

Tutkimusaineistot osana tutkimuksen julkisuutta

Tieteen julkisuus -seminaari 12.4.2013

Antti Auer

Tietohallintopäällikkö, HT

Jyväskylän yliopisto

Strateginen kehittäminen

Tutkimusaineistojen julkisuus: mikä on näkökulma?

- Tutkimusaineistoja tulisi julkaista julkaisujen tapaan ja uudelleen käytettäväksi
 - Avoin data (rahoittajienkin vaatimukset) ja uudelleenkäyttö
 - Tieteen julkisuus ja popularisointi
 - Kaupalliset intressit
 - Tutkijoiden meritoituminen
- VAI
- Pitäisikö tutkimusaineistojen olla lähtökohtaisesti julkaistuna niin, että tutkimuksen tulokset ovat verifioitavissa ja menetelmät auditoitavissa?
- Tulevaisuudessa julkaiseminen ei välttämättä tarkoita vain julkaisua, vaan tulosten esittämistä tai toistamista suoraan aineistosta
 - Tulosten läpinäkyvyys: tulokset, menetelmät jolla ja aineisto josta tulokset on saatu, muodostavat kokonaisuuden
 - Tulosten julkaisemisen nopeus

Läpinäkyvyys (verifioitavuus, auditoitavuus)

- **“Research fraud exploded over the last decade” (2012)**

<http://arstechnica.com/science/2012/10/research-fraud-exploded-over-the-last-decade/>
<http://www.pnas.org/content/109/42/17028>

2,047 biomedical and life-science research articles indexed by PubMed as retracted . 67.4% of retractions were attributable to misconduct, including fraud or suspected fraud (43.4%), duplicate publication (14.2%), and plagiarism (9.8%).

- **“Research fraud needn't happen at all” (2002)**

<http://www.apa.org/monitor/feb02/fraud.aspx>

- Experts say it is the shared responsibility of institutions, funding agencies, journal editors and principal investigators to prevent research fraud.
- **Setting clear rules for data collection and analysis.** Principal investigators ought to lay out guidelines for the type of data that will and won't be considered in a study.
- **Reviewing data.** Another pre-emptive strategy is checking the raw data collected by one's co-investigators before writing it up and submitting it for publication. Principal investigators should do this and journal editors could check raw data as well. But that would likely be a tough sell, given the amount of time it would require.
- **Explaining the data analysis process.** In any submitted journal article, researchers ought to delineate how they categorized and analyzed their data.

Julkisuus aineistojen ja tulosten läpinäkyvyytenä sekä tutkijan meritoitumisena

IQSS Dataverse Network >

POWERED BY THE **Dataverse Network** PROJECT V. 3.3

American Journal of Political Science (AJPS) Dataverse 🔍 📄 🗨️ ⚙️ [Create Account](#) [Log In](#)

ARE BLACK POLITICIANS MORE INTRINSICALLY MOTIVATED TO ADVANCE BLACKS' INTERESTS? A FIELD EXPERIMENT MANIPULATING POLITICAL INCENTIVES [< View Previous Study Listing](#)

hdl:1902.1/19187UNF:5:JWYvKzriOA3bFgKx9/av7A==
Version: 1— Released: Tue Feb 12 15:26:27 EST 2013

CATALOGING INFORMATION Data & Analysis Comments (0) Versions

Data Citation

! If you use these data, please add the following citation to your scholarly references. [Why cite?](#)

Broockman, David, 2012-11-08, "Are Black Politicians More Intrinsically Motivated To Advance Blacks' Interests? A Field Experiment Manipulating Political Incentives", <http://hdl.handle.net/1902.1/19187UNF:5:JWYvKzriOA3bFgKx9/av7A==> V1 [Version]

[Download Subset](#) [Recode & Case-Subset](#) [Descriptive Statistics](#) **ADVANCED STATISTICAL ANALYSIS**

Selected Variables

Cross-Tabulation

! [More Information about the Model](#)

Classification

> leg_black
< leg_gender

Output Options

Include Totals
 Include Statistics
 Include Percentages

Results

Study Title: Are Black Politicians More Intrinsically Motivated To Advance Blacks' Interests? A Field Experiment Manipulating Political Incentives

VDCxtabs among leg_black: and leg_gender:

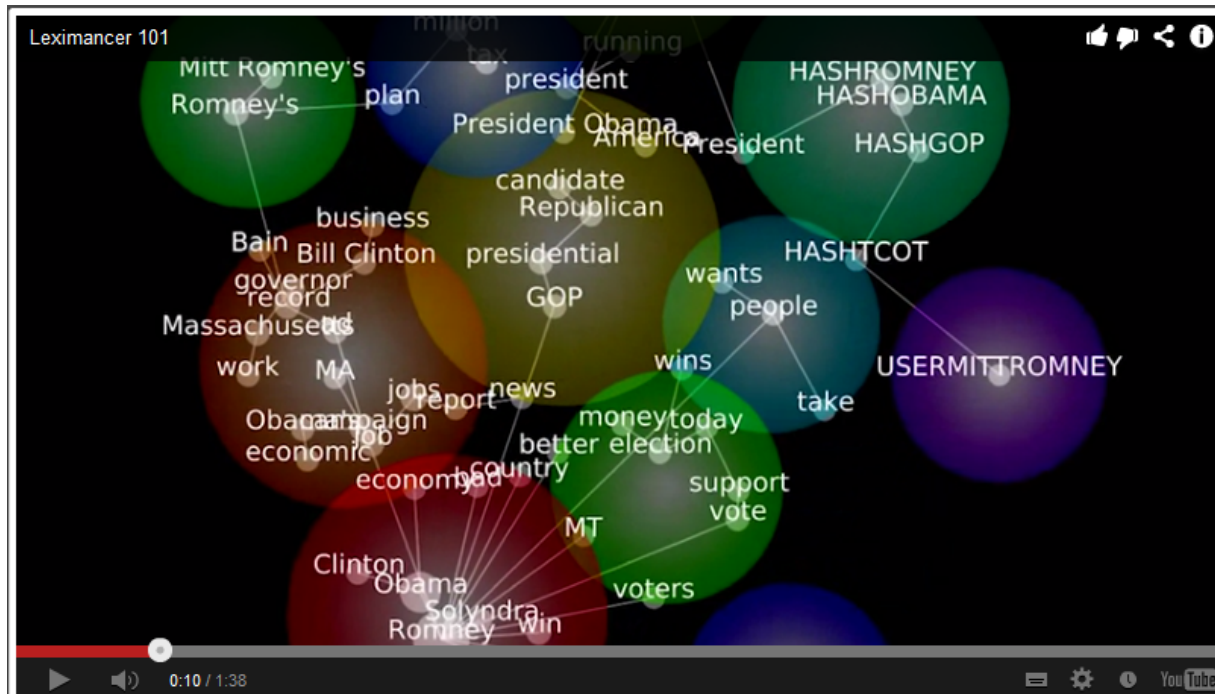
leg_black	leg_gender	"F"	"M"	Row/Total
0		1203 (21.5%) (23%) (90.7%)	4026 (72%) (77%) (94.4%)	5229(93.5%)
1		124 (2.2%) (34.1%) (9.3%)	240 (4.3%) (65.9%) (5.6%)	364(6.5%)
ColTotal		1327(23.7%)	4266(76.3%)	5593(100%)

VDCxtab Cell Legend: Frequency (total %, row %, column %)

Citation for subset analysis:

Broockman, David, 2012-11-08, "Are Black Politicians More Intrinsically Motivated To Advance Blacks' Interests? A Field Experiment Manipulating Political Incentives", <http://hdl.handle.net/1902.1/19187UNF:5:JWYvKzriOA3bFgKx9/av7A==> V1 [Version]leg_gender, leg_black [VarGrp/@var(DDI)]; UNF:3:+rRa1Ho/zaiy/2RHje4cSQ==

Entä laadulliset aineistot?



Leximancer enables you to navigate the complexity of text in a uniquely automated fashion. Our software identifies 'Concepts' within the text - not merely keywords but focused clusters of related, defining terms as conceptualised by the Author.

Leximancer's technology has now been applied to Enterprise Search in collaboration with Elsevier and its ScienceDirect article data store. Thanks to Needlz, searching for the right article in ScienceDirect is now much quicker and easier.

Julkisuus popularisointina

Ylen budjettilaskuri

http://yle.fi/uutiset/mista_sina_saastaisit_jos_olisit_ministeri_kokeile_laskurilla/6544401

Valitse säästökohteet alta

Tavoitteesi on säästää vähintään 500 milj. €

▶ NÄYTÄ KAIKKI

Työllisyys ← ALOITA TÄSTÄ	(0 / 6)
<input type="checkbox"/> Oppisopimusjärjestelmä (1) <small>Tukea vain niille, joilla ei ole ammatillista peruskoulutusta.</small>	80 milj. €
<input type="checkbox"/> Oppisopimusjärjestelmä (2) <small>Tukea vain alle 25-vuotiaille nuorille.</small>	100 milj. €
<input type="checkbox"/> Valmentava työvoimakoulutus <small>Vähennetään maahanmuuttajien koulutuksen tukea.</small>	14 milj. €
<input type="checkbox"/> Ammatillinen työvoimakoulutus <small>Vähennetään ammatillista työvoimakoulutusta. Työvoimakoulutus on suunniteltu ensi sijassa työttömille ja työttömyyden uhkaamille 20 vuotta täyttäneille.</small>	10 milj. €
<input type="checkbox"/> Aikuiskoulutustuki <small>Lasketaan tuen valtion rahoittamaa perusosaa opintotuen tasolle.</small>	10 milj. €
<input type="checkbox"/> Poistetaan vuorotteluvapaa	36 milj. €
Asuminen ja rakentaminen	(0 / 7)
Yritystuet	(0 / 5)
Koulutus, tutkimus ja kulttuuri	(0 / 10)

Mitä
näistä
seuraisi?

Viranomaisaineistot
Tutkimusaineistot
Julkaisut

Aineistojen hallinta julkisuuden edellytyksenä

- Aineistojen julkisuuden perusehtoja
 - Aineistojen omistus- ja käyttöoikeuksista on sovittu etukäteen organisaation, tutkijoiden ja tutkimuksen kohteiden kanssa
 - Aineistot on identifioitu yksikäsitteisesti (Persistent ID) niin, että niihin voidaan myös viitata
 - Aineistot on kuvattu riittävällä metatiedolla
 - Aineistot on ”puhdistettu” ja anonymisoitu avoimeen käyttöön
 - Hyväksytyt toimintapolitiikat ja -käytännöt, prosessit että tekniset järjestelmät tukevat julkisuutta ja julkisemista (Sic! julkaistu ei välttämättä ole avoin)

Hallintoa: sopimukset omistus- ja käyttöoikeuksista

(TTA-minimimetatietomalli)

1. Metatiedon tunniste 1
2. Metatiedon muokkaamisaika 1
3. Aineiston tunniste 1
4. Muokkaamisaika 1
5. Jakelijan yhteystieto 1
6. Aineiston nimi 1
7. Käyttöehdot 1
8. Aihe 1
9. Toimija 1 ("**omistaja**")
10. Aineiston kieli 0



Sopimusten yksikäsitteinen identifiointi!
Sopimushallinta järjestelmässä x

Tutkijoiden motivointi

- **Tutkijoita kiinnostaa tutkimuksen aikainen aineistojen hallinta – välitön hyöty tutkimushankkeelle**
 - Siltä osin aineistojen hallintaa pidetään tärkeänä
- Arkistointi, uudelleenkäyttö ja pitkäaikaissäilytys, popularisointi – ok, jos ei aiheuta lisätyötä
- Jos tutkimuksen aikainen aineistojen hallinta toimii, myös loppusijoitus hoituu – jälkikäteen hoidettuna hyvin hankalasti, jos ollenkaan
- Aineisto tulee normaalisti avoimeksi vasta sitten, kun se on hyödynnetty
 - Tärkein uusiokäyttökohde voi olla vertaileva tutkimus tai pitkittäistutkimus – tutkimuksen menetelmät ja käytännöt kuvattava tarkasti (so. ei vain aineisto, vaan miten se kerättiin, miten analyysit tehtiin etc.)

IKEA Job Interview



Oxfordin yliopiston tutkimusdatahanke

<http://www.ariadne.ac.uk/issue65/wilson-et-al>

Data management sharing plans	Legal & ethical	Best formats & best practice	Secure storage	Metadata	Access & discovery	Computation analysis & visualization	Restricted sharing	Data cleaning	Publication	Assess value	Preservation	Add value
Support												
Infrastructure and tools												
Policy												
Business model												

"Data repositories tend to think of it primarily in terms of preservation and curation, but researchers are more likely to associate the term with structuring data in a database, or the organisation of files and folder. At its broadest, research data management involves all the processes that information from research inputs undergoes as it is manipulated and analysed en route to becoming a research output."

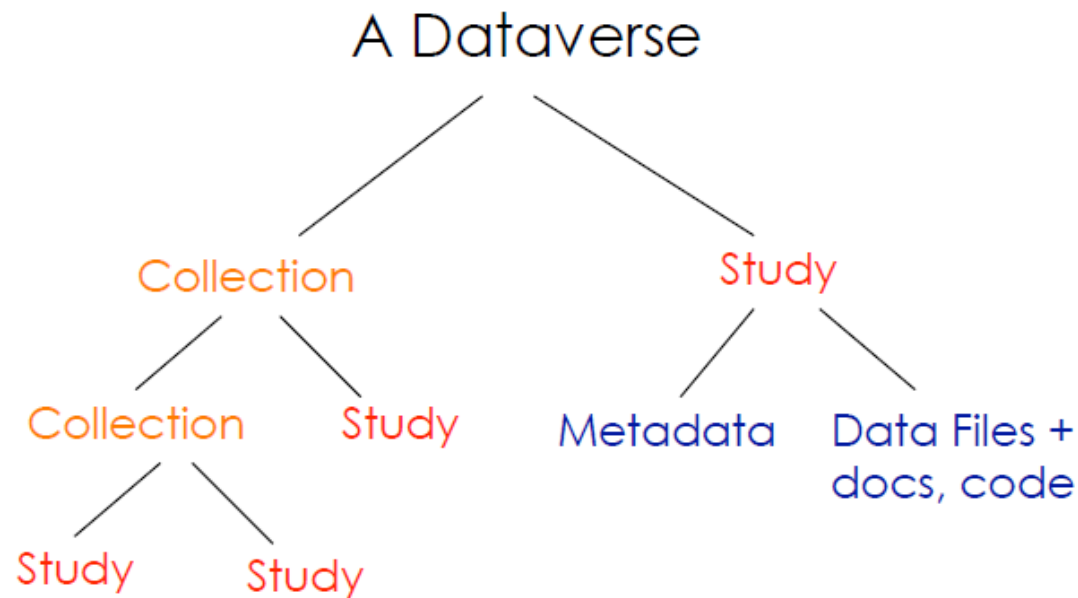
Hallintakerrokset ja -sovellukset

- Julkaisuarkisto/DSpace – digitaalisten objektien kokoelmat
- Youdata ja Yousource (IT-tdk)
- iRODS – prosessien automatisointi (mm. IDA-palvelussa)
- CKAN – avoin julkaiseminen (mm. <http://ckan.org/case-studies/helsinki/>)
- Dataverse – tilastolliset aineistot (thedata.org)
- Ala-/aineistospesifit hallintaohjelmistot
 - LIMS – laboratorio
 - CDMS – kliininen data
 - etc.
- Laadulliset aineistot
 - Esim. Atlas.ti: ”hermeneuttinen yksikkö” on kokonaisuus, joka sisältää sekä primääri- että sekundaaritekstit (koodaukset, memot etc.)
- Metatietomallit: TTA, DC, DDI, aineistospesifit....

Esim. Dataverse (thedata.org)

- It enables data archiving and preservation through re-formatting, standards and exchange protocols.
- It provides control and recognition for researchers through data management, branding and formal data citation.
- " The Dataverse Network Project standardizes the citation of data sets. Until this Project, citations of data were inconsistent or nonexistent in many publications, with future access and scholarly recognition highly uncertain. When you create a study in the Dataverse Network, the citation is generated and presented automatically."
- **Persistent URL (Handle) for permanent reference + Universal Numerical Fingerprint (UNF) for verification**
- **Viittauksen muotoesimerkki:**
Weisiger, Alex, "Replication data for: Logics of War: Explanations for Limited and Unlimited Conflicts",
<http://hdl.handle.net/1902.1/18738> UNF:5:OJCPMDOPJ96QO9V7fhXJMA== V1 [Version]
- **Tulevaisuudessa:** Seamless integration between the Open Journal Systems and The Dataverse Network.
- **Mallinnus ja visualisointi suoraan aineistosta – generoi viittauksen**

Viittaus seuraava rakennetta



Not only does this assure future researchers that they can use the same data set referenced in a years-old journal article, **it enables the data set's owner to track each iteration of the owner's research.** When an original data set is updated or incorporated into a new, related data set, the algorithm generates a unique UNF each time.

Työryhmistä toimenpiteiksi: aineistoinfrastruktuuri

- JY:ssa toiminut kaksi tutkimusaineistotyöryhmää (+ digitointiryhmä)
- Legal & Ethical: Sopimukset heti tutkimushankkeen alkuvaiheessa!
 - Työryhmässä laadittiin seitsemän mallisopimus pohjaa
- Secure storage: uudistetaan kevään aikana
 - CSC:n IDA-tallennuspalvelu jo käytössä
- Support: tietoasiantuntijan rekrytointi käynnissä
- Metadata: TTA-hankkeen minimimetatietomalli käytettävissä
 - Metatiedon tulisi tietää käyttöoikeudet ja aineiston omistaja sopimusten perusteella
- Policy: Tutkimusaineistopolitiikka/aineiston hallinnan periaatteet tekeillä
- **Tutkimusryhmille ja -hankkeille peruspalvelut (levypinta, virtuaalipalvelimet, ryhmätyövälineet) automaattisesti**
- **Hallintakerroksen rakentaminen ja käytäntöjen jalkauttaminen**
- TUTKA-ryhmässä pohditaan tutkimustietojärjestelmän (CRIS = Current Research Information System) uudistamista
 - CERIF-tietomalli - Tutkimushankkeiden ja niiden tulosten (julkaisut, patentit, infra, aineistot) välinen yhteys tärkeä osa julkisuutta

Data rules!

(even Saint Augustine knew that)



https://commons.wikimedia.org/wiki/File:Saint_Augustine_by_Philippe_de_Champaigne.jpg