



AVOIN TIEDE  
JA TUTKIMUS

# Tutkimusdatan pitkäaikaissäilytys ATT-hankkeessa

[esa-pekka.keskitalo@helsinki.fi](mailto:esa-pekka.keskitalo@helsinki.fi)



# Turun yliopiston ainutlaatuisia aineistoa päätyi roskalavalle sateeseen – osa tuhoutunut

Tuhojen laajuutta ryhdytään selvittämään tänään. Arkistokokoelmat sisältävät uskomusperinnettä ja muistitietoa useilta vuosikymmeniltä, eikä kaikkea ole digitoitu. Saamelaisperinteen kokoelma on maailman laajin.

Marttiina Sairanen

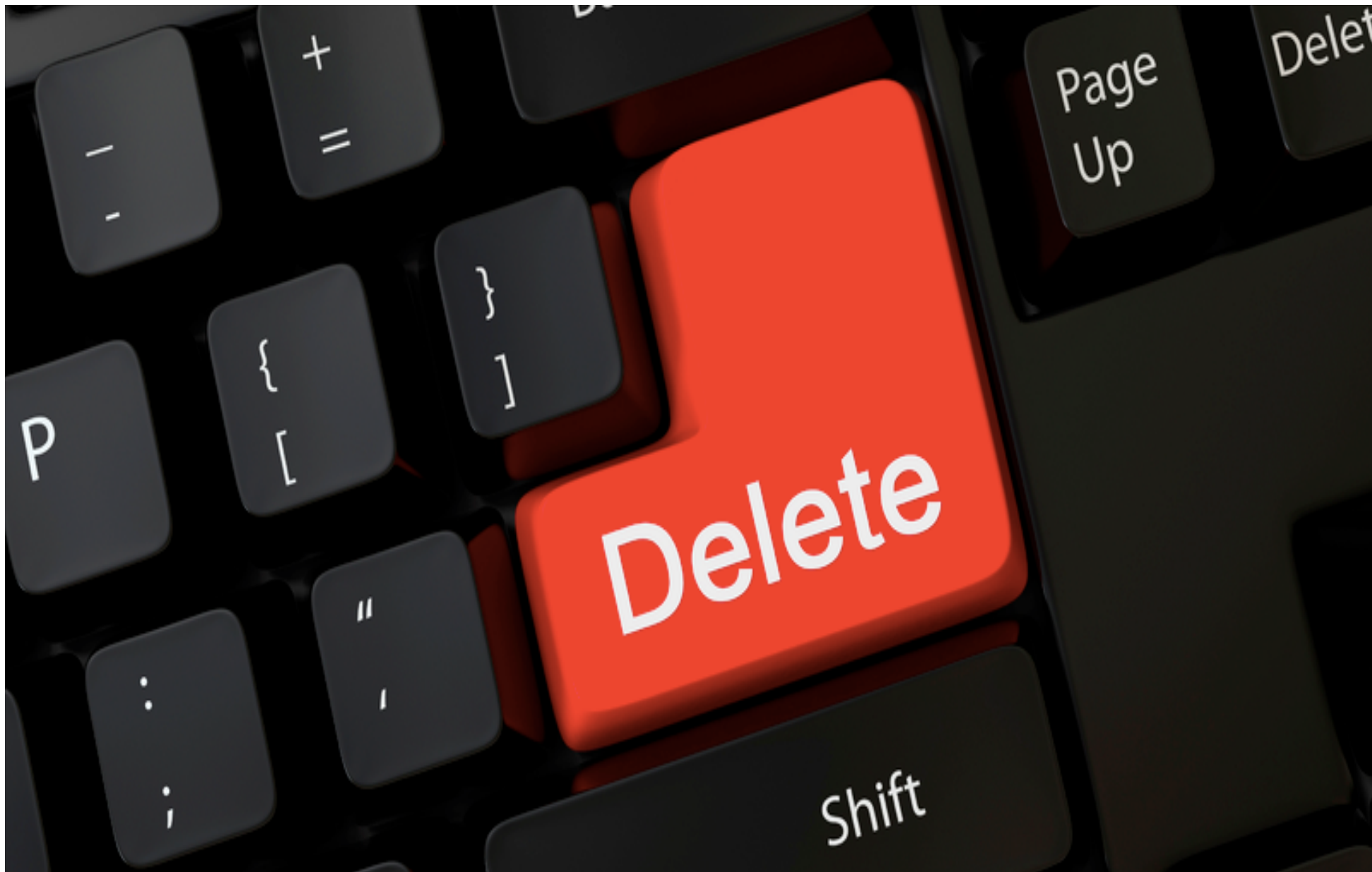


TS 22.3.2017



AVOIN TIEDE  
JA TUTKIMUS





AVOIN TIEDE  
JA TUTKIMUS



## **Datan saatavuus kuuluu tieteen perusolemuksen**

- Toistettavuus
- Todennettavuus
- Uudelleenkäyttö yllättävilläkin tavoilla
- → Sijoituksen hyödyntäminen täysimääräisesti → Rahoittajien vaatimukset



# Studying Climate Change With Old Photographs

Old family albums reveal close-ups of climate change.

By Anne Casselman | Monday, November 06, 2006

RELATED TAGS: [CLIMATE CHANGE](#)



Biologists Richard Primack and Abraham Miller-Rushing looked in an odd place



## Pitkäaikaissäilytys pähkinänkuoressa

- Ei heitetä pois.
- Pidetään järjestyksessä.
- Pidetään käyttökelpoisina.
- Annetaan käyttöön.





# Article Structure

## The Dynamics of Intraurban Quantile House Price Indexes

N. Edward Coulson and

*[Paper first received, March 2006; in final form, November 2006]*

abstract/summary

**Summary.** Estimating price indexes for different quantiles shows how prices vary for homes that are in different stages of the filtering process. The paper describes a filtering model and its implication for the dynamic interaction of housing prices at these various stages. A time-series analysis of quantile price indexes for three municipalities near Chicago supports the predictions of the filtering model.

### 1. Introduction

Despite an enormous literature on price index construction and a host of studies comparing appreciation rates across urban areas, few studies have estimated house price indexes for different locations within an urban area.

... appreciation rates in areas of intense metropolitan appreciation rates across urban areas, few studies have estimated house price indexes for different locations within an urban area. Prominent exceptions include Meese and Wallace (1991) and Case and Mayer (1996), who estimate house price indexes for the San Francisco Bay and Boston areas respectively. Other studies, such as Poterba (1991), Smith and Tesarek (1991), Mayer (1993) and Case and Shiller (1994) find that appreciation rates vary depending on a home's place in the sales price distribution. Attempts to explain differences in house price appreciation within urban areas are less common. The prominent exception is the study by Case and Mayer (1996), who argue that house appreciation rates depend on a home's spatially fixed amenities and local characteristics that vary across submarkets within an urban area. They find that house prices in single-family urban areas appreciate less rapidly in low-income municipalities that have a high percentage of manufacturing employment or middle-aged residents in 1980, and prices grew more rapidly in town-closer to Boston. Similarly, McMillen (2003) found that house prices within the City of Chicago rose more rapidly during the 1990s in locations closer to the city centre and in census tracts that in 1990 had high proportions of residents who were African American and had completed college. McMillen (2003) also found that appreciation rates were related to the characteristics of the housing stock: appreciation rates were higher in census tracts with a high proportion of vacant housing in 1990, with a housing stock of more recent construction, and with a higher proportion of

author affiliations

N. Edward Coulson is in the Department of Economics, Iowa State University, University Park, Pennsylvania, PA 16802, USA. E-mail: ecoulson@iastate.edu. David P. McMillen is in the Department of Economics (MC 146), University of Illinois at Chicago, 606 S. Morgan Street, Chicago, IL 60607, USA. Fax: 312 996 3148. E-mail: mcmlen@uic.edu

dynamic interaction. We find that the theory has fairly strong implications for the causal relationships between these time-varying variables. Section 6 concludes.

specialized vocabulary

### 2. Quantile Hedonic Price Indexes

Examples of the hedonic approach to price index construction include Kiel and Zabel (1997), Mark and Goldberg (1984), Palmquist (1980) and Thiobodeau (1989). The hedonic approach is typified by the following equation

$$y_{it} = \alpha + \beta'x_{it} + \delta_2 D_{2,t} + \dots + \delta_T D_{T,t} + u_{it} \quad (1)$$

where  $y_{it}$  is the natural logarithm of the price of home  $i$  at time  $t$ ;  $x_{it}$  is a vector of housing characteristics such as square footage and work; ref 'an' example refer to short run view the impulse responses in Oak Park, where the 90th percentile shock has a larger impact on lower quantiles than on 1 quantile housing.

### References

ALBRECHT, J., BOERKLIND, A. and VEGMAR, S. (2003) Is there a glass ceiling in Sweden? *Journal of Labor Economics*, 21, pp. 145–177.  
 BASSETT, G. W. JR. and CHEN, H.-L. (2001) Portfolio style, return-based attribution using quantile regression. *Empirical Economics*, 26, pp. 293–305.  
 BUCHENSKY, M. (1994) Changes in the US wage structure 1963–1987: application of quantile regression. *Economia*, 62, pp. 405–458.  
 BUCHENSKY, M. (1998a) The dynamics of changes in the female wage distribution in the USA: a quantile regression approach. *Journal of Applied Econometrics*, 13, pp. 1–30.  
 BUCHENSKY, M. (1998b) Recent advances in quantile regression models: a practical guideline for empirical research. *Journal of Human Resources*, 33, pp. 88–126.  
 BUCHENSKY, M. (2001) Quantile regression with sample selection: estimating women's return to education in the US. *Empirical Economics*, 26, pp. 87–113.

plifies to (1) if  $z_{it}$  always equals zero. We can rewrite equation (2) as

$$y_{it} = \alpha + \beta'x_{it} + \delta_2 D_{2,t} + \dots + \delta_T D_{T,t} + (z_{it} + \lambda_{2t} z_{it} D_{2,t} + \dots + \lambda_{Tt} z_{it} D_{T,t} + u_{it}) \quad (3)$$

The new variables measure changes in  $z$  between time  $t$  and the base period. For example, since  $D_{2,t} = 1$  while  $D_{3,t} \dots D_{T,t} = 0$  if the home sold during the second period, we have  $z_{it} = z_t + \lambda_{2t} z_{it}$  while  $z_{it} \dots z_{it} = 0$ . Thus,  $z_{it}$  in equation (1) is the same as  $z_t + \lambda_{2t} z_{it}$  in equation (2). The bracketed terms in equation (3) are the error terms when  $z$  is unobserved. The missing variables are correlated with the time variables, which leads to biased estimates of the price index.

As with any mean-based procedure, the ordinary regression model is sensitive to outliers. An obvious example in the case of house price models is depreciation, which can differ across households with the same set of  $X$ 's, and is likely to produce a relatively

*Reserve Bank of New York Policy Review*, September, pp. 63–77.

bibliography

Analysis.  
 C. (2001) ... during ... quantile ... index, 33, pp. 1027–1037.  
 JOHANSEN, S. (1995) *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. New York: Cambridge University Press.  
 KIEL, K. A. and ZABEL, J. E. (1997) Evaluating the usefulness of the American housing survey for creating housing price indexes. *Journal of Real Estate Finance and Economics*, 14, pp. 189–202.  
 KOENIGER, R. (2005) *Quantile Regression*. New York: Cambridge University Press.  
 KOENIGER, R. and BASSETT, G. W. JR. (1978) Regression quantiles. *Econometrica*, 46, pp. 33–50.  
 KOENIGER, R. and HALLOCK, K. F. (2001) Quantile regression. *Journal of Economic Perspectives*, 15, pp. 143–154.  
 LEVIN, J. (2001) For whom the reductions count: a quantile regression analysis of class size and peer effects on scholastic achievement. *Empirical Economics*, 26, pp. 221–246.



## Pidä järjestyksessä

- Omistus ja vastuut selvinä.
- Mielekkäät datakokonaisuudet.
- Riittävät kuvailutiedot.
- Hallinnan prosessit.

## **Pdä käyttökelpoisena**

- Bittitaso: torju bit rot.
- Laitteet: valvo ja vaihda.
- Formaattit: muunna tai emuloi.
- Ohjelmistot: varmista saatavuus.
- Käyttöoikeudet: selkeät ja ajantasaiset.
- Dokumentaatio: riittävä ja ajantasainen.

## Tutkimusdatan erityishaasteet

- Säilyttäminen uusi taakka.
- Eksoottiset tiedostomuodot, tieteenalojen erityiskäytännöt.
- Aineist
- Erittäin suuret data-aineistot hankalia.



## Kansallinen säilytysratkaisu

- Keskittämisen hyödyt osoitettu
  - paljoushyödyt
  - jatkuvuuden varmistaminen
  - riittävän asiantuntijayhteisön muodostuminen.
- Otetaan käyttöön 2018.

## Hyödyntää KDK-hankkeessa tehtyä työtä

- Kansallinen digitaalinen kirjasto –hanke
- Kulttuuriperinnön pitkäaikaissäilytys
- Samoja tallennustapoja, ohjelmistoja ja prosesseja voidaan soveltaa dataan.
- Kuitenkin eri palvelut erilaisille asiakaskunnille.

## Tutkimusaineiston matka pitkäaikaissäilytykseen

- Määrittele tutkimusaineisto.
- Tunnista säilytystarve ja ehdota säilyttämistä.
- Valmistele tutkimusaineisto paketiksi.
- Lähetä hyväksyttäväksi.
- Tee mahdolliset korjaukset.
- Saa tieto säilytykseen otosta.

## Tutkimusaineisto – mielekäs kokonaisuus

- Dataa
- dokumentaatio
- ohjeita, apuvälineitä ym.
- liittyvät julkaisut
- kuvailutietoa
- pysyvät tunnisteet.



## Pitkäaikaissäilytykseen valinta

- OKM antaa säilytyslupia aineistokokonaisuuksille.
- Sopimus säilyttämisestä tehdään *organisaation* kanssa.
- Sopimusprosessia tarkennetaan kesän aikana.

## METS DOKUMENTTI

### METS- OTSAKE

### KUVAILEVA METATIETO

*Standardisalkussa  
hyväksytyt kuvailevan  
metatiedon standardit*

### HALLINNOLLINEN METATIETO

Alkuperätieto

Tekninen metatieto

Käyttöoikeudet

Lähdetiedot

Säilytys suunnitelma

### RAKENTEELLINEN METATIETO

Looginen  
näkökulma

Nimiösivu  
Luvut  
Kappaleet  
....

Fyysinen  
näkökulma

Sivut  
Sarakkeet  
Sivujen osat  
....

Eri osioihin kuuluvien  
objektien linkittäminen

### TIEDOSTO- VIITTEET

Joukko 1

esim. TIFF-tiedostoja  
(alkuperäiset)

Joukko 2

esim. JPEG-tiedostoja  
(esikatseluvat)

Joukko 3

...

## Kuvailutyökalu auttaa tutkimusdatan kuvailussa

- Toteutetaan Kansalliskirjastossa.
- Tuotantoon I/2018.
- Täydentää esim. eri tieteenalojen olemassa olevia työkaluja.

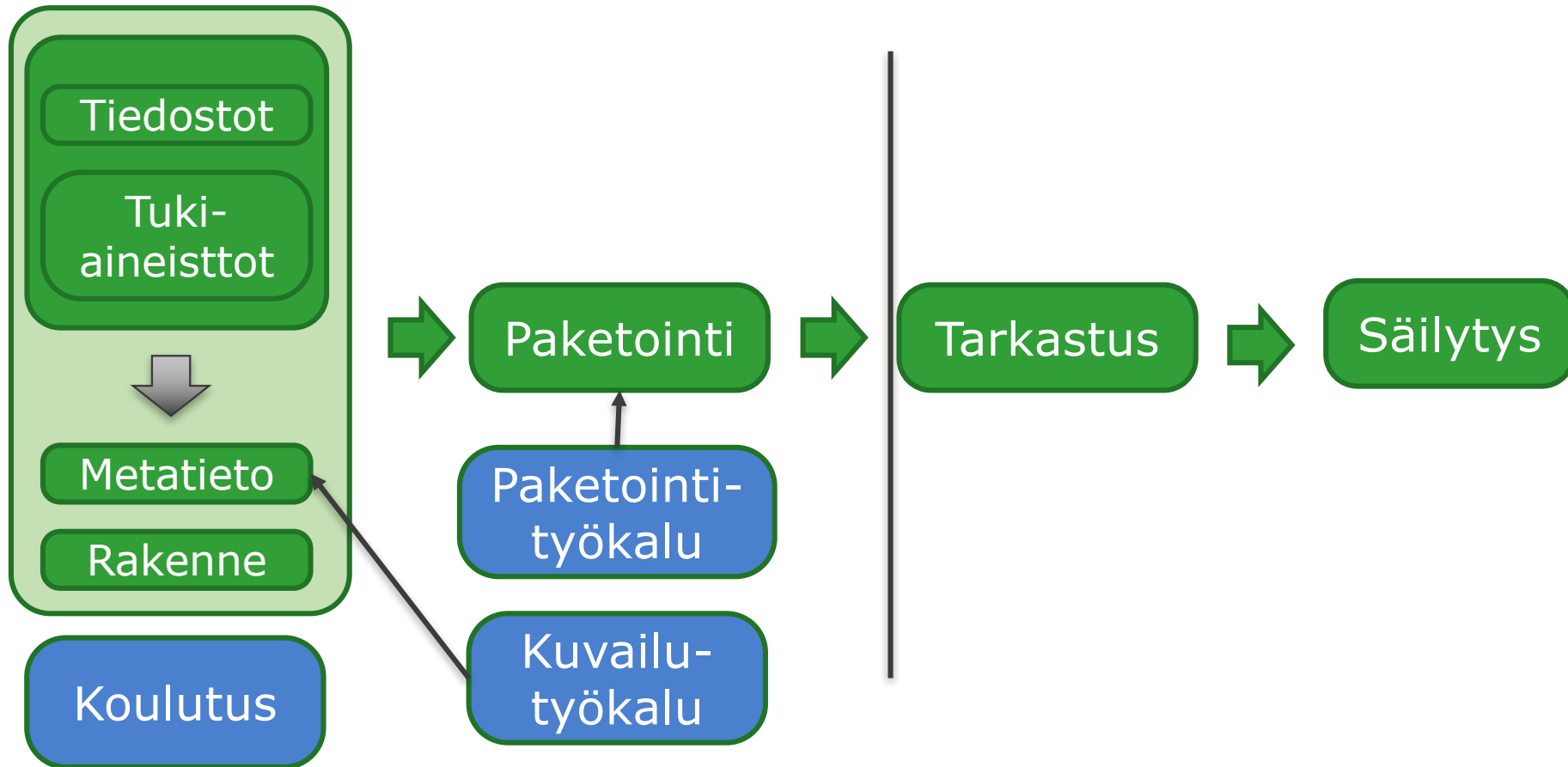
## Paketointityökalu

- Työkalu määrämuotoisten pakettien laatimiseen.
- Tuotannossa I/2018.



## Säilytykseen sopivat tiedostomuodot

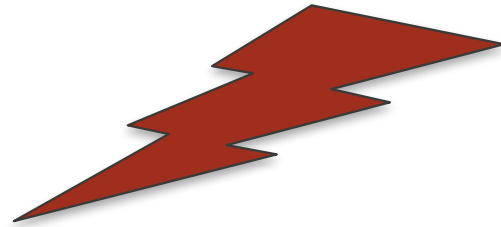
- Esim. PDF ja Docx eivät sinänsä ole säilytyskelpoisia.
- PDF/A ja Open Document Format taas ovat.
- Formaattien tulisi olla avoimia, standardeja, vakaita ja yhteensopivia, paljon käytettyjä.



## **PAS on osa palveluiden kokonaisuutta**

- Yhteys IDAan
- Yhteys Etsimeen

**Haluamme helpottaa tutkijan työtä.**



**Tutkijat tuntevat aineistot.**



**Tutkimusdatan hallinnan elinkaaren hallinta  
– tärkeää tutkimuksen alusta asti.**



AVOIN TIEDE  
JA TUTKIMUS

## Kansalliskirjaston julkaisuarkistopalvelut

## Julkaisuarkisto

- Organisaation tai tieteenalan tuottamien julkaisujen tallentaminen ja asettaminen avoimesti saataville.



## Päämääränä avoin tiede

- Tiedon avoin saatavuus
- Tiedon näkyvyys, esim. Googlen kautta
- Tiedon koneellinen siirtely avoimien rajapintojen kautta.
- Toisaalta organisaation tarve tuntea oma toimintansa.



AVOIN TIEDE  
JA TUTKIMUS



**HELDA**

HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITET  
UNIVERSITY OF HELSINKI



Jyväskylän yliopiston julkaisuarkisto

Jyväskylä University Digital Archive



Cornell University  
Library

arXiv.org

**HAL**  
archives-ouvertes.fr

## Kansalliskirjaston julkaisuarkistopalvelut

- Kansalliskirjasto tarjoaa alustan
- Asiakas käyttää alustaa ja vastaa sisällöstä.
- DSpace – avoimen lähdekoodin ohjelmisto
- Maksullista palvelutoimintaa, mutta kustannustehokasta!

## Doria.fi



- [ELY-keskus](#) [1647]
- [Kansalliskirjasto](#) [64294]
- [Lappeenrannan tekn. yliopisto](#) [10678]
- [Maanpuolustuskorkeakoulu](#) [2163]
- [Suomalaisen Kirjallisuuden Seura](#) [0]
- [Svenska litteratursällskapet i Finland](#) [46]
- [Turun yliopisto](#) [10330]
- [Åbo Akademi](#) [983]
- [Kokoelmakartta](#) [224]
- [Liikennevirasto](#) [3503]
- [Sibelius-Akatemia](#) [3045]
- [Stadia](#) [654]
- [TAMK](#) [1798]
- [Tilastokeskus](#) [1094]

**Theseus.fi**



- [www.theseus.fi](http://www.theseus.fi)
- Kaikki ammattikorkeakoulut
- Yhteiset toimintatavat

## Erilliset arkistot

- STM:n hallinnonala: [www.julkari.fi](http://www.julkari.fi)
- Luonnonvarakeskus: [www.jukuri.fi](http://www.jukuri.fi)
- Valtioneuvosto: <http://julkaisut.valtioneuvosto.fi/>
- Tampereen yliopisto: <http://tampub.uta.fi/>
- Lapin yliopisto: <http://lauda.ulapland.fi/>



## Lisäpalveluita

- Erilaiset tiedonsyöttöratkaisut
- Aineistojen massatuonnit/viennit
- Metatieto- ja aineistoputket julkaisuarkistoon ja siitä muualle
- Muita paikallisia räätälöintejä

## Kansallinen koordinaatio

- Yhteiset kuvailukäytännöt
- DSpace-kehittäjien yhteistyö
- Kansainvälinen yhteistyö

	4 / 2017	Total
<a href="#">Fluid flow in T-junction of pipes</a>	1372	<b>1372</b>
<a href="#">Lean Six Sigma in a Manufacturing Lead Time Improvement Project</a>	555	<b>555</b>
<a href="#">Heat transfer inside internal combustion engine: modelling and comparison with experimental data</a>	542	<b>542</b>
<a href="#">Design of axial-flux permanent-magnet low-speed machines and performance comparison between radial-flux and axial-flux machines</a>	482	<b>482</b>
<a href="#">Search Engine Optimization Methods &amp; Search Engine Indexing for CMS Applications</a>	459	<b>459</b>
<a href="#">Review of water electrolysis technologies and design of renewable hydrogen production systems</a>	434	<b>434</b>
<a href="#">Vasicek interest rate model</a>	414	<b>414</b>
<a href="#">Laser Cutting of Austenitic Stainless Steel with a High Quality Laser Beam</a>	391	<b>391</b>
<a href="#">Python 3 – ohjelmointiopas, versio 1.0</a>	379	<b>379</b>
<a href="#">Kraft recovery boilers – Principles and practice</a>	352	<b>352</b>



18.4.2017

# KANSALLISKIRJASTON AINEISTOJA UNESCON KANSALLISEEN MAAILMAN MUISTI- REKISTERIIN



## **KIITOS**

[esa-pekka.keskitalo@helsinki.fi](mailto:esa-pekka.keskitalo@helsinki.fi)

[www.avointiede.fi](http://www.avointiede.fi)

<https://www.kansalliskirjasto.fi/fi/palvelut/jarjestelmaalustapalvelut/julkaisuarkistopalvelu>