



# Aineistojen käsittely ja jalostaminen

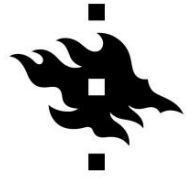
Tuula Pääkkönen

Tietojärjestelmäasiantuntija

Liikearkistopäivät 30.9.-1.10.2015, Mikkeli

<http://urn.fi/URN:NBN:fi-fe2015092714084>

# Kansalliskirjasto, Helsingin Yliopiston erillislaitos



HELSINGIN YLIOPISTO  
AJATTELUN VOIMAA  
1640–2015

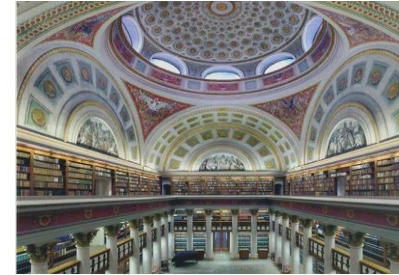
#KK375

Digitointi- ja  
Konservointi-  
keskus



2015

1990



1828

Keisarillinen Aleksanterin yo

1707

Vapaakappalelaki



1640

Turun akatemia



Akatemiatalo. Akademiäuset. The Academy. C.C. Gjørwell. 1815.



KANSALLISKIRJASTO – Digitointi- ja konservointikeskus

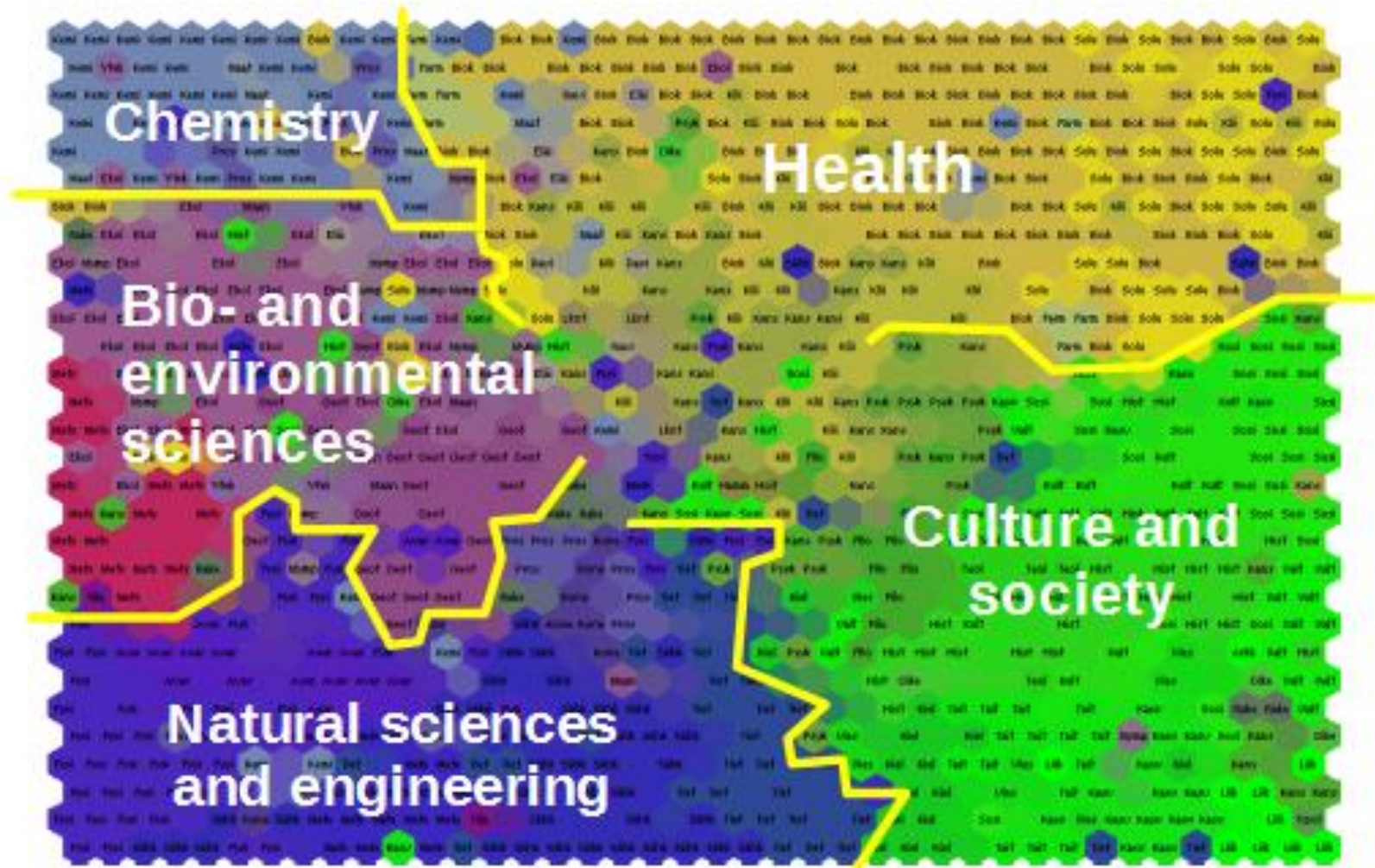
Lähde: <http://www.kansalliskirjasto.fi/yleistieto/kirjastotietoutta/historia.html>

# Sisältö

1. Aineistot
2. Käsittely
3. Jalostaminen

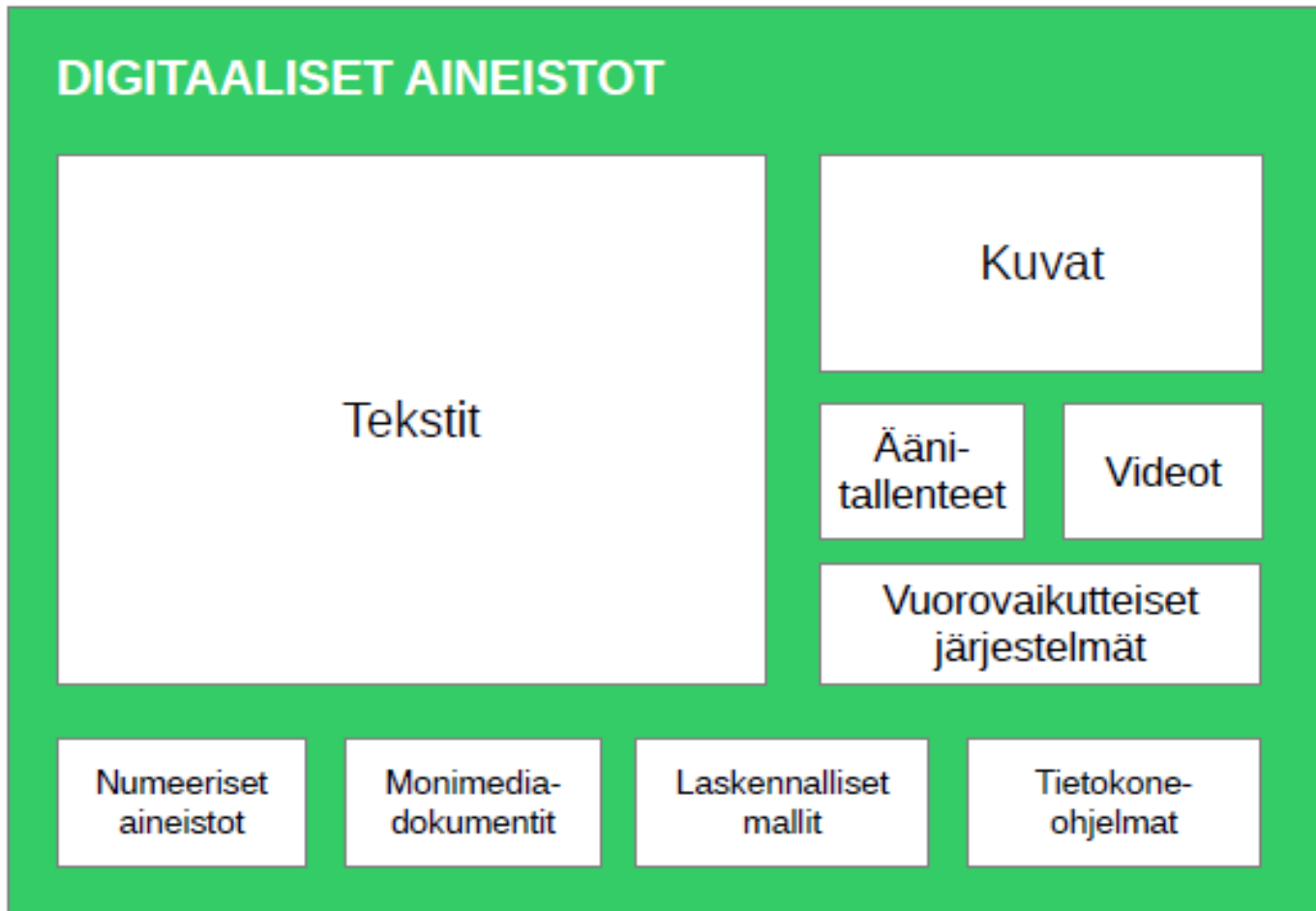


# Map of Finnish Science



(Honkela & Klami 2007)

# Moninaiset aineistot



# http://digi.kansalliskirjasto.fi



DIGI - Kansalliskirjaston digitoidut aineistot

BETA

ETUSIVU

SANOMALEHDET

AIKAKAUSLEHDET

PIENPAINATTEET

MUUT AINEISTOT

Palaute

Suomeksi  
På svenska  
In English

Kirjautu



DIGI.KANSALLISKIRJASTO.FI

OHJELU- JA TILASTOIMINTO

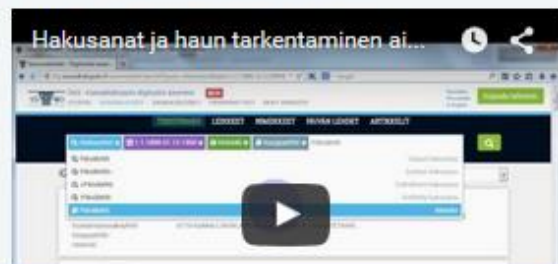
9 414 497 SIVUA

## LEIKETOIMINTO

Kuvien, mainosten ja artikkelien kerääminen digitoiduista aineistoista helpottuu. Sinulla on käytössä oma henkilökohtainen sivu, johon voit kerätä leikkeitä ja henkilökohtaisia kommentteja. Kaikki käyttäjät voivat hakea näitä leikkeitä "Leikkeet"-sivun kautta. Voit jakaa leikkeitä sosiaalisessa mediassa. Kirjautu sisään etusivulla ylhäällä oikealla. [Opastusvideo leikkeiden tekoon.](#)

## TEKSTIHAKU

Hakua voi rajata aikavälin, nimekkeiden sekä julkaisupaikkojen perusteella. [Lisäohjeet uuden haun käyttöön.](#)



## SANOMALEHDET

Digitoitu yhteensä 3 324 035 sivua.  
Vapaassa käytössä 1 960 921 sivua (58%) (-1910).  
Rajatussa käytössä 1 363 114 sivua (42%) (1911-).

Vapaa

Rajattu



## AIKAKAUSLEHDET

Digitoitu yhteensä 5 960 989 sivua.  
Vapaassa käytössä 1 133 671 sivua (19%) (-1910).  
Rajatussa käytössä 4 827 318 sivua (81%) (1911-).

Vapaa

Rajattu



## PIENPAINATTEET

Vapaassa käytössä 129 473 sivua.

Vapaa

# Digitaaliset aineistot

## SUOMI.

Suoin Suomen hyväksi. Maani marjan karvaksi.

**Mikkelissä, Keskiykökkö 28 p. Marraskuuta.**

**Tilauksilman avat:**

<b>Wuolte</b>	<b>Postin-</b>
<b>kuu</b>	<b>litti</b>
<b>kuu</b>	<b>kuu</b>
<b>kuu</b>	<b>kuu</b>
<b>kuu</b>	<b>kuu</b>
<b>kuu</b>	<b>kuu</b>
<b>kuu</b>	<b>kuu</b>
<b>kuu</b>	<b>kuu</b>
<b>kuu</b>	<b>kuu</b>
<b>kuu</b>	<b>kuu</b>
<b>kuu</b>	<b>kuu</b>
<b>kuu</b>	<b>kuu</b>

**Lehti ilmestyy:**

**Lehti ilmestyy:**

**Lehti ilmestyy:**

## THE FINLAND BULLETIN.

No. 20.                      DECEMBER 24TH.                      [1904.]

# TAITO

No 0.                      Toinen päivä: 1. PÄIVÄRINTA (päätoimittaja): K. LEHTONEN. — Toimittajapalkka ja konttori löyden kantajaja O. B. BLOMFIELD'In talossa.                      1888.

## MEETING OF THE DIET.

# arabia

### == KOULUEN VIRSKIRJA, ==

### THE SPEECH FROM THE THRONE.

vähintään 50 uudesta virskirjasta valittua vihren osaa koulun rekoushetki varten, tarjotaan täten tilattavaksi. Hinta 25 p. nollottua.

THE REPLIES OF THE ESTATES.

**A. Westerholm.**

suulke: kunnioitetavan yleisön huomion hyvin lajitellun

### THE HOUSE OF NOBLES AND THE LAND MARSHAL.

**Maanmiehesseuran** perustamisesta varten Mikkelin lähipitäjän kesken kutututavat kokoukseen Mikkelin kaupungin seurahuoneeseen ottaa tammikuun 19 p. Mikkelissä 27 p. Marrask. 1888.

### ENTHUSIASTIC WELCOME.

**Maanmiehesseuran** perustamisesta varten Mikkelin lähipitäjän kesken kutututavat kokoukseen Mikkelin kaupungin seurahuoneeseen ottaa tammikuun 19 p. Mikkelissä 27 p. Marrask. 1888.

### THE POSITION OF PARTIES.

**Maanmiehesseuran** perustamisesta varten Mikkelin lähipitäjän kesken kutututavat kokoukseen Mikkelin kaupungin seurahuoneeseen ottaa tammikuun 19 p. Mikkelissä 27 p. Marrask. 1888.

**Preliminary Arrangements.**

**Maanmiehesseuran** perustamisesta varten Mikkelin lähipitäjän kesken kutututavat kokoukseen Mikkelin kaupungin seurahuoneeseen ottaa tammikuun 19 p. Mikkelissä 27 p. Marrask. 1888.

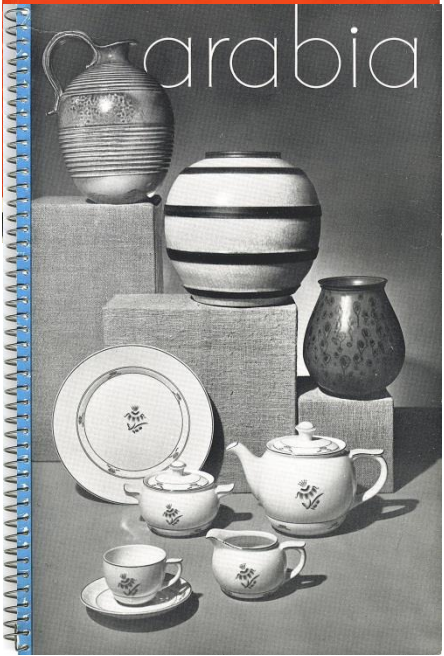
**Unpopular Appointments.**

**Maanmiehesseuran** perustamisesta varten Mikkelin lähipitäjän kesken kutututavat kokoukseen Mikkelin kaupungin seurahuoneeseen ottaa tammikuun 19 p. Mikkelissä 27 p. Marrask. 1888.

**The Nobles and the Land Marshal.**

**Maanmiehesseuran** perustamisesta varten Mikkelin lähipitäjän kesken kutututavat kokoukseen Mikkelin kaupungin seurahuoneeseen ottaa tammikuun 19 p. Mikkelissä 27 p. Marrask. 1888.

**One of the most important items in the many preliminaries**



**Maanmiehesseuran** perustamisesta varten Mikkelin lähipitäjän kesken kutututavat kokoukseen Mikkelin kaupungin seurahuoneeseen ottaa tammikuun 19 p. Mikkelissä 27 p. Marrask. 1888.

**Optimists had cherished hopes that Prince Obolensky, by making these resolutions, would pay due regard to the majority in the Houses, which, as we point out in another article, is decidedly "constitutional" in three of the four houses, whilst in the House of Clergy both parties are about**

**Maanmiehesseuran** perustamisesta varten Mikkelin lähipitäjän kesken kutututavat kokoukseen Mikkelin kaupungin seurahuoneeseen ottaa tammikuun 19 p. Mikkelissä 27 p. Marrask. 1888.

**equally represented.** The Governor-General, however, no doubt on the advice of the Senate, selected persons who were on good terms with that body, and indeed, some of their own members.

**Maanmiehesseuran** perustamisesta varten Mikkelin lähipitäjän kesken kutututavat kokoukseen Mikkelin kaupungin seurahuoneeseen ottaa tammikuun 19 p. Mikkelissä 27 p. Marrask. 1888.

**It would scarcely have been possible to appoint to the posts of Land Marshal and Vice-Land Marshal, as the Speaker and the Vice-Speaker are termed in the House of Nobles, two men more widely unpopular, and less to the liking of the House of Nobles, than Mr. Constantin Linder and Senator Oskari Vuorehimo.**

**Maanmiehesseuran** perustamisesta varten Mikkelin lähipitäjän kesken kutututavat kokoukseen Mikkelin kaupungin seurahuoneeseen ottaa tammikuun 19 p. Mikkelissä 27 p. Marrask. 1888.

**The former is one of the Vice-Presidents of the Senate, so which post he was appointed in 1901, notwithstanding his notorious ignorance, which can only be matched by his incapacity. It was hoped by those who appointed him that he, by reason of his large connections with the Russian aristocracy, and his sympathetic personal qualities, would be a willing tool in carrying through General Bobrikoff's programme.**

**Maanmiehesseuran** perustamisesta varten Mikkelin lähipitäjän kesken kutututavat kokoukseen Mikkelin kaupungin seurahuoneeseen ottaa tammikuun 19 p. Mikkelissä 27 p. Marrask. 1888.

**The expectations thus formed he has amply justified, and his appointment as Land Marshal has been accepted in Finland as a sign that the former policy would not be abandoned. Senator Vuorehimo, the Vice-Land Marshal, has a big share in the unpopularity**

**Maanmiehesseuran** perustamisesta varten Mikkelin lähipitäjän kesken kutututavat kokoukseen Mikkelin kaupungin seurahuoneeseen ottaa tammikuun 19 p. Mikkelissä 27 p. Marrask. 1888.

**6M, 20%**

**130k, 100%**

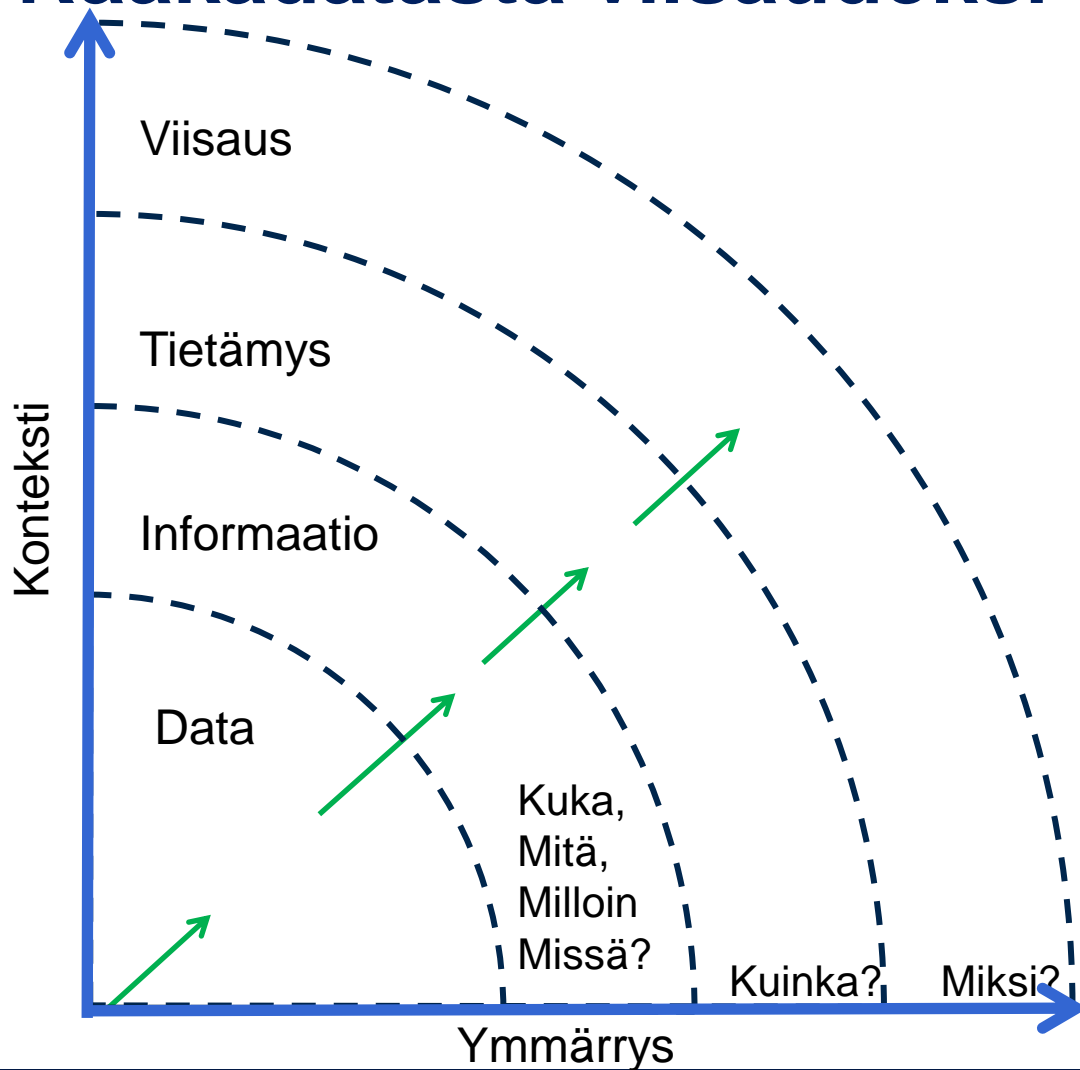
**3,2M sivua, 60% vapaassa verkkokäytössä**

# Käsittely





# Raakadatasta viisaudeksi



Käännetty alkuperäisestä [Lähde](#) (Forsythe, Implementing KM, Part I: Concepts & Approach)

konservointikeskus

# Moninaiset käyttäjäryhmät

Kansalaiset

Taiteilijat

Opettajat

Tutkijat

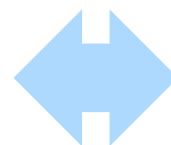
Toimittajat

Informaatikot

Päätöksentekijät



Digitaaliset aineistot



Museot

Arkistot

Kirjastot

Yliopistot

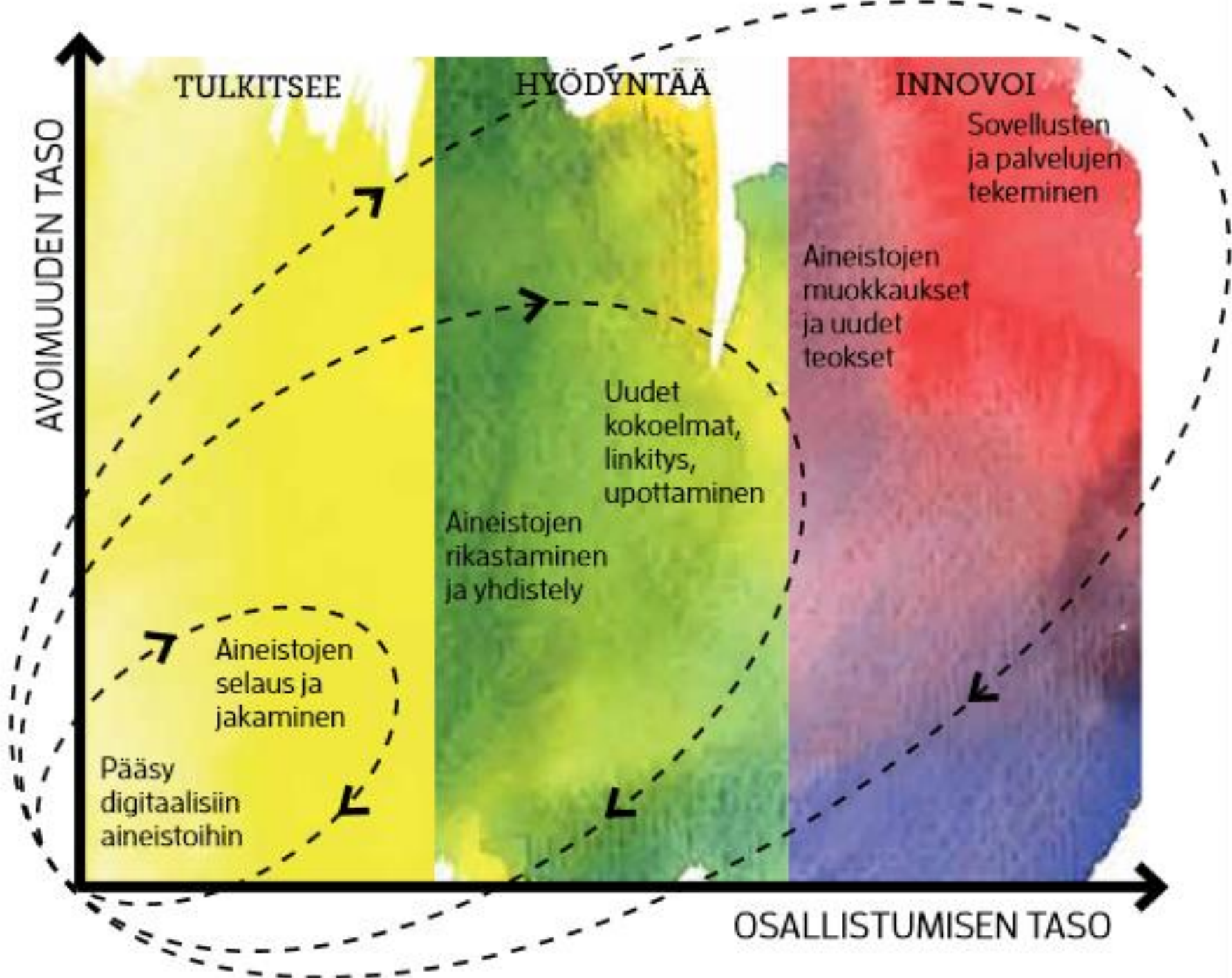
Yhdistykset

Media

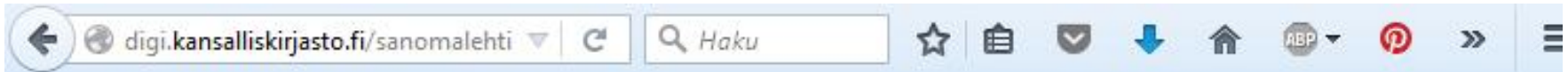
Yritykset

Kunnat

Valtio



# Joukkoistaminen



Putoamisen syynä oli se, että laivan kannella olevan aitaufsen portti oli jäänyt jotenkin auki. Wainaja oli nojautuessaan porttia vasten, jonka hän tietenkin kuuli olevan kiinni, horjasti arvaamatta portti-aukosta joteen. Vuolasta hän oli jäitähdyksestä saanut halvauksen, kostapa wainaja ei voinut laukaakaan aittaa veden päällä pysytellä, waikkakin omasi joukunlaisen uimataidon.

Muuten tahtoisimme tässä kaittien vastaisten jattumien waralta walawasti huomattaa, että poliisilaitoksen oli ehdottomasti hantitawata paremmat pelastusvehkeet. Myllään ei esim. ollut kuin yksi pahanpäiwäinen haka. Olemme mellein warmoja siitä, että jos wain oli ollut useampia weneitä ja hakoja, niin wainaja oli saatu paljon pikemmin maihin, ja näin ollen oli myöskin henkiinberättäminen toisinaan ollut vielä mahdollistakin.

Poliisilaitoksen on aiwan ensi tilassa täydenmittävä puutteellista pelastusvälinevarastoaan, ja myöskin nyt on jo tehtävä päätös, että jokainen poliisikonstaapeli pamaan ensi lesänä uimakouluun oppimaan uimataitoa ja hengenpelastuskiikkeitä.

**Warowammuin on hevosilla ajettava wilkkaimmilla liikepaikoilla.**

Silen illalla noin puoli 8 aikaan ajoi erään kauppiaan renki-poika laivaramassa niin hurjaa wauhtia, että aiwanpa oli lähemata kailli pyörönsä alle. Ommelji ei jattunut muuta wahinloa, kuin eräs wieni woika sai merkin nil-

**Mlioppilasfirjoituffissa**  
on hyväksytty Nurmeksen yhteiskoulusta Martta Sämäläinen, Elwi Härkönen ja Otto Kärkkäinen.

**Wanginfuljettajalta karammit wanki.**

Pielisjärwen piiriin nimismies J. S. Kuivalainen ilmotti kirjeesjään wiime elokuun 8 p:stä läänin kuvernöörille, että hän on wanginnut ja Kuopion lääninwanakilaan päästittanut erään Mi Heikki Makosen, mutta kun mainittua ei ollut lääninwanakilaan tullutkaan, niin tiedusteli kuvernöörinwirasto nimesmies Kuivalaiselta mihinkä wanki oli joutunut. Mystemmin on nimismies Kuivalainen ilmoittanut, että wanginfuljettaja oli päästännyt wangin karklaamaan ja senjälkeen salannut asiaa kolme wiikkoa hänelle siitä ilmoittamatta.

Nimismies Kuivalaista on läsfetty pitämään poliisitutkintoa asian johdosta.

**Maalaispoliisit.**

Rautawaaran nimismies poliisikonstaapeli Juho T. on liwuloisuuden talia saanut wirtawapautta wiiden ja puolen kuukauden ajaksi, lullen tl. 15 p:stä, jolla ajalla tointa on määrätty hoitamaan kihlakunnan lautamies J. Karppa.

**Murhapolstoista tiomittu.**

Sortawalan wälitäräjillä on irtolainen Aleks Simonpoika Riisinen kotoisin Utranchahalta tuomittu murhapolstoista 2 w. wankuteen ja sen jälkeen 3 wuodeksi wuottamään joulukuuhun.

hoidoissa:  
Stoulu wihittim tl. 10 pna.

**Ilkomaalaisia sanomalehtimie-**

**Leikkeen tiedot**

Otsikko \*

Maalaispoliisit Rautawaaralla

Tyyppi \*

Artikkeli tai uutinen

Aihealue

Laki, oikeus ja rikokset

Avainsanat

maalaispoliisi x poliiskonstaapeli x  
virkavapaus x kihlakunta x lautamies x  
nimismiespiiri x Rautawaara x  
Juho Timonen x Juho Karppa x

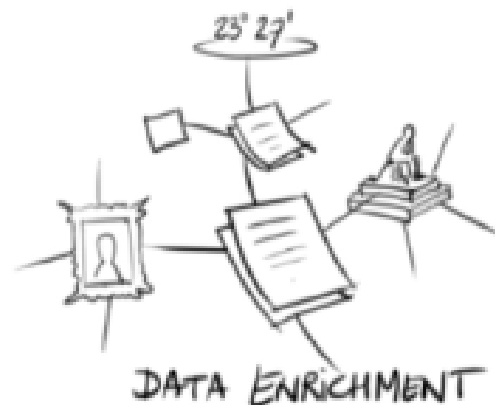
Peruuta Tallenna



Kuvaaja: Tuula Pääkkönen, 2015

# Miksi digitoida?

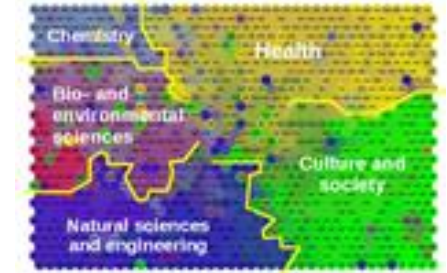
- Vaikuttavuus
- Monikanavaisuus
- Datan rikastaminen \*
- Brändi
- Uudenlaiset rahoitusmallit
- Löydettävyys \*
- Uudet asiakkaat
- Julkinen tehtävä \*
- Osaamisen kasvattaminen
- Onnekkait sattumat



Lähde: Benefits of open data (image by JAM/Europeana. CC BY via <http://mw2013.museumsandtheweb.com/paper/open-culture-data-opening-glam-data-bottom-up/>)

# Tekstinlouhinta

- “Sehän on tiedon hakua algoritmilla”



Kyllä, ja...

- Keino löytää **merkityksiä** suurista tekstimassoista automaattisesti
- Käytetään yleensä tilastollisia ohjelmia, jotka etsivät tekstimassoista **toistuvia hahmoja tai malleja**

Luokittelu

Tapahtumat, henkilöt, paikat -> Linkitetty data!

Sentimenttianalyysi

...

# Haku tekstimassasta/lähiluku

1640 DIGI - National Library's Digital Collections BETA  
FRONTPAGE NEWSPAPERS JOURNALS EPHEMERA OTHER DIGITAL COLLECTIONS

TEXT SEARCH CLIPPINGS TITLES PAPERS FOR DAY ARTICLE INDEX

Q London~ x 1.1.1816-31.12.1816 x Abo |

Q Abo

Q "Abo "

Tidningar Utgifne af et Sällskap i Abo

Uleåborgs Tidning

Uleåborgsbladet

Abo Allmänna Tidning

Abo Morgonblad

Abo Allmänna Tidning

Search term

Phrase

Publication

Publication

Publication

Publication

Publication

Publication

Publication

ABO ALLMÄNNA TIDNINGEN

...Den 5 Maji s- st. »ar EVNiftn: På *London* 5 Jr I Pence. Amsterdam 9 -s- Stäoer. Hamburz 8 i, 44 sch. Banc». Paris eoz Eentimes. Hollcindsta Dncatsr u Rubel...

...innehaft pil wenfira och högra RhensttMdej» kommer att tillhöra Bayern. *London* den 26 Aprilil. En Officerare wid Brittiska Flottan i Medelh^fwet berättar...

...ett härwarande handelshus. ? Hertigen af Wellington återwantas snart till *London*. ? Lord Exmouth har ftutat ett fördrag med Algier för England , utan att...

...Den 1 September g. si. war Coursen: «d *London* 103 i. tentimes. (M Tt. Petelsbmsst»...

...personer och de saker dem hon haft tillfäll bwlilka, dels fur sitt nöje iTu((...

“Tehokkuusero digitoitujen hakujen ja paperihakujen välillä on todella päätähuimaava!”

Kirjoita hakutermin, ja hae

- julkaisuaika
- Sanomalehden nimi
- fraasi
- Etc.



# Haaste: sivutekstimöykky

mt">6>? \* \* \*? " ». »\*.Tm?'^?mffmmraor-N:o 0^\* **HELSINGIN**  
SANOMATVastaava toimittaja: Paavo Warén.Kuudesti viikossa  
ilmestyvä aamulehti.**Helsingissä**, torstaina heinäkuun 7  
p.PäivällBt2 **Helsingin** postiimmistot pidetään»voin! a : r  
äänostikor.ttoriiläklevien postien toimisto: arkipäivinä klo II) a. p. ? « i,  
p., aimn- ja juhlapäiv. klo 9?II a. p.Tulevien postien toimisto:  
arki/iivinä klo 8 a. p. ? (? i. p., sunii -ja juulapäiv. klo o? II a.  
p.Postiosotusten ja po>tis:i;istöpankin Kmi.it ii, arkipäivinä klo 10 a. p. ?  
7p., sunn.- ja juhlapäivinä klo il ? 11a. p.Frankkomorkkcjä myydään  
arkipäivinä klo 7 a. p. ? 10 i. p., suun.- ja juhla{.uv. klu 7 a. p. 11 a. p ja  
?>,?><?> -7,30 i, p.POSTiuV MGtBulevardi... iivinä klo  
10a. p, ? 3 i, p. 4 G i. p., tunnun,- jai...  
p.Vuorimiehenkatu 1, arkip...  
jajuhlapäiv. klo ? 10 a. p.Kaivoka...  
lälietysten sekä vakuuttainaUomi...  
arkipäivinä klo 7,3(J 'J,3H a. p.2-3...  
juhlapäiv. klo 7.30? i),30 a. p.Frankkom...  
arkipäivinä klo 7,;)0 a. p.? 8.10 i. p.sunn.- ja juhlapäivinä klo I,'M) 'J,30

Mikä tämä on?  
Mistä lehdestä?  
Mikä on lehden rakenne?

# Sanojen analysointia

Kiitos!

<https://github.com/flammie/omorfi>

ja [WORD\_ID=ja][POS=PARTICLE][SUBCAT=CONJUNCTION][SUBCAT=COORDINATING]

Vwerst +?

Lumantin +?

armosa +?

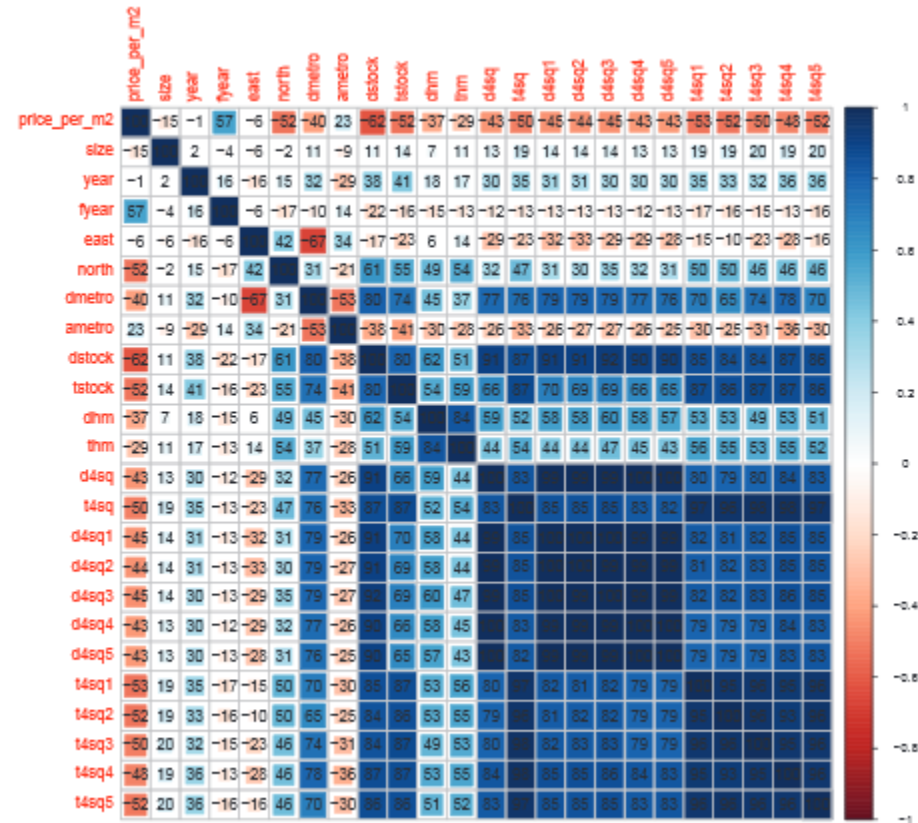
**ollen**

**[WORD\_ID=oilla][POS=VERB][VOICE=ACTIVE][INFINITIVE=E][NUMBER=SINGULAR][CASE=INSTRUCTIVE]**

Herran

[WORD\_ID=Herra][POS=NOUN][SUBCAT=PROPER][NUMBER=SINGULAR][CASE=GENITIVE]

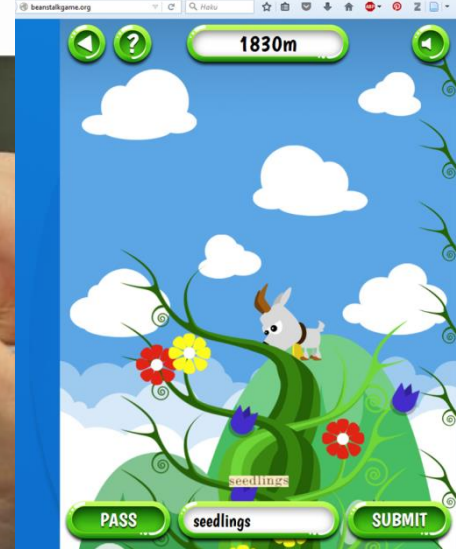
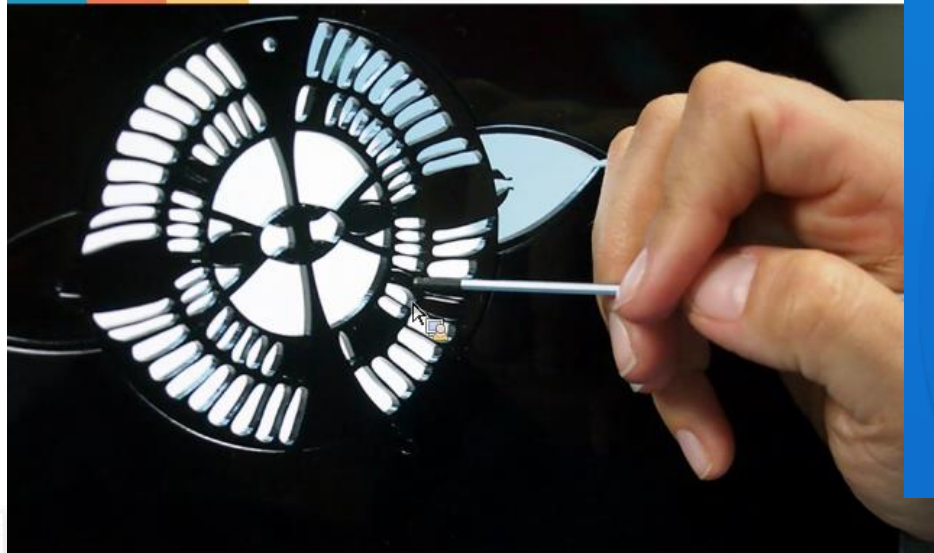
# Jalostaminen tuotteita & sovelluksia



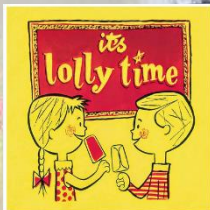
Lähde: Accessibility by public transport predicts residential real estate prices

Fig 5. <http://ceur-ws.org/Vol-1392/paper-09.pdf>

# Tuotteita



**HACK YOUR HERITAGE!**  
#hack4fi



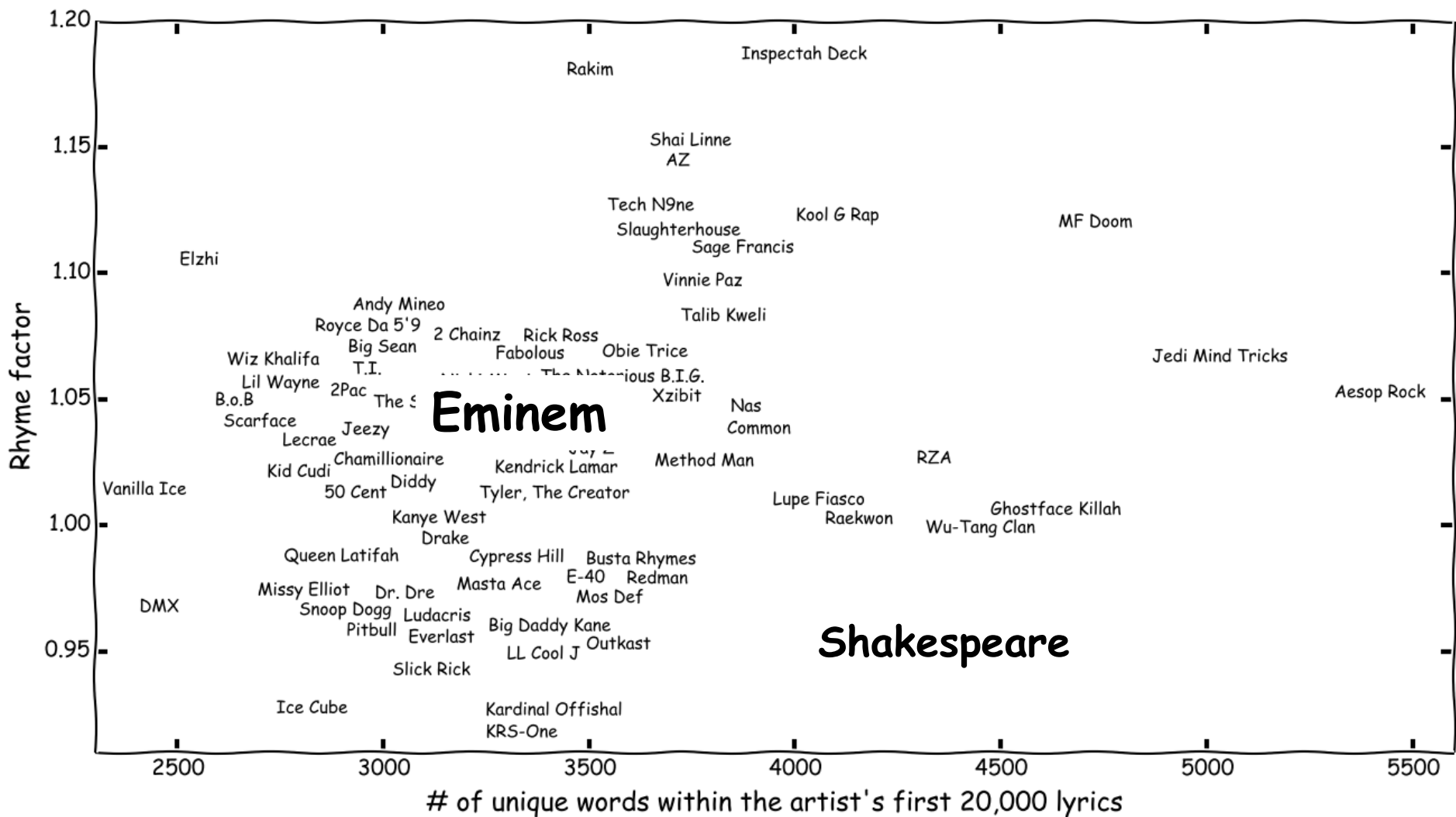
Myyräjahdissa metsästetään väärin kirjoitettuja sanoja. Kuva: Microtask.



Source: <https://pbs.twimg.com/media/CPf3S1FWIAADz6t.jpg?large>

Source:  
<http://melissaterras.blogspot.fi/2014/10/reuse-of-digitised-content-2-heres-one.html>

# Sanoitusten analyysiä

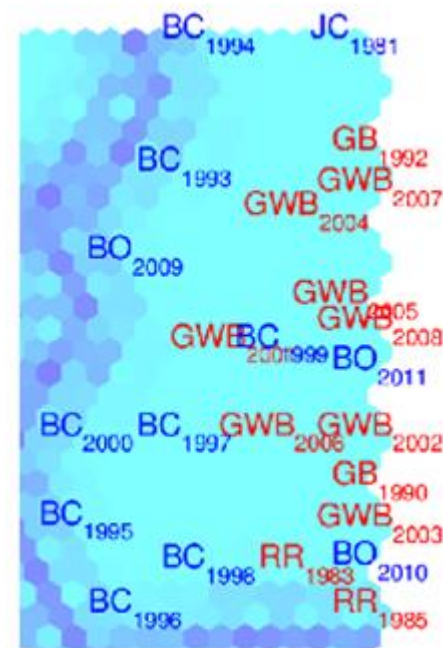


# Fraktuuramerkit itseorganisoivalla kartalla

À a	U		N		V	P		G		B
Ä ä	Ü							H		D
ß b			M							
Ɔ c	Q		W				Y			S
Đ đ										
Ɛ e					X					À
Ƒ f										Ä
Ɔ g	O						R			
Ɔ h	Ö									
Ɔ i									Z	-
Ɔ j										
…	C		T	L		K	F			I
	E									J

# Datasta eteenpäin

- **Talletettua, digitoitua** aineistoa voi jalostaa eteenpäin
- Käsittelyn ja jalostamisen vaatimukset
  - Aineisto
  - Digitointi/digitaalisuus
    - Standardit (esim. METS/ALTO)
  - Analyysi (yksityiskohdat/massa)
    - Lähiluku & tiedonlouhinta



**Subjects on objects in contexts:  
Using GICA method to quantify  
epistemological subjectivity.**  
Timo Honkela, Juha Raitio, Krista Lagus,  
Ilari T. Nieminen, Nina Honkela, and Mika Pantzar.  
*Proc. of IJCNN 2012.*

# Tulevia tapahtumia / EU koodiviikko

- Kansalliskirjasto osallistuu EU:n Koodiviikko-tapahtumaan viikonloppuna: 10.-11.10.2015
- Paikka : MAMK/Xinno, Mikkeli

