

Kansalliskirjaston aineistot

HRI Loves Developers: Avoimen tekstidatan merkitys
digitalisaatiossa 23.5.2024

Juha Rautiainen, Kansalliskirjasto

Kulttuuriaineistolain mukaiset tehtävät

Kansalliskirjasto on Suomen suurin ja vanhin tieteellinen kirjasto sekä yksi Helsingin yliopiston suurimmista erillislaitoksista.

Vastaamme kulttuuriaineistojen tallettamisesta ja säilyttämisestä annetussa laissa (1433/2007) mainituista tehtävistä.

Kotimaisen julkaisu-
tuotannon
tallentaminen tutkijoiden ja
muiden tarvitsijoiden käyttöön

Yhteistyössä julkaisualan
toimijoiden kanssa

Kotimaisen julkaisu-
tuotannon
dokumentointi –
kansallisbibliografia



Kansalliskokoelman kartunnan perusta



- VAPAAKAPPALELAKI 1707
 - painettu aineisto
- ÄÄNITTEET MUKAAN 1980
 - musiikki ja puheäänitteet
- TEKIJÄNOIKEUSLAKI 2006
 - lupa digitoida kokoelmiin kuuluvia aineistoja ilman oikeudenhaltijan lupaa; lupa välittää digitaalisia aineistoja ns. vapaakappaletyöasemilla kulttuuriaineistolaissa mainituissa kirjastoissa
- KULTTUURIAINEISTOLAKI 2008 (Laki kulttuuriaineistojen tallettamisesta ja säilyttämisestä)
 - verkkoaineistot
 - Kansallinen audiovisuaalinen instituutti (KAVI) tv- ja radio lähetykset

Käyttöoikeudet



Kansalliskirjaston itsensä tuottama data ja metadata ovat vapaasti kaikkien saatavilla.

Käytettävissä on useita tietokantoja ja ohjelmistorajapintoja CC0-lisenssillä.



Digitoidut aineistot

Noin 30 prosenttia digi.kansalliskirjasto.fi:n aineistosta on avoimesti saatavilla, joko vapaasti tai sopimusperusteisesti.

Kaikki aineistot ovat käytettävissä vapaakappalekirjastoissa.

Osa rajoitetussa käytössä olevasta aineistosta on käytettävissä tutkimukseen sopimusperusteisesti korkeakouluissa ja yliopistoissa.



Verkkoarkisto ja sähköiset vapaakappaleet

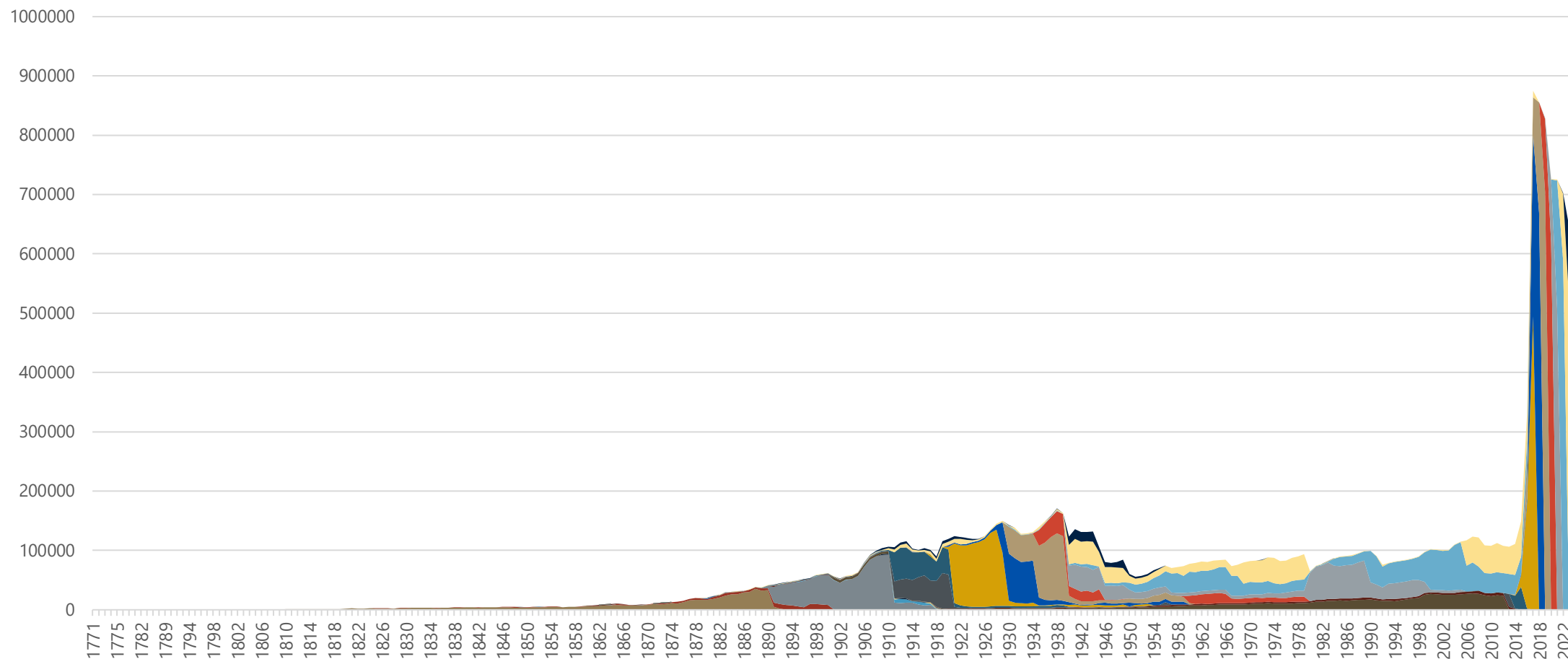
Nämä aineistot ovat käytettävissä vain vapaakappalekirjastoissa tarkoitukseen varatuilla työasemilla.

Digitoidut aineistot – digi.kansalliskirjasto.fi

Aineistolaji	Niteitä	Sivuja	Joista vapaasti verkossa saatavilla
Sanomalehdet	2 006 957	18 758 103	3 756 953 20 %
Aikakauslehdet	296 579	7 190 591	2 599 134 36 %
Pienpainatteet	11 868	146 801	146 801 100 %
Kirjat	24 399	2 142 337	1 578 370 73 %
Nuotit	44	1 322	550 41 %
Kartat	444	1 149	872 75 %
Kortistot	269	215 249	215 249 100 %
Käsikirjoitukset	23	5 954	3 493 58 %
Yhteensä	2 340 583	28 461 506	8 301 422 29 %

<https://digi.kansalliskirjasto.fi/stats> haettu 15.5.2024

Digitoitujen sanomalehtisivujen määrä julkaisuvuosittain (1771–2023)



digi.kansalliskirjasto.fi

Tekstisisältöön kohdistuva haku on käytössä kaikissa aineistotyypeissä. Hakua voi kohdentaa useilla metatiedoilla ja muokkaamalla hakulausetta.

Haku Kokoelmat Lehdet Päivän lehdet Leikkeet

Hae sivujen tekstisisällöstä

☒ vaadi kaikki hakusanat ☐ sumea haku hae: ☒ tekstisisällöstä ☐ kuvailutiedoista Kokeile uutta Nimiapuria!

Rajaa hakua: ☐ sanomalehdet ☐ aikakauslehdet ☐ pienpainatteet ☐ kirjat ☐ käsikirjoitukset ☐ kartat ☐ nuotit ☐ kuvat ☐ kortistot

kaikki nimekkeet kaikki kokoelmat

Aikaväli pp.kk.vvvv 01.01.2017

kaikki tekijät (kirjat) kaikki julkaisupaikat

kaikki julkaisijat kaikki avainsanat

kaikki kielet Sivut

Aineisto lisätty palveluun Kaikki ajat

☐ Sivu sisältää kuvituksia ☐ Hae kokonaisia niteitä ☐ Hae viimeiseltä sivulta ☐ Näytä hakutulokset, joihin ei ole luku-oikeutta Aiemmat hakuni

228724

« 1 2 3 4 5 6 7 »

Uusimmat ensin

ajatus 72

Ajatus : Suomen filosofisen yhdis...
nro 72, 2015
Helsinki, Suomen filosofinen yhdistys

ajatus 71

Ajatus : Suomen filosofisen yhdis...
nro 71, 2014
Helsinki, Suomen filosofinen yhdistys

ajatus 70

Ajatus : Suomen filosofisen yhdis...
nro 70, 2013
Helsinki, Suomen filosofinen yhdistys

ajatus 69

Ajatus : Suomen filosofisen yhdis...
nro 69, 2012
Helsinki, Suomen filosofinen yhdistys

Koneellinen tekstintunnistus kehittyi

*) Nykyiset kauppasäännöt, jot'a ovat kaupan pidättäneet etuus-oikeuksilla varustettu kouri n, myöntävät kruunun-palvelialle oikeuden, että milloin päälle sattuu, poisottamisen kaiken sen kalun, mitä ei ole sen myötävänä, jolla on sanottu etuus-oikeus, joskin se sama kama olisi ostettu samanlaatuiselta, mutta sattuis ostamaan enempi, kuin hänen nykyiset tarpeensa vaativat, jolloin liikkain poisotetaan. Tähän on näinä vuosina täällä nähty paljon esimerkkejä,

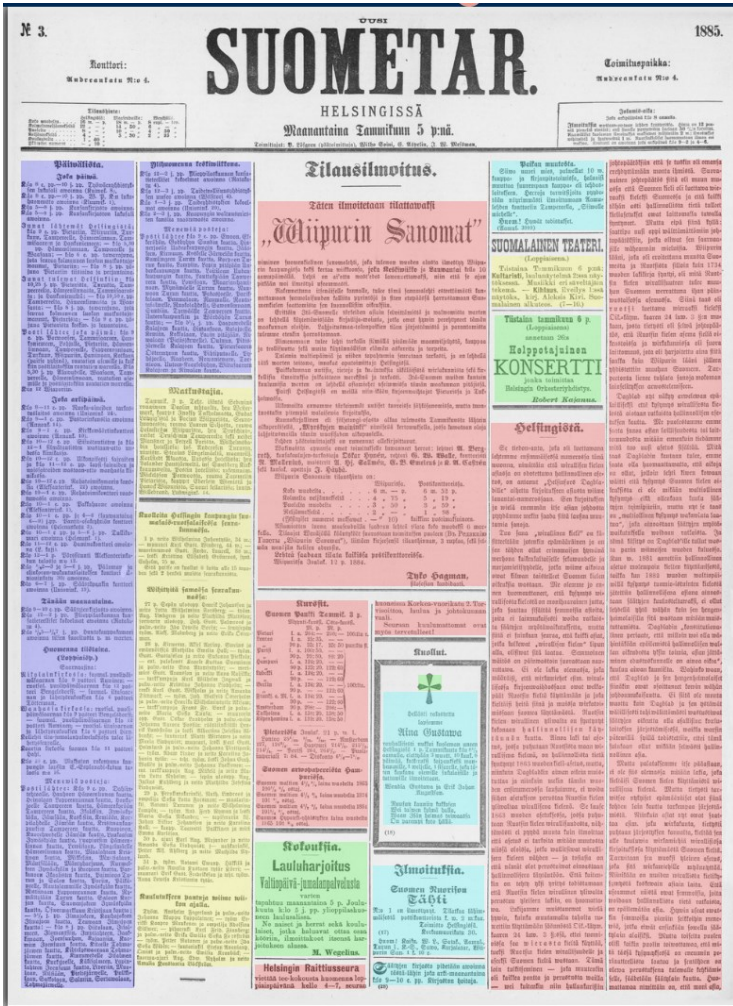
Pohjan-Tähti, 06.02.1867, nro 6, s. 1
<https://digi.kansalliskirjasto.fi/sanomalehti/binding/455330?page=1>
Kansalliskirjaston digitaaliset aineistot
Viitattu: 15.05.2024

*) 9?i;ft)jjet fau^afSHntiät, {ot*a omt knipan pi«
bhtmttt ctime«otftuff;fla ipantStettuinn touvin,
innoin tiimat frumuit^nhvcUafle etfeubcii/ että
miCoui päälle fattittt> „p'^",l'inns,u taiten sen
taliin, mifä ei rle sen itnjotKtönnä/ totta on
[anottu etnitB*oifeus, jceiin se sama tama elisi
ostettu fainanfaafuifel-tti/ idutta fntutis
cateinaan enempi, kuin hänen mifyijct
tatbeejifa »vaati.» wat/ jolloin Ui tatin
poii,'ctctaa. Tähän on näif^ vuosina täällä
nähty paljon esimerkkejä'.



*)Nykyiset kauppasäännöt, jot'a ovat kaupan
pidättäneet etuus-oikeuksilla varustettu
kouri n, myöntävät kruunun-palvelialle
oikeuden, että milloin päälle sattuu,
poisottamisen kaiken sen kalun, mitä ei ole sen
myötävänä, jolla on sanottu etuus-oikeus,
joskin se sama kama olisi ostettu
samanlaatuiselta, mutta sattuis ostamaan
enempi, kuin hänen nykyiset tarpeensa
vaativat, jolloin liikkain poisotetaan. Tähän on
näinä vuosina täällä nähty paljon esimerkkejä,

ALTO XML



<SP HEIGHT="51" WIDTH="22" VPOS="1342" HPOS="157"/>
<String ID="t_115_w2" HEIGHT="51" WIDTH="75" VPOS="1342" HPOS="179" CONTENT="5—8"/>
<SP HEIGHT="51" WIDTH="28" VPOS="1346" HPOS="254"/>
<String ID="t_115_w3" HEIGHT="51" WIDTH="23" VPOS="1346" HPOS="282" CONTENT="j."/>
<SP HEIGHT="51" WIDTH="17" VPOS="1349" HPOS="305"/>
<String ID="t_115_w4" HEIGHT="51" WIDTH="56" VPOS="1349" HPOS="322" CONTENT="pp."/>
<SP HEIGHT="51" WIDTH="29" VPOS="1350" HPOS="378"/>
<String ID="t_115_w5" HEIGHT="51" WIDTH="261" VPOS="1350" HPOS="407" CONTENT="Kansankirjasto"/>
<SP HEIGHT="51" WIDTH="34" VPOS="1350" HPOS="668"/>
<String ID="t_115_w6" HEIGHT="51" WIDTH="159" VPOS="1350" HPOS="702" CONTENT="awoinna."/>
<TextLine>
<TextLine ID="t_116" BASELINE="91,1416 163,1416 245,1413 263,1414 306,1423 376,1424 697,1426 858,1427" HEIGHT="46" WIDTH="769" VPOS="1382" HPOS="90"/>
<String ID="t_116_w1" HEIGHT="51" WIDTH="61" VPOS="1382" HPOS="92" CONTENT="K:lo"/>
<SP HEIGHT="51" WIDTH="22" VPOS="1380" HPOS="153"/>
<String ID="t_116_w2" HEIGHT="51" WIDTH="78" VPOS="1380" HPOS="175" CONTENT="5—9"/>
<SP HEIGHT="51" WIDTH="28" VPOS="1386" HPOS="253"/>
<String ID="t_116_w3" HEIGHT="51" WIDTH="23" VPOS="1386" HPOS="281" CONTENT="j."/>
<SP HEIGHT="51" WIDTH="17" VPOS="1389" HPOS="304"/>
<String ID="t_116_w4" HEIGHT="51" WIDTH="56" VPOS="1389" HPOS="321" CONTENT="pp."/>
<SP HEIGHT="51" WIDTH="27" VPOS="1391" HPOS="377"/>
<String ID="t_116_w5" HEIGHT="51" WIDTH="284" VPOS="1391" HPOS="404" CONTENT="Kansankirjaston"/>
<SP HEIGHT="51" WIDTH="40" VPOS="1392" HPOS="688"/>
<String ID="t_116_w6" HEIGHT="51" WIDTH="133" VPOS="1392" HPOS="728" CONTENT="lukusali"/>
<TextLine>
<TextLine ID="t_117" BASELINE="131,1457 291,1457" HEIGHT="30" WIDTH="162" VPOS="1428" HPOS="130"/>
<String ID="t_117_w1" HEIGHT="51" WIDTH="162" VPOS="1423" HPOS="130" CONTENT="awoinna."/>
<TextLine>
<TextLine ID="t_118" BASELINE="93,1505 246,1505 510,1511 856,1516" HEIGHT="51" WIDTH="765" VPOS="1466" HPOS="92"/>
<String ID="t_118_w1" HEIGHT="50" WIDTH="146" VPOS="1472" HPOS="96" CONTENT="Junat"/>
<SP HEIGHT="51" WIDTH="42" VPOS="1475" HPOS="242"/>
<String ID="t_118_w2" HEIGHT="51" WIDTH="227" VPOS="1475" HPOS="284" CONTENT="lähtewä"/>
<SP HEIGHT="51" WIDTH="36" VPOS="1480" HPOS="511"/>
<String ID="t_118_w3" HEIGHT="51" WIDTH="311" VPOS="1480" HPOS="547" CONTENT="Helsingistä:"/>
<TextLine>
<TextLine ID="t_119" BASELINE="133,1543 188,1543 228,1540 274,1543 344,1549 535,1549 751,1559 857,1549" HEIGHT="52" WIDTH="726" VPOS="1508" HPOS="132"/>
<String ID="t_119_w1" HEIGHT="51" WIDTH="49" VPOS="1509" HPOS="130" CONTENT="k:lo"/>
<SP HEIGHT="51" WIDTH="29" VPOS="1506" HPOS="179"/>
<String ID="t_119_w2" HEIGHT="51" WIDTH="15" VPOS="1506" HPOS="208" CONTENT="9"/>
<SP HEIGHT="51" WIDTH="22" VPOS="1507" HPOS="223"/>
<String ID="t_119_w3" HEIGHT="51" WIDTH="28" VPOS="1507" HPOS="245" CONTENT="e."/>
<SP HEIGHT="51" WIDTH="21" VPOS="1513" HPOS="273"/>
<String ID="t_119_w4" HEIGHT="51" WIDTH="51" VPOS="1513" HPOS="294" CONTENT="pp."/>
<SP HEIGHT="50" WIDTH="21" VPOS="1516" HPOS="345"/>
<String ID="t_119_w5" HEIGHT="50" WIDTH="171" VPOS="1516" HPOS="366" CONTENT="Pietariin."/>
<SP HEIGHT="50" WIDTH="21" VPOS="1521" HPOS="537"/>
<String ID="t_119_w6" HEIGHT="50" WIDTH="193" VPOS="1521" HPOS="558" CONTENT="Wiipurin."/>
<SP HEIGHT="51" WIDTH="22" VPOS="1519" HPOS="751"/>
<String ID="t_119_w7" HEIGHT="51" WIDTH="85" VPOS="1519" HPOS="773" CONTENT="Tur." SUBS_TYPE="HypPart1" SUBS_CONTENT="Turkuun."/>
<TextLine>



METS

-	VOLUME	Fredrik Pacius Kotimaana ...	1 -228
+	FRONT		1 -19
-	MAIN		20 -217
+	CONTRIBUTION	Pacius Helsingfor...	20 -29
+	CONTRIBUTION	Fredrik Pacius - s...	30 -97
+	CONTRIBUTION	Pacius och körs...	98 -119
+	CONTRIBUTION	Fredrik Pacius ...	120 -163
+	CONTRIBUTION	Musik, idrott o...	164 -179
-	CONTRIBUTION	Fredrik Pacius ...	180 -205
L	TITLE		180
L	AUTHOR		180
-	BODY		180 -205
+	BODY_CONTENT		181 -202
-	ILLUSTRATION	Korpral Möttöne...	180
L	IMAGE		180
L	CAPTION		180
+	ILLUSTRATION	Korpral Möttöne...	185
+	ILLUSTRATION	Ballad (regi, Toiv...	189
+	ILLUSTRATION	Kvinnobilder (re...	197
+	ILLUSTRATION	namnlös	202
+	ILLUSTRATION	namnlös	202
+	ILLUSTRATION	Ballaadi (ohj. Toi...	203
+	ILLUSTRATION	Melodramen To...	204
+	ILLUSTRATION	Vastaava kohta s...	205
+	CONTRIBUTION	Om Fredrik Pa...	206 -217
+	BACK		218 -228

```

<?xml version="1.0" encoding="UTF-8" ?>
<!-- METS -->
<!-- METS xsi:schemaLocation="http://www.loc.gov/METS/ http://digitalpreservation.fi/schemas/mets/mets.xsd" PROFILE="http://digitalpreservation.fi/mets-profiles/cultural-heritage" OBJID="URN:NBN:fi-fd2019-00022799" TYPE="PAS-
METS_v1-4_NLF-3" LABEL="Fredrik Pacius Kotimaana musiikki Musiken som hemland Pacius Fredrik Pacius" fi:SPECIFICATION="1.7.0" fi:CATALOG="1.7.0" fi:CONTRACTID="urn:uuid:ee012eec-1a1e-4132-8cb2-b5d7252cb16a">
  <!-- METS:agent ROLE="CREATOR" TYPE="OTHER" OTHERTYPE="SOFTWARE">
    <mets:agent ROLE="CREATOR" TYPE="OTHER" OTHERTYPE="SOFTWARE">
      <mets:name>CCS docWorks/METAE Version 7.3-0</mets:name>
    </mets:agent>
  <!-- METS:agent ROLE="CREATOR" TYPE="ORGANIZATION">
    <mets:agent ROLE="CREATOR" TYPE="ORGANIZATION">
      <mets:name>The National Library of Finland</mets:name>
    </mets:agent>
  </mets:metsHdr>
  <!-- METS:dmdSec ID="dmd-marc" GROUPID="1" CREATED="2023-08-31T13:42:30+03:00">
    <mets:dmdSec ID="dmd-marc" GROUPID="1" CREATED="2023-08-31T13:42:30+03:00">
      <mets:mdWrap MDTYPE="MARC" MDTYPEVERSION="marcxml=1.2; marc=marc21" LABEL="Bibliographic meta-data of the digital version">
        <mets:xmlData>
          <marc21:record>
            <marc21:leader>03852nam a22008414i 4500</marc21:leader>
            <marc21:controlfield tag="007">cr|||||||||||||</marc21:controlfield>
            <marc21:controlfield tag="008">100122r20192009fi |||eo 011 0dfin </marc21:controlfield>
            <!-- METS:subfield code="a" -->
            <marc21:subfield code="a">fd2019-00022799</marc21:subfield>
            <marc21:subfield code="2">skl</marc21:subfield>
            </marc21:subfield>
            <!-- METS:subfield code="a" -->
            <marc21:subfield code="a">URN:NBN:fi-fd2019-00022799</marc21:subfield>
            <marc21:subfield code="2">urn</marc21:subfield>
            </marc21:subfield>
            <!-- METS:subfield code="a" -->
            <marc21:subfield code="a">FI-NL</marc21:subfield>
            <marc21:subfield code="b">fin</marc21:subfield>
            <marc21:subfield code="e">rda</marc21:subfield>
            </marc21:subfield>
            <!-- METS:subfield code="a" -->
            <marc21:subfield code="a">fin</marc21:subfield>
            <marc21:subfield code="a">swe</marc21:subfield>
            </marc21:subfield>
            <!-- METS:subfield code="a" -->
            <marc21:subfield code="a">e-fi---</marc21:subfield>
            <marc21:subfield code="a">e-gw---</marc21:subfield>
            <marc21:subfield code="a">e-sw---</marc21:subfield>
            </marc21:subfield>
            <!-- METS:subfield code="a" -->
            <marc21:subfield code="a">ML410.P112</marc21:subfield>
            <marc21:subfield code="b">F</marc21:subfield>
            </marc21:subfield>
            <!-- METS:subfield code="a" -->
            <marc21:subfield code="a">52</marc21:subfield>
            <marc21:subfield code="2">kkaa</marc21:subfield>
          </marc21:record>
        </mets:xmlData>
      </mets:mdWrap>
    </mets:dmdSec>
  </mets:mets>

```

Lataustyökalu

Työkalu saatavissa:
<https://digi.kansalliskirjasto.fi/odata/submit>

Ohjevideo
lataustyökalun
käyttöön:
<https://www.youtube.com/watch?v=o-ykh9zTOdI>

Hae sivujen tekstisisällöstä

Lataustyökalu

1640 KANSALLISKIRJASTO NATIONALBIBLIOTEKET

EUROOPAN UNIONI European Union

Vipuvoimaa EU:lta 2014–2020

Digin hakutulokset sisältävän sivun osoite
<https://digi.kansalliskirjasto.fi/search?coll=...>

Tekstiaineiston formaatti
☒ ALTO xml ☐ teksti

Kuva-aineistot
☐ jpg kuvat ☐ pdf

Lataa kaikki niteen sivut
☐

Aineiston tallennussijainti
C:\Hyapp\Aineistolataaja

Hakemistopuun rakenne
Valitse sijainti BindingTitle, BindingID

Latauksen tila

Lataa aineisto Peruuta lataus

doista

Kokeile uutta Nimiapuria!

äsi kirjoitukset ☐ kartat ☐ nuotit ☐ kuvat ☐ kortistot

kokoelma: Taide. Urheilu.

kaikki julkaisupaikat

kaikki avainsanat

kaikki kielet

Sivut

ytä hakutulokset, joihin ei ole luku-oikeutta

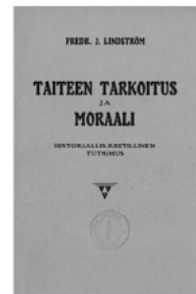
Aiemmat hakuni

19

« 1 »

☰ ☱ ☲ ☳ ☴ ☵ ☶ ☷

Vilmeiksi lisätyt



Esimerkkejä aineistojen käytöstä



Tieteellisessä tutkimuksessa, esimerkiksi
TurkuNLP, NewsEye

Hackatoneissa, mm. Helsinki Digital Humanities
Hackathon

Sukututkimuksessa, paikallishistorioissa

Somessa, kuten Wanhoja aikoja
(<https://www.tiktok.com/@ohonojopas>)

...

www.kansalliskirjasto.fi

Juha Rautiainen