

International Seminar:

Open Research Data - the FAIRest Data is the Future of Science

Tallinn University of Technology, 20 April 2017

Lithuanian SSH Data Archive: starting DA before FAIR

dr. Vaidas Morkevičius

Senior researcher

Institute of Public Policy and Administration, Kaunas UT

Lithuanian Data Archive for SSH www.lidata.eu

LiDA: how we started

- **2006-2008** project funded by EU Structural Funds, implemented by Kaunas University of Technology (**librarians and political scientists**) in collaboration with Vilnius University and Institute for Social Research (**quantitative sociologists**)
 - Big Data, Open Science, Open Data, FAIR principles etc. - only in the far horizon
 - Even then we started with an idea to become open, findable/reusable and interoperable DA

LiDA: how we started

- **2009-2011** another project funded by EU Structural Funds, implemented by Kaunas University of Technology in collaboration with Vilnius University, Institute of History (**historians**) and Vytautas Magnus University (**qualitative sociologist**)
 - The infrastructure that is currently available was fully developed and installed

LiDA: who we are and what we do (www.lidata.eu)

LiDA Lietuvos HSM duomenų archyvas

LiDA Data Training Research News Login

Guest

Logo: MOKSLAS • EKONOMIKA • SAUGLAUKIA
Kuriame Lietuvos ateitį

Logo: ŠVIETIMO IR MOKŠLO MINISTERIJA

Logo: KAUNO TECHNOLOGIJOS UNIVERSITETAS



Lietuvos HSM duomenų archyvas (LiDA) yra virtuali skaitmeninė empirinių HSM duomenų kaupimo, ilgalaikio saugojimo ir sklaidos infrastruktūra.

NEW DATA

- 2017-04-06
 - ESS7, atrankos dizaino duomenys, Lietuva, 2015 m. balandis – birželis, 2 leidimas
- 2017-02-17
 - Socialinė politika VI, 2015 m. spalio - gruodis, 2 leidimas
- 2017-02-06
 - ESS7, žiniasklaidos (naujiųjų) tyrimo duomenys Lietuva, 2015 m. balandis – birželis, 2 leidimas
- 2016-12-13
 - Lietuvos visuomenės istorinės samonės ir

DATA SEARCH

New ■ CESSDA welcomes a new member: LiDA **More ... New** ■ October, 2011 - The first edition of the data and documentation from ESS Round 5 **More ... New** ■ May 11, 2011 - LiDA started sea

About LiDA

LiDA provides virtual digital infrastructure where researchers can deposit, search, browse, make online analyses, and download

Training

LiDA serves as a tool to increase methodological competence of researchers by providing methodological assistance, and

Data

As the National Member Institution at the Inter-university Consortium for Political and Social Research (ICPSR) LiDA also provides access

Research

LiDA coordinates international data collection initiatives in Lithuania, e.g. European Social Survey (ESS), European Election Study (EES), and

LiDA: who we are and what we do

- LiDA provides **virtual digital infrastructure** for acquisition, preservation and dissemination of digital SSH data in Lithuania
 - SSH researchers can search, browse, make online analyzes and download data sets of more than 250 **surveys**
 - LiDA has modules for archiving data of the Lithuanian **political system**, **Qualitative studies** and **Historical statistics** of the Lithuania.

LiDA: who we are and what we do

- LiDA also serves as a hub to increase **methodological competence** of researchers by providing methodological assistance and training:
 - Distance learning solutions
 - Data confrontation seminars and methodological training
- LiDA also aims to become a national point of **access to international SSH data** stored in other archives (ICPSR, CESSDA)

LiDA: who we are and what we do

- Main **activities**:
 - SSH data acquisition, documentation and publication for free access to academic community
 - Proper documentation is the key
 - Methodological training
 - Data without knowledgeable users are useless

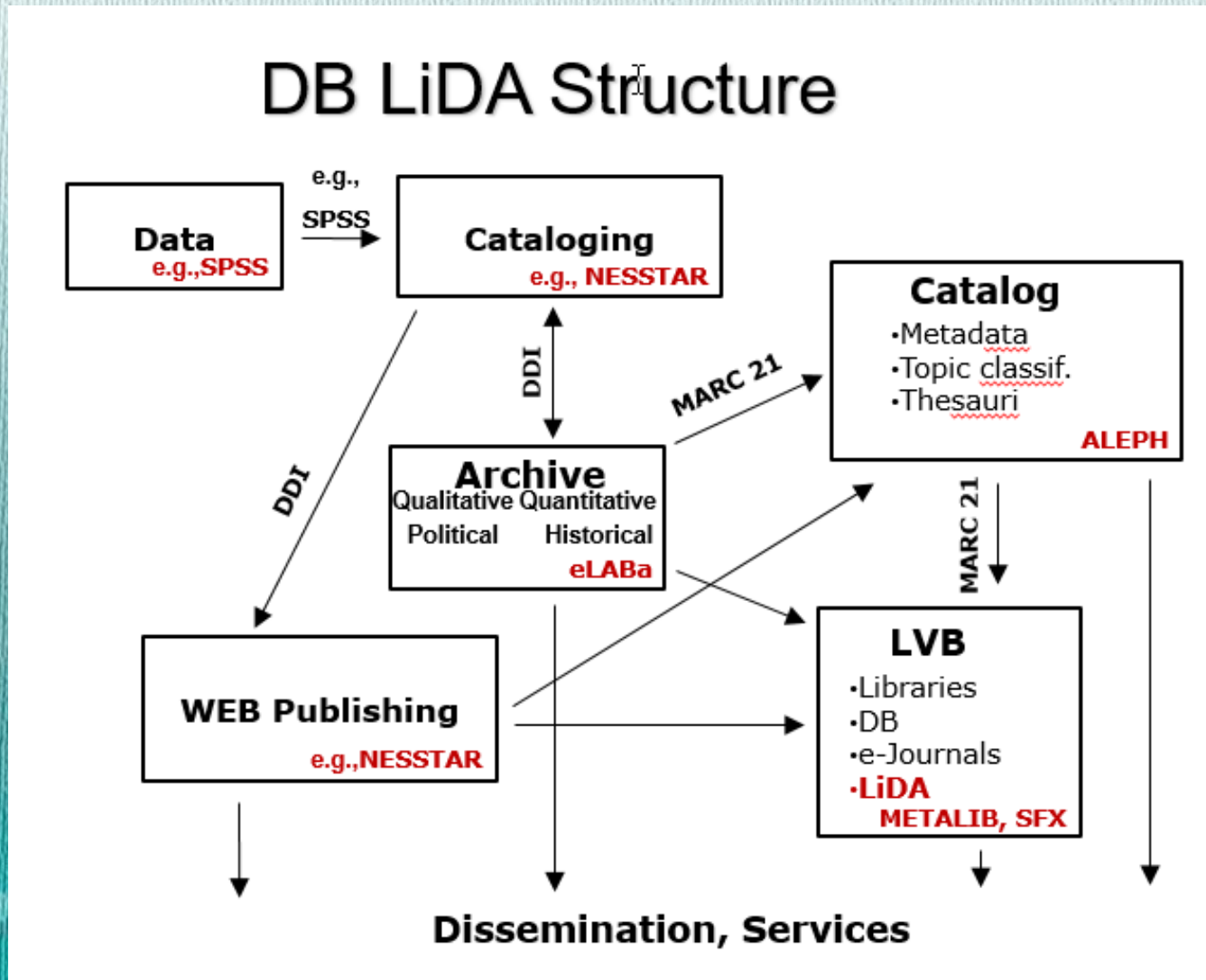
LiDA: IT infrastructure

- LiDA has **three level** IT infrastructure:
 - **Archiving** (repository of data objects) → **FEDORA**:
 - XML (DDI)
 - NESSTAR Server
 - SPSS
 - EXCEL
 - CSV
 - etc.

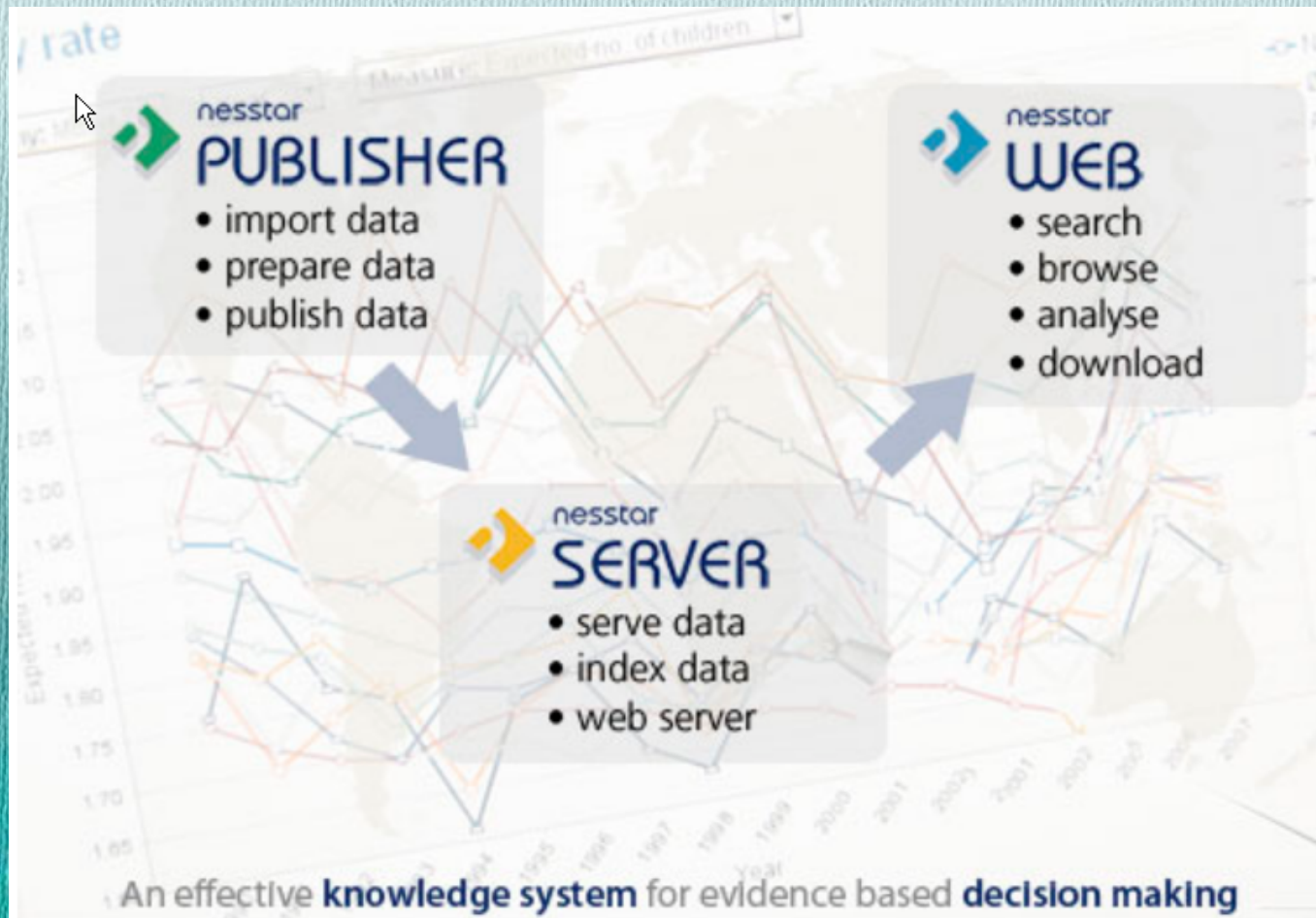
LiDA: IT infrastructure

- LiDA has **three level** IT infrastructure:
 - **Services:**
 - Data documentation → **NESSTAR Publisher**
 - Thesauri → **HASSET/ELSST**
 - Indexing and Searching
 - Visualization → **WEB, NESSTAR WebView**
 - Analysis → **NESSTAR WebView**
 - **Web portal**

LiDA: IT infrastructure (survey data)



LiDA: IT infrastructure – NESSTAR



LiDA: IT infrastructure (survey data)

- Ingest → **SPSS** data files (not open, but most common data format in survey research)
 - Metadata added with **NESSTAR Publisher**
 - **DDI** (1.2.2), in Lithuanian and English
 - **Keywords** (thesauri)
 - **Topic** classification
 - **PID**
 - Etc.

NESSTAR

- NESSTAR Publisher 3.54

Nesstar Publisher v3.54 - [LiDA_ESS_0150.NESSTAR.V3]

File Edit Documentation Variables Variable Groups Data Publishing Tools Window Help

Document Description Study Description Other Study Materials File Description Variables Variable Groups Data Entry Other Materials

Fields:

- 2.1.1.1-2.1.2.1
 - Title
 - Parallel Title
 - ID Number
- 2.1.2.1-2.1.3.1
 - Authoring Entity / Primary Investigator:
 - Other Identifications / Acknowledgments:
- 2.1.3.2-2.1.3.6
 - Copyright
 - Date of Production
 - Place of Production
 - Fundings
- 2.1.4.1-2.1.4.3
 - Distributors
 - Contact Persons
 - Depositor
 - Depositor Affiliation
- 2.1.4.4-2.1.5.2
 - Date of Deposit
 - Date of Distribution
 - Series Name

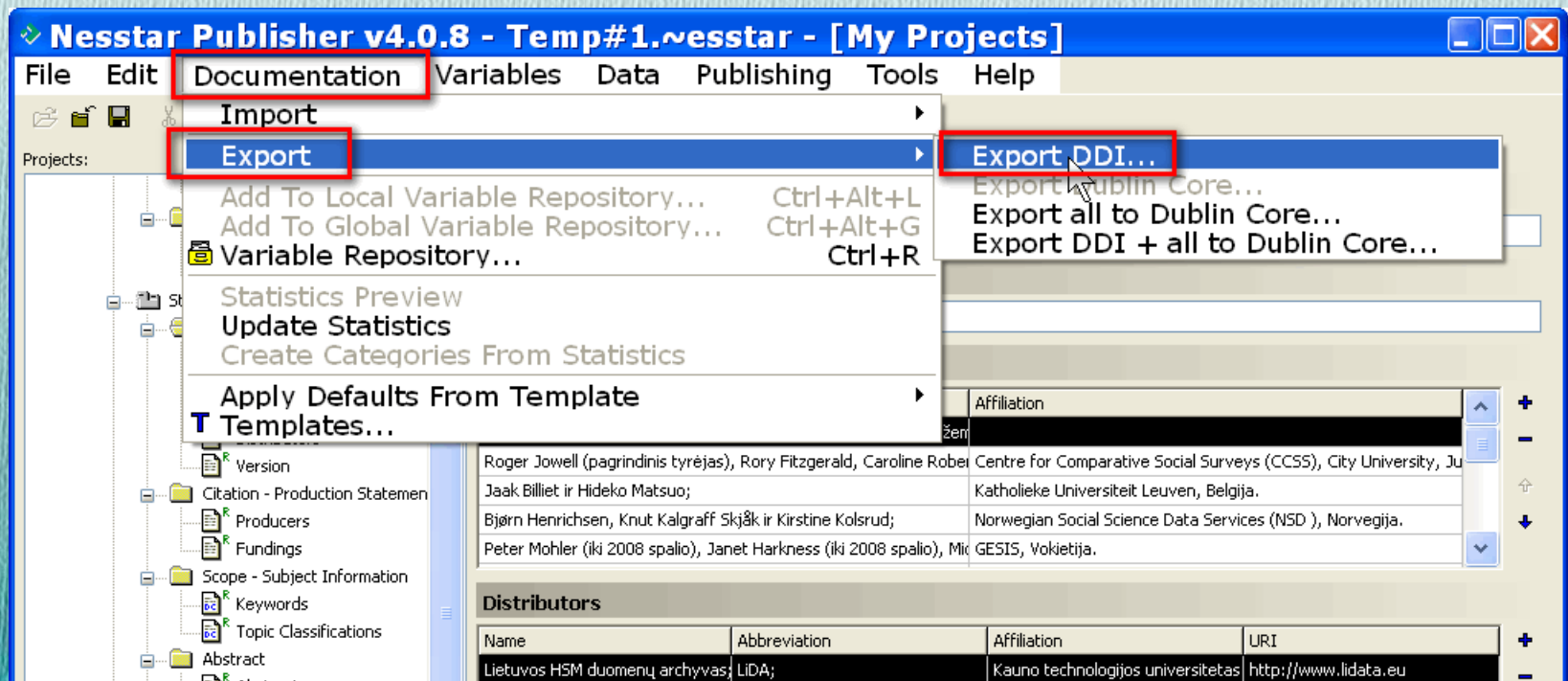
Field information:

Authoring Entity / Primary Investigator:	
Name	Affiliation
Jowell, Roger;	Centrinė koordinuojančioji taryba
Fitzgerald, Rory;	Centrinė koordinuojančioji taryba
Roberts, Caroline;	Centrinė koordinuojančioji taryba
Harrison, Eric;	Centrinė koordinuojančioji taryba
Eva, Gillian;	Centrinė koordinuojančioji taryba
Widdop, Sally;	Centrinė koordinuojančioji taryba
Keane, Mary;	Centrinė koordinuojančioji taryba
Billiet, Jaak;	Centrinė koordinuojančioji taryba

Other Identifications / Acknowledgments:	
Name	Affiliation
Kaase, Max;	ESS Mokslinė patariamoji taryba = The Scientific Advisory Board
Lievesley, Denise;	ESS Metodologijos grupė = The Methods Group of the ESS;
Häder, Sabine;	ESS Atrankos sudarymo specialistai = The Sampling Panel of the ESS;
Harkness, Janet;	ESS Vertimo priežiūros grupė = The Translation Taskforce of the ESS;
Abrams, Dominic;	ESS 4 bangos klausimyno bloko "Diskriminacija dėl amžiaus" koordinavimas
Svallfors, Stefan;	ESS 4 bangos klausimyno bloko "Socialinė gerovė" konstravimas
Curtice, John;	
Eriksson, Robert;	

NESSTAR

- NESSTAR Publisher 4.09



- Alternatives are available: SDA, Dataverse

LiDA: IT infrastructure (survey data)

- Archiving and Publication → **FEDORA repository** and **NESSTAR Server**
 - Metadata: **DDI** → DC, MARC21, etc.
 - Metadata: **OAI-PMH** → Lithuanian Virtual Library (LVB), LABT, Google, etc.
- Data access is provided on the **Web portal** and **NESSTAR WebView**

LiDA: IT infrastructure (survey data):

OAI-PMH

```

- <OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2013-10-20T23:59:13Z</responseDate>
  <request verb="ListRecords" from="2000-01-01" metadataPrefix="ddi" set="quant">http://www.lidata.eu/oaiprovider/</request>
- <ListRecords>
  - <record>
    - <header>
      <identifier>oai:lida.lt:LiDA:quant.LiDA_CCEB_0045</identifier>
      <timestamp>2011-12-01T17:05:02Z</timestamp>
      <setSpec>quant</setSpec>
    </header>
  - <metadata>
    - <codeBook ID="LiDA_CCEB_0045_STUDY_01" version="1.2.2">
      - <docDscr>
        - <citation>
          - <titlStmt>
            - <titl>
              Candidate Countries Eurobarometer 2001.1, October 2001: DDI document
            </titl>
            <subTitl> Version 2 = 2 versija </subTitl>
          - <altTitl>
              CCEB: Candidate Countries Eurobarometer 2001.1, October 2001
            </altTitl>
          - <parTitl>
              <b>Šlit:</b> Šalių kandidačių Eurobarometras 2001.1, 2001 m. spalio: DDI dokumentas
            </parTitl>
            <IDNo> LiDA_CCEB_0045_XML_02 </IDNo>
          </titlStmt>
        - <rspStmt>
          <AuthEnty affiliation="Kauno technologijos universitetas = Kaunas University of Technology">Krupavičius, Algis</AuthEnty>
        </rspStmt>
      </codeBook>
    </docDscr>
  </codeBook>
  </metadata>
</record>
</ListRecords>
</OAI-PMH>
  
```

LiDA: IT infrastructure (survey data): NESSTAR WebView



DESCRIPTION TABULATION



- LIDAKAT
 - ⊕ Apklausos
 - ⊕ Istoriniai duomenys
 - ⊕ Mokomieji duomenys
 - ⊕ Lietuvos politinės sistemos duomenys (test)
 - ⊕ Bandomasis Lietuvos statistikos departamento duomenų rinkinys (test)

LIDA internetinės analizės katalogas (LiDAKAT)

Sveiki prisijungę prie [Lietuvos HSM duomenų archyvo](#) internetinio duomenų analizės katalogo LiDAKAT. Tai socialinių ir humanitarinių mokslų duomenų portalas, kuriame galima naršyti, analizuoti ir parsisiųsti humanitarinių ir socialinių mokslų tyrimų duomenis. LiDAKAT sistema suteikia nesudėtingą prieigą prie duomenų analizės - nuo paprastų procentinių dažnių pasiskirstymų iki koreliacinių lentelių ir grafikų kūrimo. LiDAKAT esančių duomenų analizės rezultatai lengvai gali būti perkeltami į įvairaus formato bylas (Excel, Acrobat Reader etc.).

Šiuo metu esantys LiDA archyve [duomenų rinkiniai](#) yra pateikiami kairiajame LiDAKAT meniu. Ateityje numatoma plėsti prieigą prie duomenų rinkinių.

LiDAKAT paslauga yra nemokama, tačiau norint atlikti duomenų analizę ar parsisiųsti duomenis, būtina [registracija](#).

LiDAKAT naudojimo taisykles galite rasti mūsų svetainės [šoniniame meniu](#).

[Mums labai svarbi Jūsų nuomonė ir komentarai](#) apie LiDAKAT !

[Nesstar vartotojo vadovas](#)

Recent Datasets

2013-12-12

- [Tarptautinė socialinio tyrimo programa \(2012\): Šeima, darbas ir lyčių vaidmenys IV, 2013 m. sausis - vasaris](#)

2013-11-28

- [Tarptautinė socialinio tyrimo programa: Lietuvos socialinių problemų stebėseną. Bandomasis tyrimas, 2012 m. lapkritis](#)

2013-10-28

- [Virtualūs socialiniai tinklai I, 2010 m. lapkritis - 2011 m. vasaris](#)

2013-09-17

- [Lietuvos Respublikos liberalų sąjūdžio įstatai](#)

LiDA: IT infrastructure (survey data): Web portal

DATA SET

[Download](#) |  Nesstar

[Documentation](#) | [Variable List](#) | [Keywords](#) | [Publications](#)

Object URL:

http://www.lidata.eu/data/quant/LIDA_ESS_0282



Like Be the first of your friends to like this.

Title:

ESS7, pagrindinio ir papildomo klausimyno duomenys, Lietuva, 2015 m. balandis – birželis, 2 leidimas

Parallel Title:

ESS7, Main and Supplementary Questionnaire Data, Lithuania, April - June 2015, edition 2

Series Name:

ESS: Europos socialinis tyrimas = ESS: European Social Survey

Theme:

ŽINIASKLAIDA = MASS MEDIA

SOCIALINĖS NUOSTATOS IR ELGSENA = SOCIAL ATTITUDES AND BEHAVIOUR

POLITINĖ ELGSENA IR NUOSTATOS = POLITICAL BEHAVIOUR AND ATTITUDES






VALDŽIA IR POLITINĖS SISTEMOS = GOVERNMENT AND POLITICAL SYSTEMS

I YGYBĖ IR NELYGYBĖ = EQUALITY AND INEQUALITY

LiDA: IT infrastructure (survey data)

DOWNLOAD

Data Files:

Format	Name	Size in Kilobytes
	SPSS (SAV)	0
	SPSS (SPS)	0
	DDI	0
	DC	0
	NESSTAR (V3)	0



LiDA: IT infrastructure (survey data)

VARIABLE LIST

Group: VG5F1 (Kita)

V1 (Tyrimo identifikacinis numeris (LiDA))
 V2 (Versija (LiDA))
 V3 (Duomenų rinkinio URL (LiDA))
 V4 (ESS banga (ESS))
 V5 (Versija (ESS))
 V6 (Sukūrimo data (ESS))
 V7 (Respondento identifikacinis numeris)
 V396 (Regionas, Lietuva (pagal NUTS klasifikacijos 3 lygmenį – apskritis))
 V397 (Regioninio vieneto lygmuo)
 V398 (Interviu pradžia: mėnesio diena)

Show Variable

Group: VG6F1 (A. Žiniasklaida ir socialinis pasitikėjimas)

V8 (A1|TV žiūrėjimas, iš viso, įprastą darbo dieną)
 V9 (A2|TV žiūrėjimas, žinių arba politinių laidų, įprastą darbo dieną)
 V10 (A3|Dauguma žmonių galima pasitikėti, ar reikia būti labai atsargiam)
 V11 (A4|Dauguma žmonių stengtųsi jumis pasinaudoti, ar stengtųsi būti sąžiningi)
 V12 (A5|Žmonės dažniausiai stengiasi padėti kitiems, ar rūpinasi tik savimi)

Show Variable

LiDA: IT infrastructure (survey data)

VARIABLE

Variable V8 (tvttot) : A1|TV žiūrėjimas, iš viso, įprastą darbo dieną

Value	Category	N	%
0	Visai tam neskiriu laiko	87	3.9%
1	Mažiau negu 0,5 val.	71	3.2%
2	Nuo 0,5 val. iki 1 val.	192	8.5%
3	Daugiau negu 1 val., iki 1,5 val.	290	12.9%
4	Daugiau negu 1,5 val., iki 2 val.	412	18.3%
5	Daugiau negu 2 val., iki 2,5 val.	334	14.8%
6	Daugiau negu 2,5 val., iki 3 val.	297	13.2%
7	Daugiau negu 3 val.	533	23.7%
77	Atsisakau atsakyti	0	0%
88	Nežinau	34	1.5%
99	Nėra atsakymo	0	0%

Summary Statistics:

Valid cases: 2216

Missing cases: 34

Variable Format: numeric

LiDA: IT infrastructure (survey data)

NESSTAR WebView online analysis

- LiDA data catalogue (**LiDAKAT**) allows inspecting the data, processing it as well as elementary statistical **analysis online**
 - Results can easily be **exported**

LiDA: data sets

- **Four types** of data can be stored:
 - Survey data
 - Historical statistics
 - Data on Lithuanian political system (prototype)
 - Qualitative data (prototype)

LiDA: **survey** data sets

- **Public opinion** data is the biggest and most frequently used data collections (also in other national SSH data archives)
 - This data is mainly used for **secondary analysis**
 - Historical and/or cross-cultural research
 - **Almost 300** data sets available

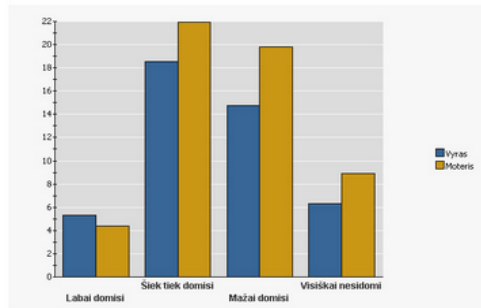
LiDA: **survey** data sets

- ☐ Duomenys
 - Paieška
 - Kintamųjų paieška
 - Paieška su tezauru
 - Duomenų katalogai
 - **Apklausų duomenys**
 - Istoriniai duomenys
 - Politiniai duomenys
 - Kokybiniai duomenys
 - Asociuoti ištekčiai
 - ICPSR
 - CESSDA
 - Kiti archyvai
- ☒ Naujoms vartotojams
- ☒ Duomenų depozitas
- ☒ Publikacijos

APKLAUSŲ DUOMENYS

Dataset: EES2009, rinkėjų tyrimo duomenys, Lietuva, 2009 m. birželis

Q102[Lytis] Q78[Domėjimasis politika] Type
 Q102[Lytis] Q78[Domėjimasis politika] Row and column percentage



© The European Union FP 7 PIREDEU project, The European University Institute



Visuomenės nuomonės apklausų ir tyrimų duomenų rinkiniai yra didžiausia ir dažniausiai naudojama duomenų archyvų kolekcija. Didžiąją dalį šios kolekcijos sudaro kitų mokslininkų ir tyrėjų jau įgyvendintų tyrimų duomenys, kuriuos kiti vartotojai gali naudoti atlikdami savo antrinę analizę.

2006 metais sukurtame LiDA kataloge saugomų apklausų duomenų rinkinių skaičius – 258. Jie padalinti į du katalogus:


- Lietuvoje atliktų tarptautinių tyrimų duomenų rinkiniai anglų kalba. Duomenų rinkinių skaičius kataloge - 33. Jie papildomai suskirstyti į tyrimų rinkinius (8).
- Lietuvoje atliktų tyrimų duomenų rinkiniai. Duomenų rinkinių skaičius kataloge – 225. Jie papildomai suskirstyti į tyrimų rinkinius (17).

Bendras LiDA kataloge patalpintų duomenų rinkinių kintamųjų skaičius yra 31051.

Kiekvienas apklausų duomenų rinkinys yra aprašytas (dokumentuotas) lietuvių ir anglų kalbomis.

Prieiga prie apklausų duomenų rinkinių ir su jais susijusios informacijos **nekomerčiais tikslais (mokslinių tyrimų, savišvietos ir mokymo) yra nemokama!**

Norėdami surasti jus dominančių apklausų duomenų rinkinį galite pasinaudoti **paieška**. Paiešką galima atlikti ir pasitelkiant **kintamųjų paiešką** arba kontroliuojamą žodyną (**HASSET tezaurą**).

Peržiūrėti informaciją apie apklausų duomenų rinkinius galite naršydami  **nesstar** kataloge arba LiDA portale:

- Visi rinkiniai
- Pagal autorių
- Pagal temas
- Pagal tyrimų rinkinius
- Pagal depozitorius
- Pagal finansavimo šaltinį
- Nauji rinkiniai

Atkreipiame dėmesį, kad norėdami atsisiųsti duomenis jums patogiu formatu (SPSS ir pan.) ar analizuoti juos online (t. y. atlikti **LiDA duomenų analizę Nesstar aplinkoje**), turite **prisiregistruoti**.

Apklausų duomenys į LiDA buvo deponuoti bendradarbiaujant su partneriais Lietuvoje ir užsienyje:



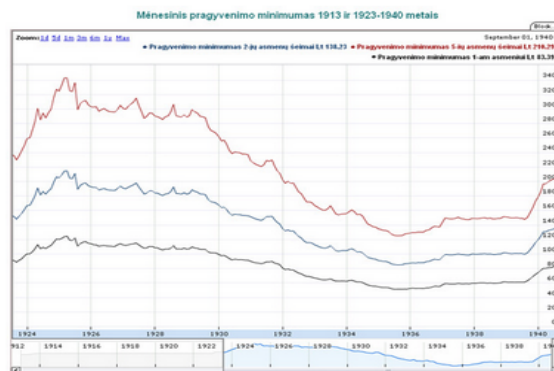
LiDA: data sets of **historical statistics**

- **Historical statistics** (mainly data tables) include data about Lithuanian population (some census data), economy, trade, socio-economic indicators, culture, education, public health etc.
 - Mainly, pre-II World War data (**old** data, not readily available at Statistics Lithuania)
 - **More than 60** data sets available

LiDA: data sets of **historical statistics**

- ☐ Duomenys
 - Paieška
 - Kintamųjų paieška
 - Paieška su tezauru
 - Duomenų katalogai
 - Apklauso duomenys
 - **Istoriniai duomenys**
 - Politiniai duomenys
 - Kokybiniai duomenys
 - Asocijuoti ištekliai
 - ICPSR
 - CESSDA
 - Kiti archyvai
- ☒ Naujiems vartotojams
- ☒ Duomenų depozitas
- ☒ Publikacijos

ISTORINIAI DUOMENYS (ISTORINĖ STATISTIKA)



Istorinės statistikos duomenų rinkiniai yra specializuota duomenų archyvų kolekcija, kurią savo tyrimuose dažniausiai naudoja istorikai. Šią kolekciją paprastai sudaro istoriniai duomenys apie gyventojų skaičių ir sudėtį, ūkį, prekybą, transportą ir ryšius, finansus, socialinius ir ekonominius santykius, sveikatos apsaugą ir sveikatos istoriją, socialinį gerbūvį, nelaimės (gaisrus, karų padarytus nuostolius ir pan.), švietimą, kultūrą, valstybės raidą, bažnyčias, miestus ir kt.

2010 metais sukurtame LiDA HISTAT kataloge saugomų istorinių statistinių duomenų rinkinių skaičius – **61**. Istoriniai duomenų rinkiniai, suskirstyti pagal laikotarpius:

- statistiniai duomenys iki XIX amžiaus,
- XIX amžiaus statistika,
- XX amžiaus statistika.

Duomenų rinkiniai taip pat suskirstyti į rinkinių grupes (9).

Kiekvienas istorinių duomenų rinkinys yra aprašytas (dokumentuotas) lietuvių ir anglų kalbomis.

Prieiga prie istorinių duomenų rinkinių ir su jais susijusios informacijos **nekomerciniais tikslais (mokslinių tyrimų, savivietos ir mokymo) yra nemokama!**

Norėdami surasti jus dominančių istorinių duomenų rinkinį galite pasinaudoti **paieška**. Paiešką galima atlikti ir pasitelkiant **kintamųjų paiešką** arba kontroliuojamą žodyną (**HASSET tezaurą**).

Peržiūrėti informaciją apie istorinių duomenų rinkinius galite naršydami  nesstar kataloge arba LiDA portale:

- Visi rinkiniai
- Pagal istorinius statistinius šaltinius
- Pagal temas
- Pagal rinkinių grupes
- Pagal depozitorius
- Pagal autorių
- Nauji rinkiniai

Atkreipiame dėmesį, kad norėdami atsisiųsti duomenis jums patogiu formatu (Excel, SPSS ir pan.), turite **prisiregistruoti**.

Istorinės statistikos duomenų rinkiniai parengti bendradarbiaujant su Lietuvos istorijos instituto ir Vilniaus universiteto Istorijos fakulteto mokslininkais.



LiDA: data access

- **Data** – freely available for non-commercial use (still not open)
 - Data provided in open and not open (but most commonly used) formats
 - Plans to make data open in the future
 - No API
- **Metadata** – freely available (already open)
 - Plans to make available in other standards

LiDA: data access

- **Search** → metadata search, variable search, search using thesauri (HASSET, ELSST)
- **Data inspection and analysis** → **NESSTAR**
 - Proprietary software
 - Other (and open source) platforms are becoming more popular (**Dataverse**)

LiDA: challenges

- **Technological** changes (progress)
 - IT infrastructures
 - Data and metadata standards
- **Funding and support**
 - Academic community still not fully aware of the benefits related to having specialized data archives (and open data, in general)
 - National funding agencies still hesitant about importance of data archives
 - What are the advantages to have them?
- **International** cooperation/integration

Thank you !