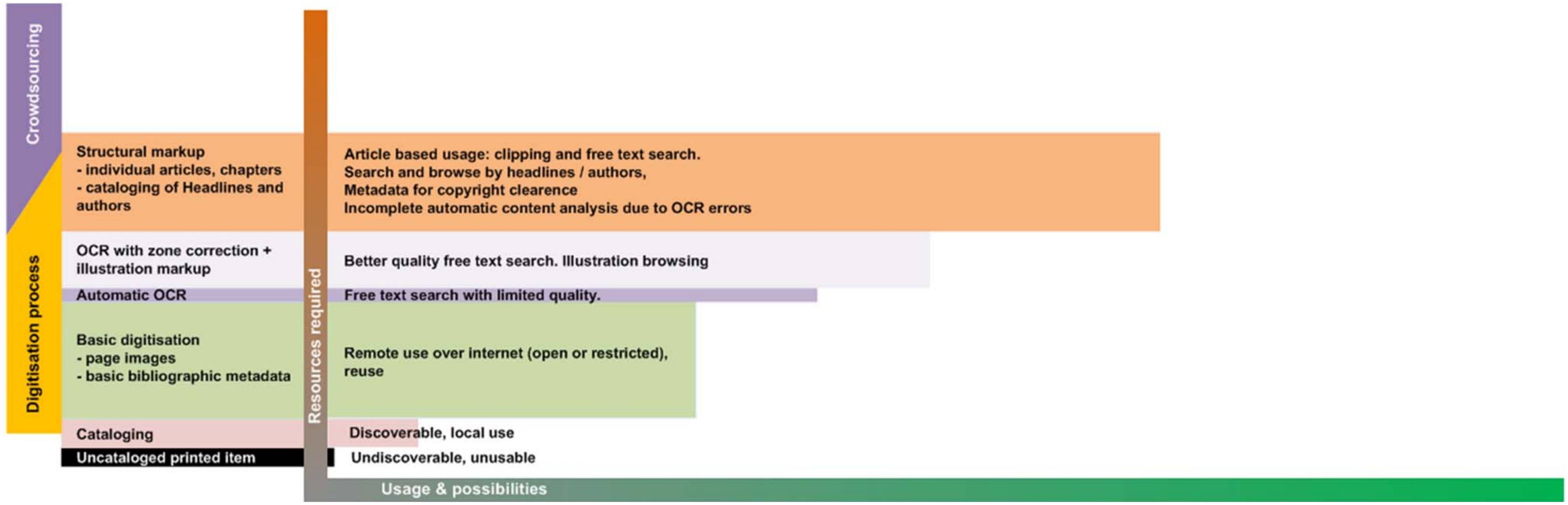
Uncataloged printed item

Undiscoverable, unusable

Usage & possibilities

Talkoot

A talkoot is a Finnish custom involving a group of people gathering to work together unpaid, for instance to build or repair something, similar to a bee or a barn raising in English.



Crowdsourcing

Digitisation process

Resources required

Usage & possibilities

Structural markup
- individual articles, chapters
- cataloging of Headlines and authors

Article based usage: clipping and free text search.
Search and browse by headlines / authors,
Metadata for copyright clearence
Incomplete automatic content analysis due to OCR errors

OCR with zone correction + illustration markup

Better quality free text search. Illustration browsing

Automatic OCR

Free text search with limited quality.

Basic digitisation
- page images
- basic bibliographic metadata

Remote use over internet (open or restricted), reuse

Cataloging

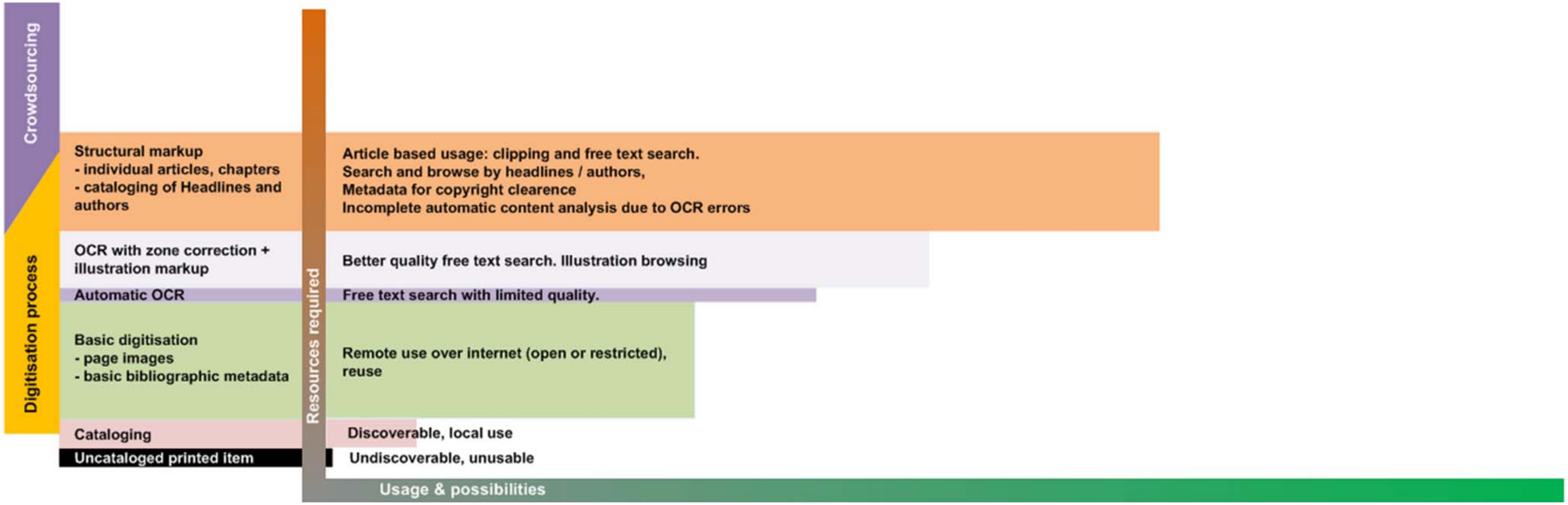
Discoverable, local use

Uncataloged printed item

Undiscoverable, unusable

KUVATALKOOT, 2011-2013

- Extend the digital object presentation system
- Use crowdsourcing for marking articles/images
- For casual users, enthusiasts (genealogy etc) and researchers
- Adding metadata to marked up items
  - Different levels of metadata
  - Leveraging ONKI.FI (National Onthology Service)



Crowdsourcing

Digitisation process

Resources required

Structural markup
- individual articles, chapters
- cataloging of Headlines and authors

Article based usage: clipping and free text search.
Search and browse by headlines / authors,
Metadata for copyright clearence
Incomplete automatic content analysis due to OCR errors

OCR with zone correction + illustration markup

Better quality free text search. Illustration browsing

Automatic OCR

Free text search with limited quality.

Basic digitisation
- page images
- basic bibliographic metadata

Remote use over internet (open or restricted), reuse

Cataloging

Discoverable, local use

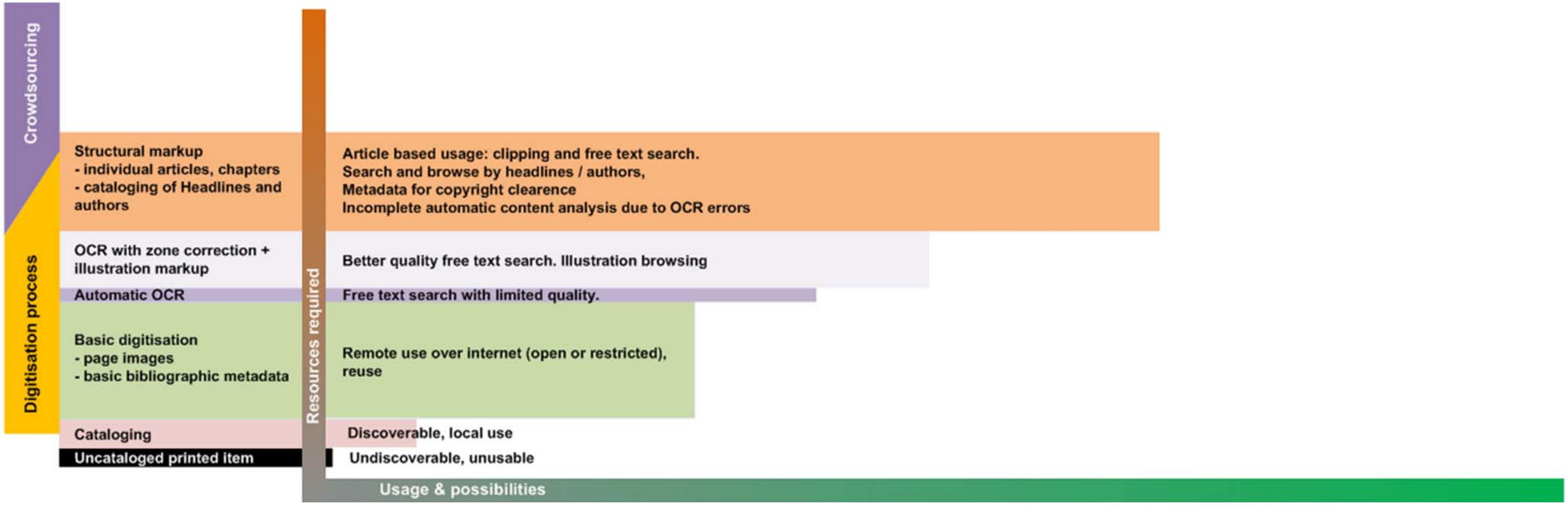
Uncataloged printed item

Undiscoverable, unusable

Usage & possibilities

KUVATALKOOT, 2011-2013 (2)

- Results shared with the Library and the Community
- 2nd order crowdsourcing?
  - Allow users to create markup tasks for other users.
- Youtube –like embedded player that allows easy reuse on web?
- Under construction



Structural markup
- individual articles, chapters
- cataloging of Headlines and authors

Article based usage: clipping and free text search.
Search and browse by headlines / authors,
Metadata for copyright clearence
Incomplete automatic content analysis due to OCR errors

OCR with zone correction + illustration markup

Better quality free text search. Illustration browsing

Automatic OCR

Free text search with limited quality.

Basic digitisation
- page images
- basic bibliographic metadata

Remote use over internet (open or restricted), reuse

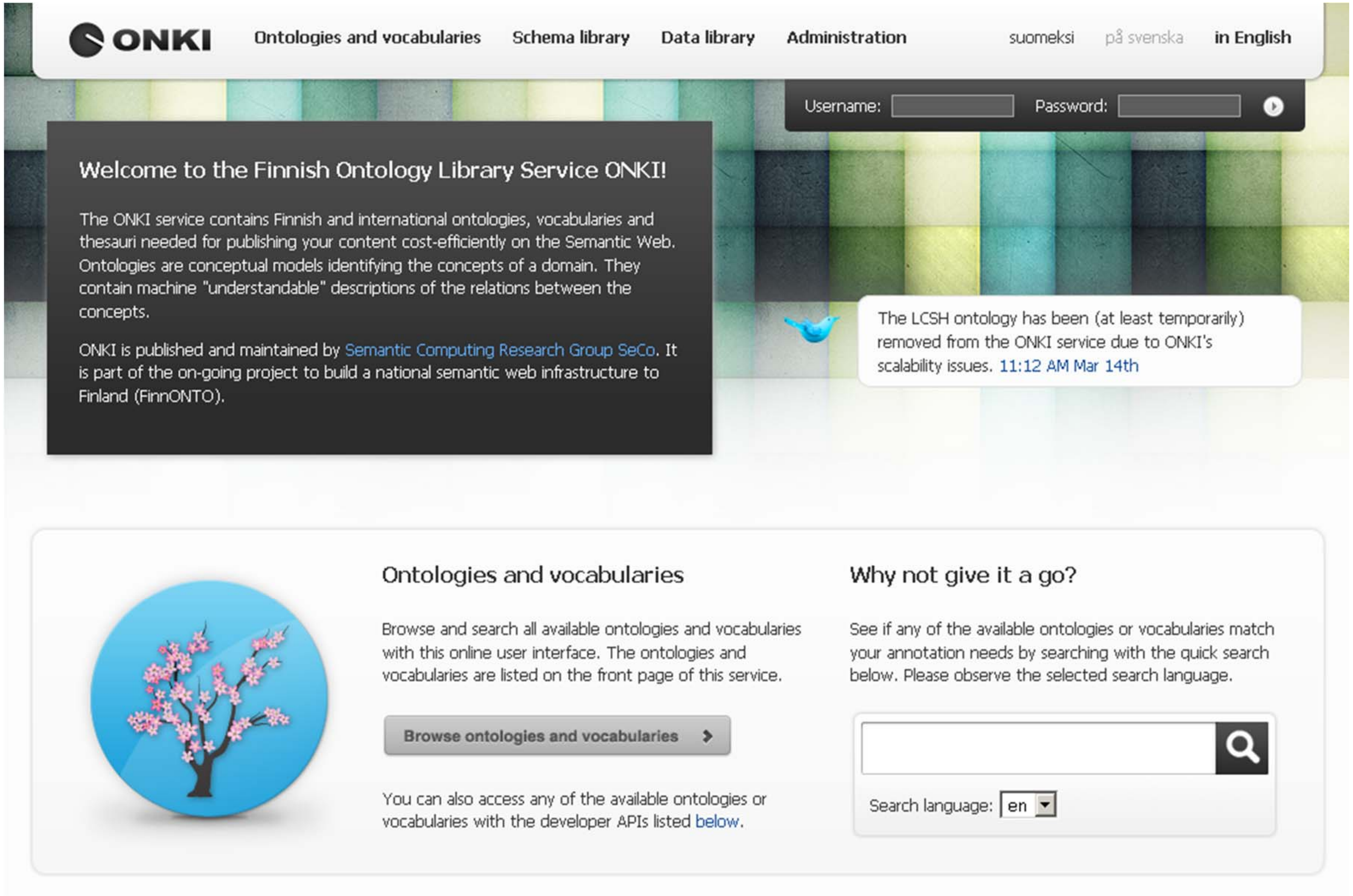
Cataloging

Discoverable, local use

Uncataloged printed item

Undiscoverable, unusable

National Ontology Service ONKI.FI



Welcome to the Finnish Ontology Library Service ONKI!

The ONKI service contains Finnish and international ontologies, vocabularies and thesauri needed for publishing your content cost-efficiently on the Semantic Web. Ontologies are conceptual models identifying the concepts of a domain. They contain machine "understandable" descriptions of the relations between the concepts.

ONKI is published and maintained by Semantic Computing Research Group SeCo. It is part of the on-going project to build a national semantic web infrastructure to Finland (FinnONTO).

The LCSH ontology has been (at least temporarily) removed from the ONKI service due to ONKI's scalability issues. 11:12 AM Mar 14th

Ontologies and vocabularies

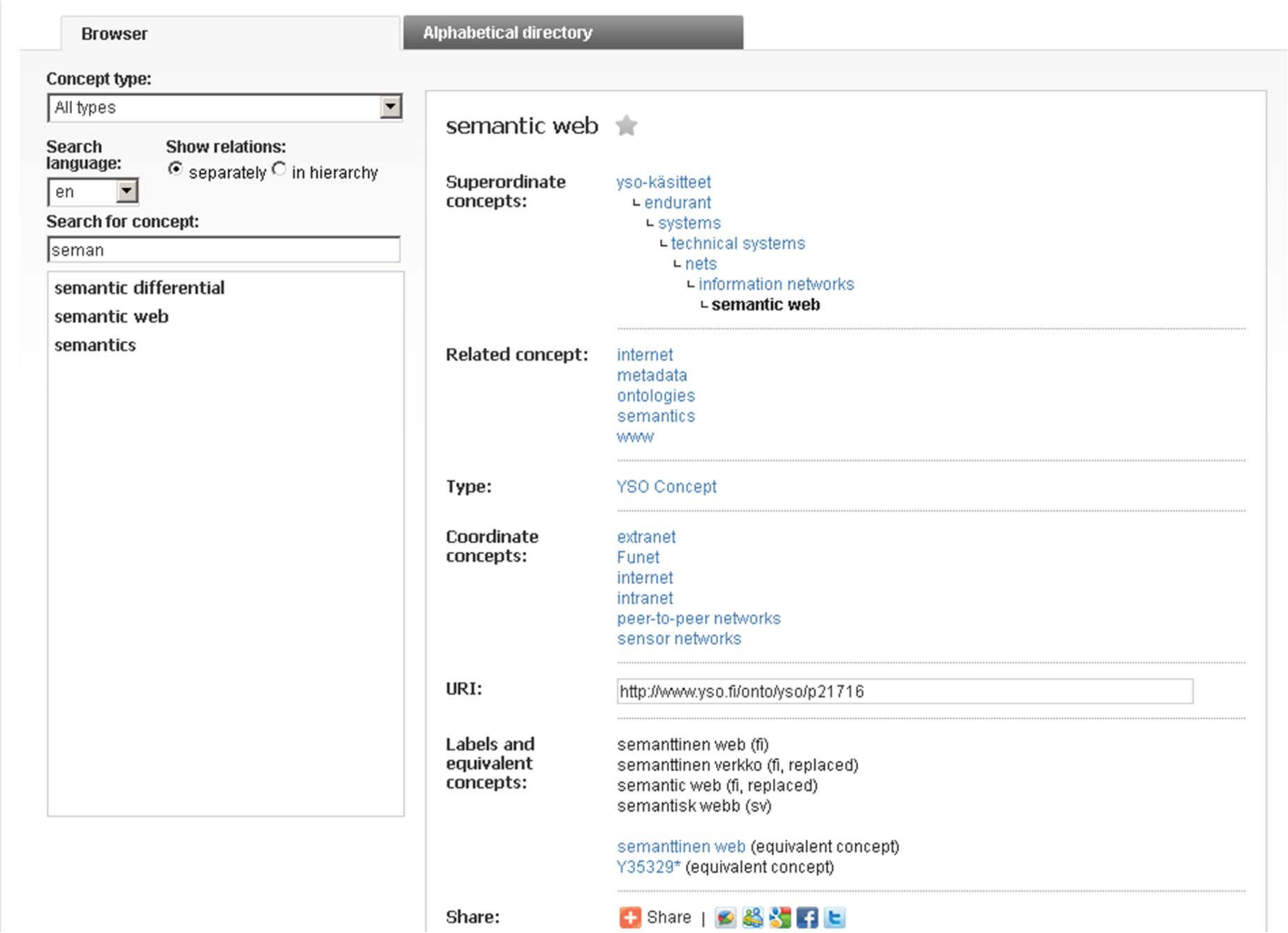
Browse and search all available ontologies and vocabularies with this online user interface. The ontologies and vocabularies are listed on the front page of this service.

You can also access any of the available ontologies or vocabularies with the developer APIs listed below.

Why not give it a go?

See if any of the available ontologies or vocabularies match your annotation needs by searching with the quick search below. Please observe the selected search language.

Search language: en



Concept type: All types

Search language: en

Show relations: ● separately ○ in hierarchy

Search for concept: seman

semantic differential
semantic web
semantics

semantic web

Superordinate concepts:
yso-käsitteet
└ endurant
└ systems
└ technical systems
└ nets
└ information networks
└ **semantic web**

Related concept:
internet
metadata
ontologies
semantics
www

Type: YSO Concept

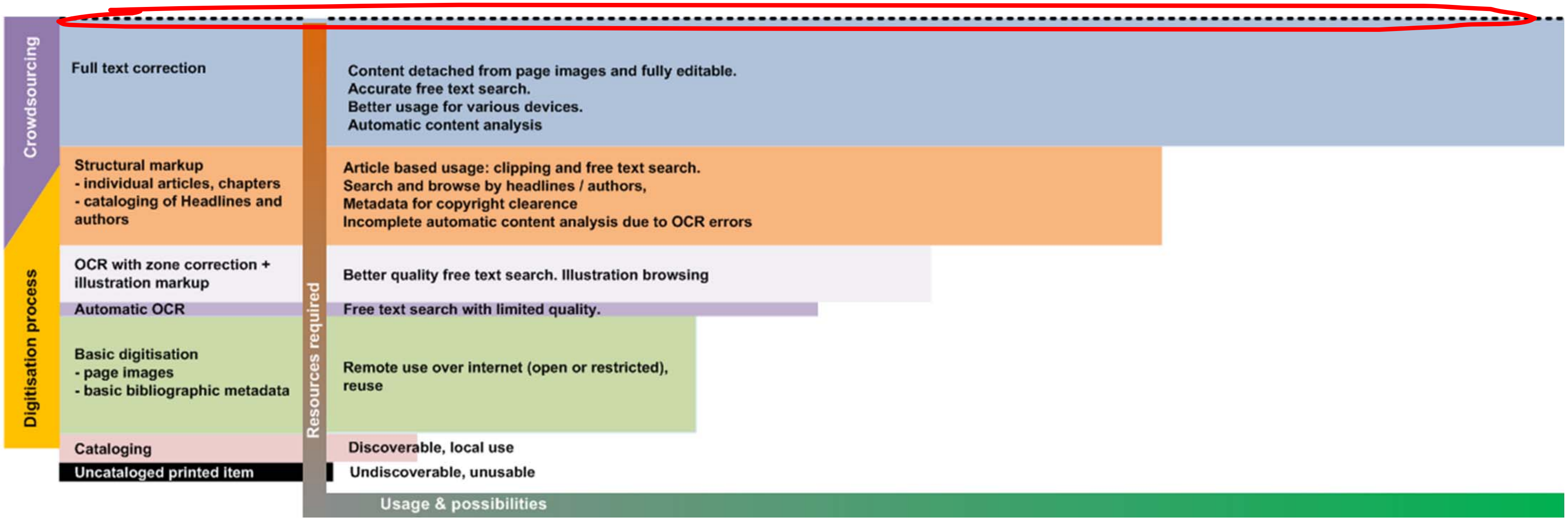
Coordinate concepts:
extranet
Funet
internet
intranet
peer-to-peer networks
sensor networks

URI: http://www.yso.fi/onto/yso/p21716

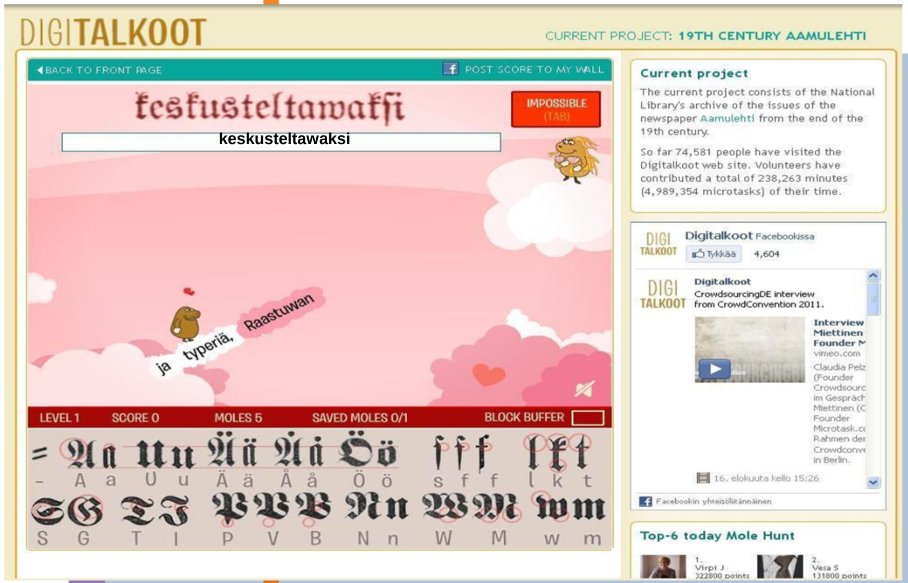
Labels and equivalent concepts:
semanttinen web (fi)
semanttinen verkko (fi, replaced)
semantic web (fi, replaced)
semantisk webb (sv)

semanttinen web (equivalent concept)
Y35329* (equivalent concept)

Share:


Crowdsourcing
Digitisation process
Resources required
Full text correction
Content detached from page images and fully editable.
Accurate free text search.
Better usage for various devices.
Automatic content analysis
Structural markup
- individual articles, chapters
- cataloging of Headlines and authors
Article based usage: clipping and free text search.
Search and browse by headlines / authors,
Metadata for copyright clearence
Incomplete automatic content analysis due to OCR errors
OCR with zone correction + illustration markup
Better quality free text search. Illustration browsing
Automatic OCR
Free text search with limited quality.
Basic digitisation
- page images
- basic bibliographic metadata
Remote use over internet (open or restricted), reuse
Cataloging
Discoverable, local use
Uncataloged printed item
Undiscoverable, unusable
Usage & possibilities

Gamified text correction with Microtask



DIGITALKOOT

CURRENT PROJECT: **19TH CENTURY AAMULEHTI**

◂ BACK TO FRONT PAGE — POST SCORE TO MY WALL

keskusteltawaksi

IMPOSSIBLE (TAB)

ja typeriä, Raastuwan

LEVEL 1 — SCORE 0 — MOLES 5 — SAVED MOLES 0/1 — BLOCK BUFFER

Current project

The current project consists of the National Library's archive of the issues of the newspaper Aamulehti from the end of the 19th century.

So far 74,581 people have visited the Digitalkoot web site. Volunteers have contributed a total of 238,263 minutes (4,989,354 microtasks) of their time.

Digitalkoot Facebookissa — Tykkää 4,604

Digitalkoot
CrowdsourcingDE interview from CrowdConvention 2011.

Interview Miettinen Founder M
vimeo.com
Claudia Pelz (Founder Crowdsourc im Gespräch Miettinen (C Founder Microtask.c Rahmen der Crowdconve in Berlin.

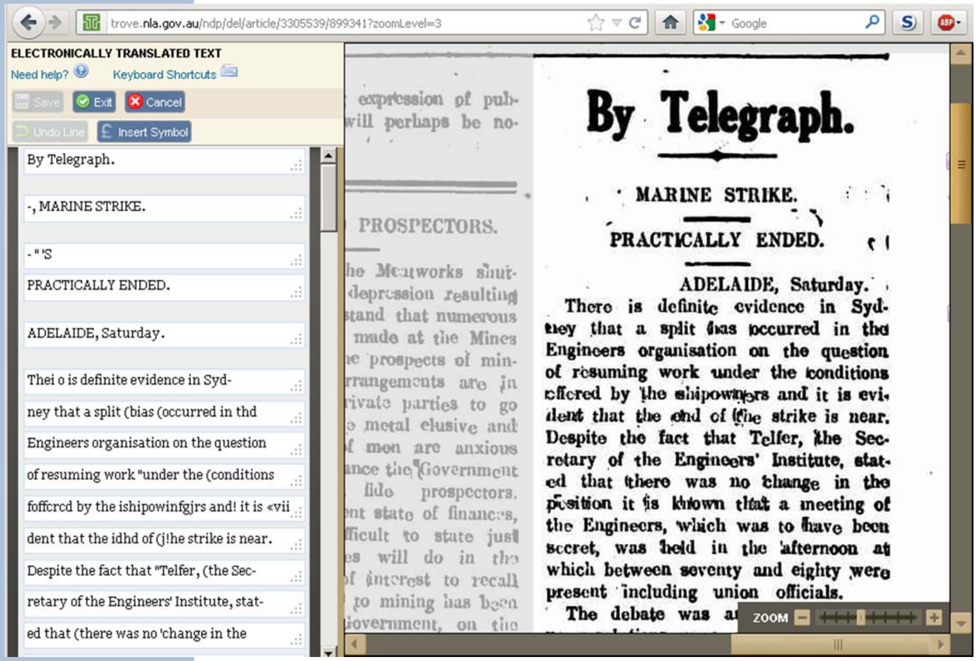
16. elokuuta kello 15:26

Facebookin yhteisöliitännäinen

Top-6 today Mole Hunt

1. Virpi J 322800 points
2. Vesa S 131800 points

Full text correction example from National Library of Australia



trove.nla.gov.au/ndp/del/article/3305539/899341?zoomLevel=3

ELECTRONICALLY TRANSLATED TEXT

Need help? — Keyboard Shortcuts

Save — Exit — Cancel — Undo Line — Insert Symbol

By Telegraph.
-, MARINE STRIKE.
- " 'S
PRACTICALLY ENDED.
ADELAIDE, Saturday.
Thei o is definite evidence in Syd-
ney that a split (bias (occurred in thd
Engineers organisation on the question
of resuming work "under the (conditions
fofferd by the ishipowinfgjrs and! it is «vii
dent that the idhd of (j!he strike is near.
Despite the fact that "Telfer, (the Sec-
retary of the Engineers' Institute, stat-
ed that (there was no 'change in the

By Telegraph.

MARINE STRIKE.

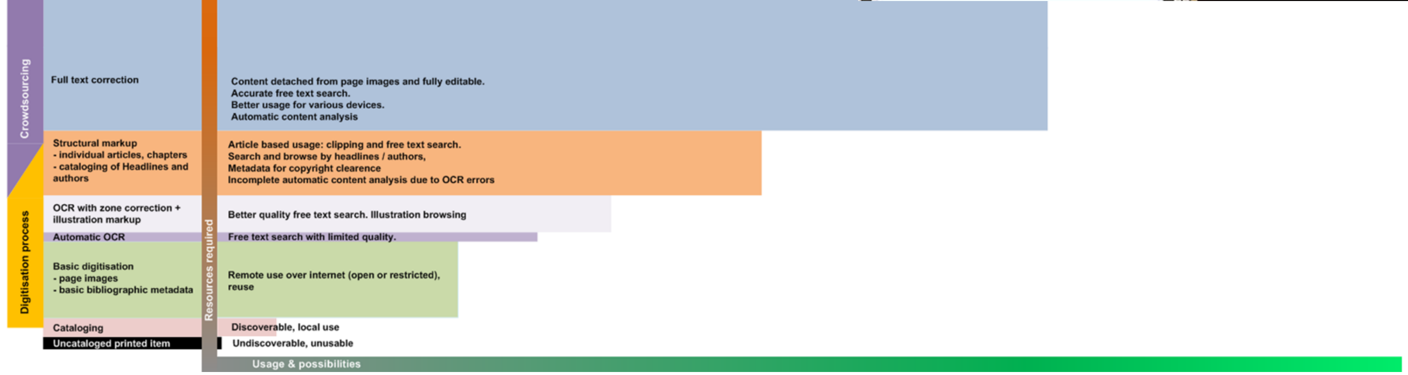
PRACTICALLY ENDED.

ADELAIDE, Saturday.

There is definite evidence in Sydney that a split has occurred in the Engineers organisation on the question of resuming work under the conditions offered by the shipowners and it is evident that the end of the strike is near. Despite the fact that Telfer, the Secretary of the Engineers' Institute, stated that there was no change in the position it is known that a meeting of the Engineers, which was to have been secret, was held in the afternoon at which between seventy and eighty were present including union officials. The debate was a

PROSPECTORS.

ZOOM



Full text correction

Content detached from page images and fully editable. Accurate free text search. Better usage for various devices. Automatic content analysis

Structural markup
- individual articles, chapters
- cataloging of Headlines and authors

Article based usage: clipping and free text search. Search and browse by headlines / authors, Metadata for copyright clearence Incomplete automatic content analysis due to OCR errors

OCR with zone correction + illustration markup

Better quality free text search. Illustration browsing

Automatic OCR

Free text search with limited quality.

Basic digitisation
- page images
- basic bibliographic metadata

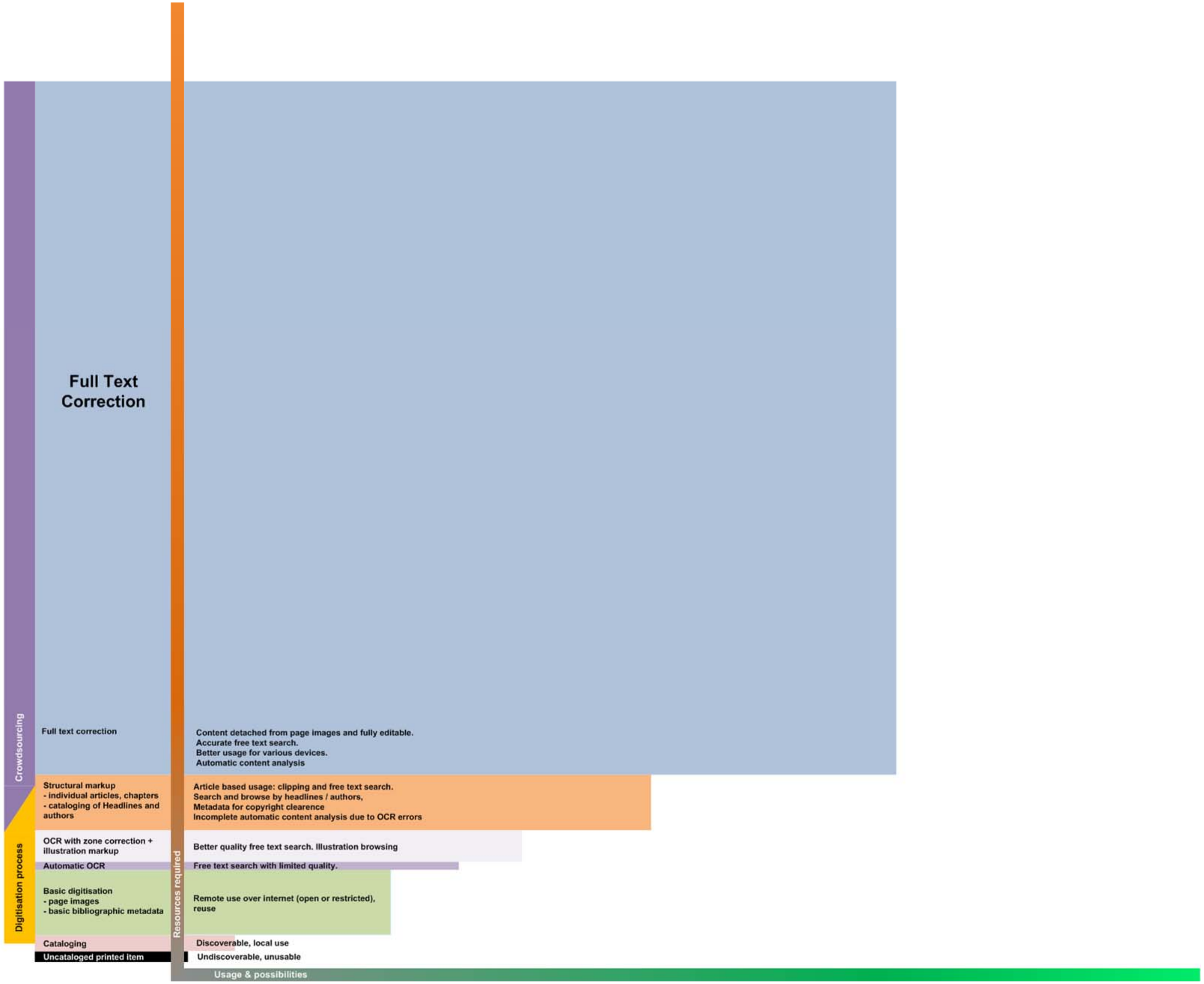
Remote use over internet (open or restricted), reuse

Cataloging

Discoverable, local use

Uncataloged printed item

Undiscoverable, unusable



Full Text Correction

Crowdsourcing

Digitisation process

Resources required

Full text correction

Content detached from page images and fully editable.
Accurate free text search.
Better usage for various devices.
Automatic content analysis

Structural markup
- individual articles, chapters
- cataloging of Headlines and authors

Article based usage: clipping and free text search.
Search and browse by headlines / authors,
Metadata for copyright clearence
Incomplete automatic content analysis due to OCR errors

OCR with zone correction + illustration markup

Better quality free text search. Illustration browsing

Automatic OCR

Free text search with limited quality.

Basic digitisation
- page images
- basic bibliographic metadata

Remote use over internet (open or restricted), reuse

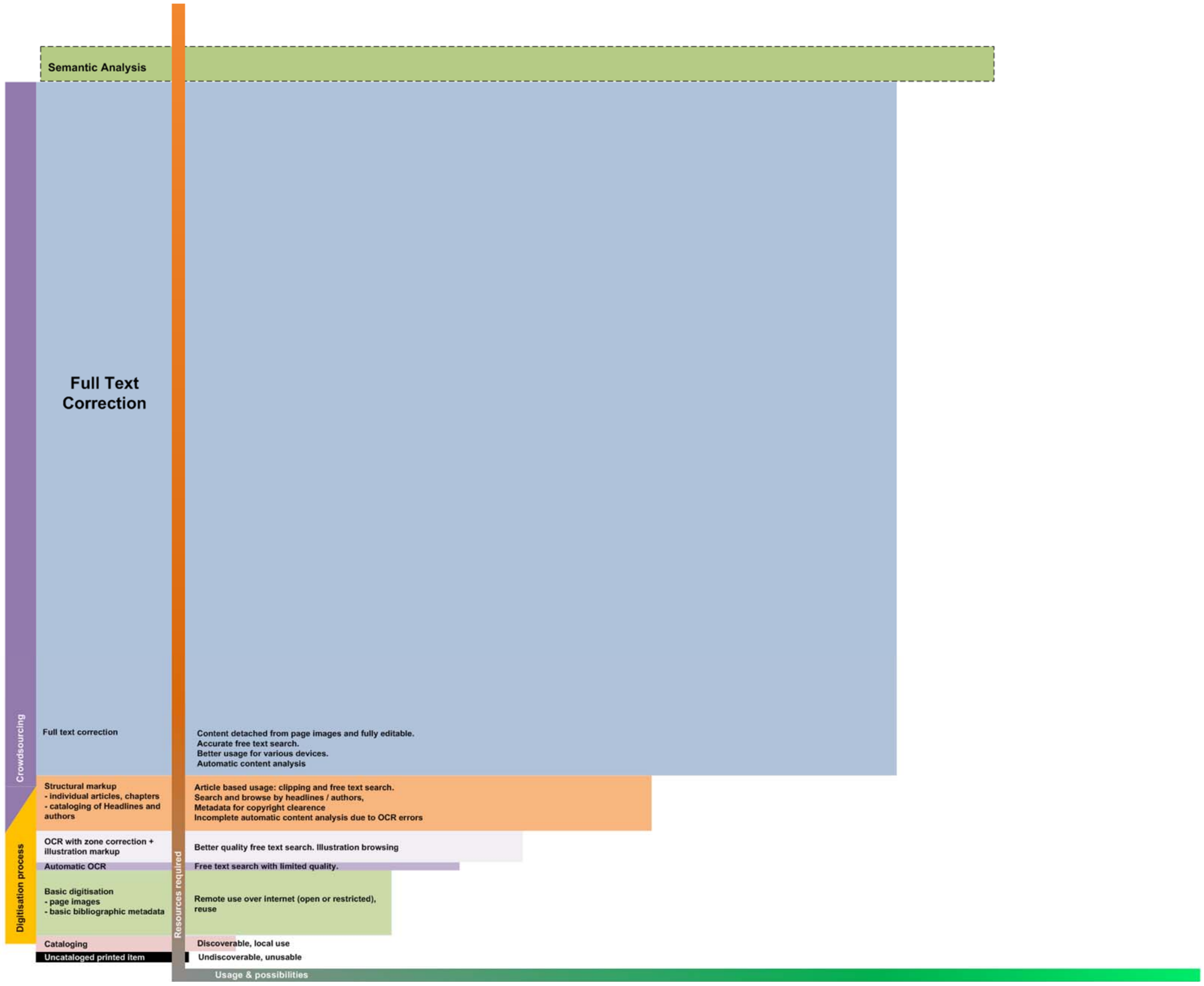
Cataloging

Discoverable, local use

Uncataloged printed item

Undiscoverable, unusable

Usage & possibilities



Semantic Analysis

Full Text Correction

Crowdsourcing

Digitisation process

Resources required

Full text correction

Content detached from page images and fully editable.
Accurate free text search.
Better usage for various devices.
Automatic content analysis

Structural markup
- individual articles, chapters
- cataloging of Headlines and authors

Article based usage: clipping and free text search.
Search and browse by headlines / authors,
Metadata for copyright clearence
Incomplete automatic content analysis due to OCR errors

OCR with zone correction + illustration markup

Better quality free text search. Illustration browsing

Automatic OCR

Free text search with limited quality.

Basic digitisation
- page images
- basic bibliographic metadata

Remote use over internet (open or restricted), reuse

Cataloging

Discoverable, local use

Uncataloged printed item

Undiscoverable, unusable

Usage & possibilities

Professorship in Digital Content Research

- The Information is primarily to be sought in the net
- The importance of mathematical methods and electronic research content is growing
- It is important to develop and forsee new ways of use:
- Research in many fields: Language technology, IT, Faculty of Arts, Electronic Archiving

Professorship in Digital Content Research (2)

- to intensify the search possibilities of digitised content:
    - Data mining from large collections /automation
    - The influence of the digital recording, format, on the user experience
    - The impact of the content on search possibilities
- To improve content management
    - E-content and longterm preservation
    - Linked open data and other such enrichment possibilities
    - Content and context information created by automation
- New ways, crowdsourcing

Co-operation

- Helsinki University (HU)/ Faculty of Arts, Modern Language, Language Technologies
- Partners: HU/Mathematics natural science, IT- science,
- Partners: National Library/Centre for Preservation and Digitisation, University of applied science in Mikkeli, (HU) Mikkeli University Consortia et al
- 5-year period 2013-2017, after that probably permanently

BENEFITS OF DATA AND ENRICHMENT

- Comprehensive digital collections in the ownership of the Nation /NLF
  - Case: Newspaper Publishers and NLF
- Use and reuse of digital content, new users and purposes
  - Community enrichment: retrospectively improve the quality through distributed efforts.
  - Professorship to be established: Digital collections research: Helsinki University/Modern Languages/ language technology and NLF and other partners
  - Descriptive data created during digitisation is attached to the library catalogues improving searchability

CASE NEWSPAPER PUBLISHERS & NLF

- THE WIDEST POSSIBLE USE OF NEWSPAPERS
  - The NLF has had Legal Deposit Right since 1707, including all newspapers since the first one in 1771-
  - Users. the newpaper customers, researchers, students, library customers

- USER ANGLE:
  - Newspaper houses: sites with exclusive supply of newspaper content
  - The copyright society Kopiosto has the right to make extended collective licensing agreements (2012): newspapers, journals, journalists, photographs…
  - NLF: a wider use of the newspaper content: via the library sectors in Finland
  - Long term preservation: METS, National digital library (KDK)

VISIBILITY OF THE 20 CENTURY

- Newspapers interests:
  - Extended use, to benefit the 20th century golden age in electronic format
  - Sophisticated services via the newspaper´s own site
  - Publicity and a small income via the Historical Newspaper Library / National Digitial Library
  - Partnership in digitisation
  - Long term preservation
  - METS-ALTO, metadata development together with the NLF
  - Benefits of the crowdsourcing projects of the NLF

VISIBILITY OF THE 20 CENTURY (2)

- Library:
  - Extends the supply of newspapers via the Historical Newspaper Library
    - The older part of the 20th century via the NLF:s service to the Library sectors: ”the moving wall”-principle.
  - Electronic delivery of newspapers
  - Computer output microfilming
  - Rediness to develop and share: crowdsourcing, metadata- development and enrichment, Kuvatalkoot, professour-results
- Kopiosto:
  - Clears the rights from the rightholders
  - Distributes the proceeds to the rightholders
- Project planning:
- National newspapers.
  - Maaseudun tulevaisuus, HBL
- Regional newspapers ongoing:
  - Etelä-Suomen Sanomat, Länsi-Savo

Leverage from the EU 2007–2013

European Union
European Social Fund

Challenges

- Organisational culture
  - MARC21, cataloging rules, ”Sacred National Bibliography”
  - ”That’s how it has always been done”
- Copyrights and licensing
- Data protection and privacy laws
- Language tools for Finnish

Challenges

- Software infrastructure needed
  - Preservation, presentation and APIs needed.
  - National and international cooperation and open source projects
  - Developers, developers, developers
  - How to do agile software development in library?