



Data management in the social sciences

*Eleonora Serafimovska, Aneta Cekikj,
Jovan Bliznakovski MK DASS, UCiM-ISPJR, Skopje*

*Webinar for social science researchers: **Data management in the social sciences**
October, 26, 2022*

DOI: [10.5281/zenodo.7250138](https://doi.org/10.5281/zenodo.7250138)

 cessda.eu

 [@CESSDA_Data](https://twitter.com/CESSDA_Data)



Licence: CC-BY 4.0

УПРАВУВАЊЕ СО ИСТРАЖУВАЧКИ ПОДАТОЦИ ВО ОПШТЕСТВЕНИТЕ НАУКИ

„животен“ циклус на истражувачките податоци

Елеонора Серафимовска (УКИМ-ИСППИ)

26 октомври 2022 година

 cessda.eu

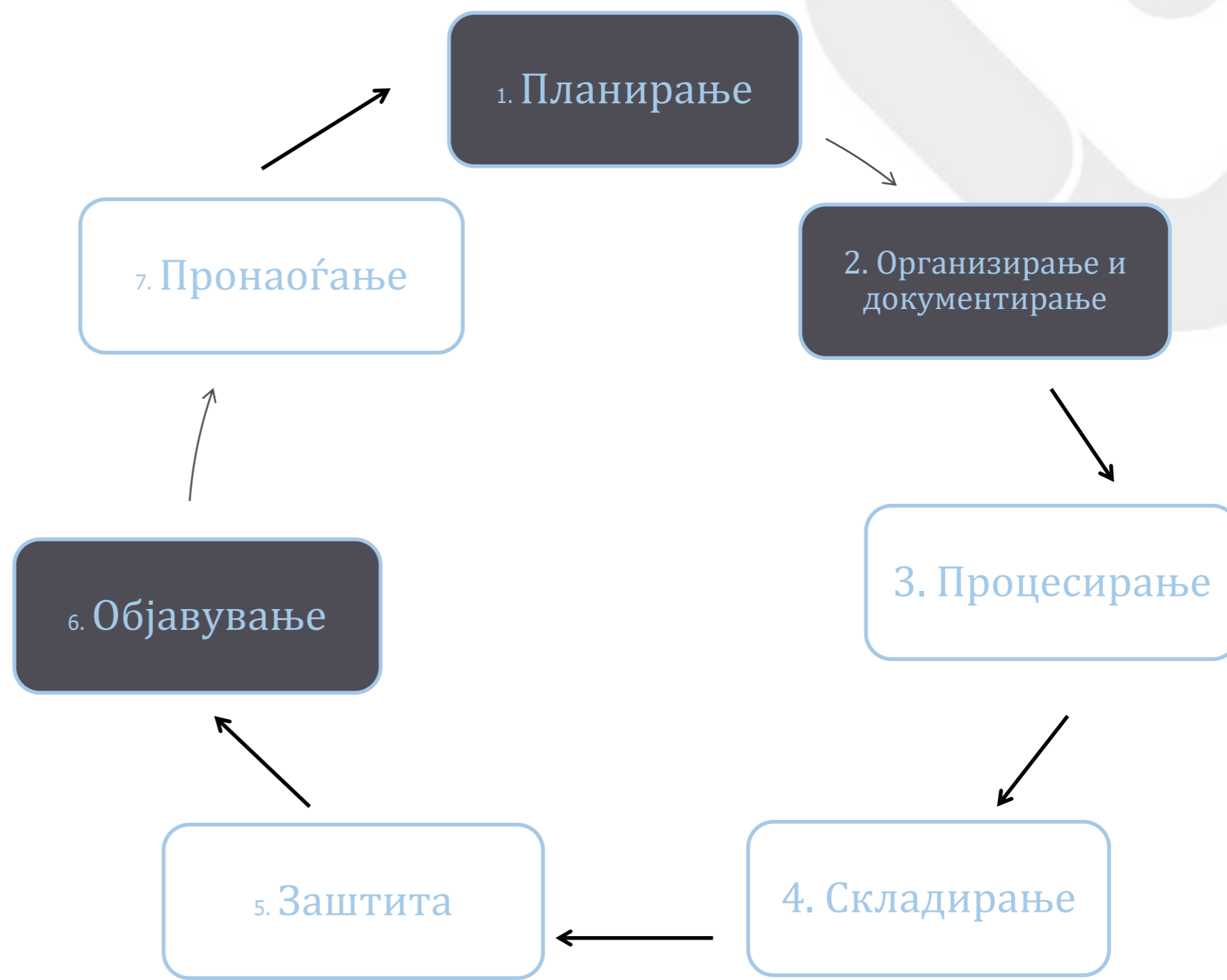
 @CESSDA_Data



Кој е животниот циклус на истражувачките податоци

„животниот“ циклус на податоците се однесува на целиот процес на управување, од започнувањето на планирањето на одреден проект и проектни активности, па сè до завршните етапи на архивирање и публикување на податоците и наодите од истражувачкиот проект.

Животниот циклус на истражувачките податоци се состои од седум последователни фази: планирање; организирање и документирање; процесирање; складирање; заштита; објавување и пронаоѓање на податоците



Планирање

Се објаснува што се подразбира под поимот „управувањето со податоци“, се дефинира „планот за управување/менаџирање на податоците“ (ПУП) и се нагласува во што се состои нивната важност. Во овој дел се дефинира основниот поим за „податоци од општествени науки“ и се објаснуваат принципите на „FAIR“ податоци.

Организирање и документирање

Приказ на најдобрите практики на креирање на соодветна структура на папки и датотеки на податоци, како и начините на правилно именување, документирање и организирање на датотетките со податоци.

Процесирање

Се посочуваат стандардните практики на внес и кодирање на податоците. Се даваат и основни насоки за форматирање на датотеката со податоци што треба да го осигура интегритетот на прибавените податоци.

Складирање

Запознавање со различни решенија за складирање и креирање на резервна копија, како и нивните предности и недостатоци. Во овој дел се посочува кои се начините на заштита од неовластено пристапување до податоците, како на пр. преку силен клучен збор или преку криптирање (шифрирање).

Заштита

Во овој дел се обработуваат правните и етичките прашања поврзани со истражувачките податоци. Се претставуваат практики на заштита како што се прибавување согласност од испитаниците или др. правни субјекти; највисоко можно ниво на анонимизирање на податоците; утврдување на носителот на авторски права на податоците; дефинирање на степените на пристап до податоците.

Објавување

Запознавање со можностите за архивирање и објавување на истражувачките податоци и наоди. Се претставуваат можните начини на промовирање на податоците во постоечките сервисии за архивирање и публикување.

Пронаоѓање

Се дискутираат начините на пронаоѓање и повторна, секундарна употреба (анализа) на веќе постоечки податоци или сетови на податоци кои се претходно прибавени и генерирани од други истражувачи или институции.



I фаза: ПЛАНИРАЊЕ

Содржина

*Вебинар: Управување со истражувачки податоци од општествените науки
26 октомври 2022*

Елеонора Серафимовска (УКИМ-ИСППИ)

- 1.1 Придобивки од управувањето со податоци**
- 1.2 Истражувачки податоци**
- 1.3 Податоци во општествените науки**
- 1.4 FAIR податоци**
- 1.5 Европски диверзитет**

 cessda.eu

 @CESSDA_Data



План за управување со податоци

Со цел да се поедностави управувањето со податоците, Планот за управување/менаџирање со податоци (ПУП а оригинално препознатлив како DMP-Data Management Plan) треба да се **создаде рано** во процесот на истражување.

ПУП е **формален документ** кој обезбедува рамка за управување со податоците за време на истражувањето, но и после тоа.

Нема универзален изглед како ПУП ќе изгледа откако истиот ќе биде завршен.

Тој е еден вид на „**жив**“ **документ** кој се менува заедно со потребите на проектот и на неговите учесници. Се ажурира во текот на целиот проект за да се обезбеди сигурност дека ги следи промените со текот на времето и дека ја рефлектира моменталната состојба на истражувачкиот проект.

Понекогаш има посебни барања од **засегнатите страни** на кои треба да се одговори во ПУП.



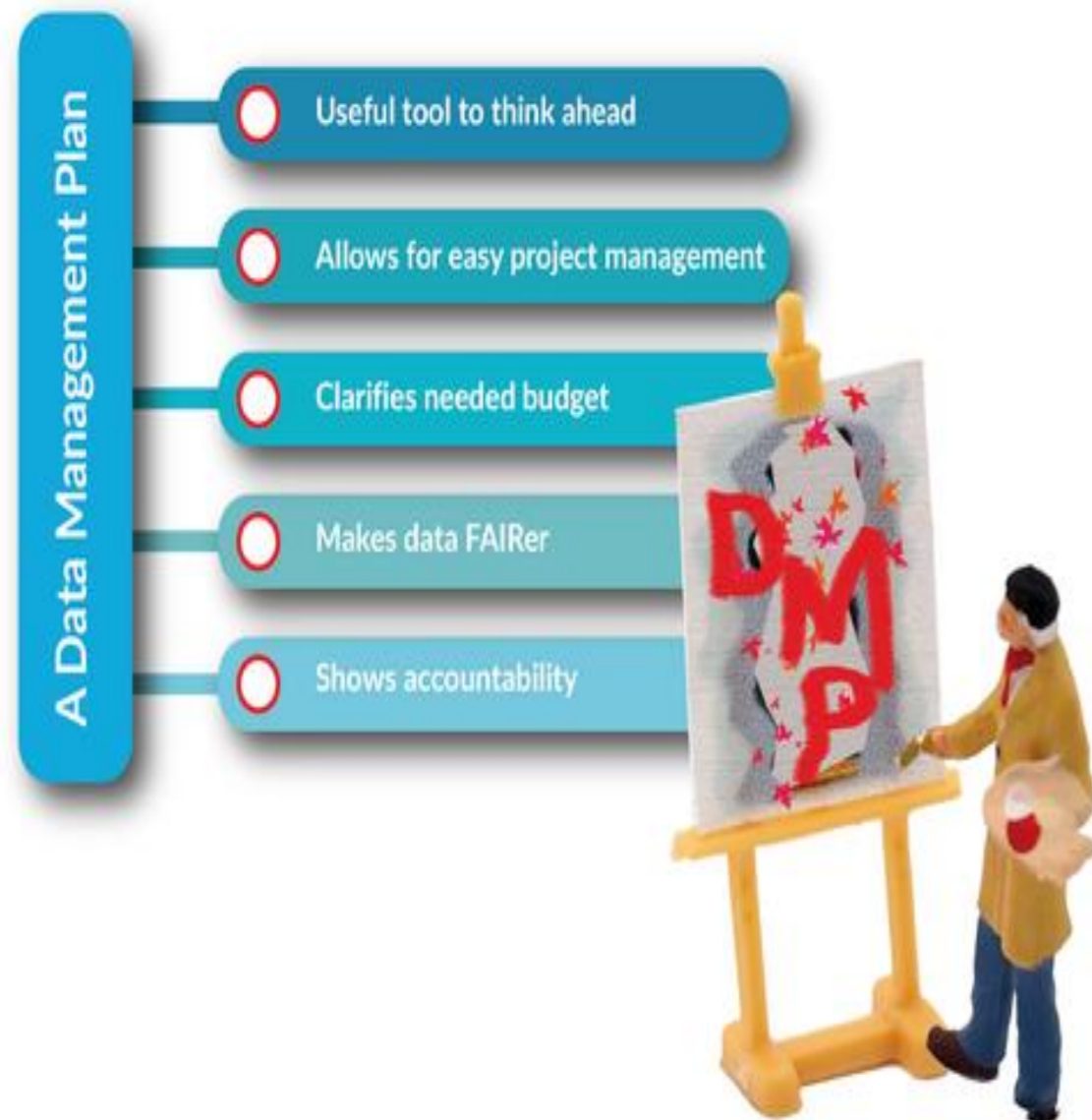
Финансиери

Финансиерите може да бараат од истражувачите План за управување со податоци за да добијат информации какви податоци тие имаат намера да соберат и дали (и како) ќе ги направат истите податоци достапни за другите.

Во овој случај, всушност агенцијата која финансира се обезбедува со секакви информации што им се потребни, до степенот што тие го определуваат.

Институцијата што го спроведува истражувањето

1.1 Придобивки од управувањето со податоци



Придобивка 1. Корисна алатка за размислување однапред

Со инвестирање време во планирање на проектот однапред може да превенираат низа проблеми откако проектот веќе ќе биде во тек и ќе се работи на истиот.

Придобивка 2. Овозможува лесно управување со проекти

Важна функција на ПУП е да работи како едношалтерски систем за да се пронајдат сите информации поврзани со проектот. Истражувањето станува многу полесно ако сите прашања во врска со управувањето со податоците се собираат на едно место и деталите поврзани со проектот се лесно достапни, отколку само нејасно запомнети или едноставно заборавени.

Придобивка 3. Одредување и појаснување на потребниот буџет за управување

Управувањето со податоци не е бесплатно. Сигурно не би сакале да наидете на финансии кои недостасуваат пред крајот на проектот, затоа што сте ги игнорирале или потцениле трошоците за структурно, детално и безбедно управување со податоци. Затоа, важен аспект на ПУП е негово внесување во пресметката, односно, колку пари ќе бидат потребни за управување со истражувачките податоци за време на истражувачкиот проект.

Придобивка 4. Правење на податоците повеќекратно употребливи

ПУП дозволува претходно да се размисли како да се обезбеди база на податоци до/за репозиториум на податоци што е можно по ФАИР.

Придобивка 5. Покажување одговорност

Со подготвувањето на ПУП, на институцијата која го спроведува истражувањето, финансиерите и партнерите на проектот, се покажува сериозен пристап во врска со управувањето со податокот од истражувањето, што вклучува и одговорен пристап кон фондовите за истражување и учесниците во истражувањето.

1.2 Истражувачки податоци

„Материјал што е поткрепа на истражувачките тврдења“ (Универзитет во Шефилд, 2013).

Опиливите форми што може да ги има овој „материјал“ се на пр. „факти, набљудувања, интервјуа, снимки, мерења, експерименти, симулации и софтвер; нумерички, описни и визуелни; сирови (row), исчистени и обработени“ (Van Berchum & Grootveld, 2017).

◆ Типови на податоци

Според изворот

регистри (административни, историски регистри, резултати од гласање, медицински, итн.)

Според физичкиот формат

нумерички, текстуални формати, слики, геопросторни, аудио, видео формати и софтвери

Според тоа како тие се создадени/ забележани

електронски текстуални документи, табеларни пресметки, лабораториски тетратки, раковници и дневници, прашалници, записници и шифри, аудио-ленти и видео-снимки, фотографии и филмови, резултати од испити, артефакти, слајдови, шеми за бази на податоци, содржина на база на податоци, модели, алгоритми и скрипти, работни текови (workflows), стандардни оперативни процедури и протоколи, експериментални резултати, метаподатоци и други датотеки со податоци како на пр. евиденција за преглед на литература и архиви на е-пошта

„Големи податоци“ (Big Data)

Научната литература обично ги опишува Големите податоци според нивните атрибути. Сите овие атрибути на англиски јазик започнуваат со буквата „V“ и тие се волумен (Volume), брзина (Velocity) и разновидност (Variety) (Купер, 2013)

- » Волуменот значи дека Големите податоци се многу големи и дека нивната обработка бара голема пресметковна моќ.
- » Брзината се залага за фактот дека големите податоци се произведуваат последователно и секој момент се појавуваат нови податоци.
- » Разновисноста нè потсетува дека Big Data се неструктурирани и неуредни и затоа не се подготвени за итна анализа
Некои автори додаваат уште две V, Вистинитост (Veracity) и Вредност (Value) (Wamba et al, 2015):
 - » Вистинитоста ни кажува дека Големите податоци мора внимателно да се испитаат од перспектива на нивната доверливост. Со други зборови, истражувачите треба да внимаваат на квалитетот на големите податоци.
- » Вредноста значи дека Големите податоци потенцијално генерираат вредни сознанија кои се важни за носителите на одлуки, креаторите на политики, истражувачите и различните организации.

Во зависност од нивниот извор, OECD (Organization for Economic Cooperation and Development. дефинира шест категории на Големи податоци:

A: Податоци кои произлегуваат од трансакциите на владата, на пример, даночните и социјалното осигурување.

B: Податоци кои ги опишуваат барањата за официјална регистрација или лиценцирање.

C: Комерцијални трансакции направени од поединци и организации.

D: Интернет податоци, кои произлегуваат од активностите за пребарување и социјално вмрежување.

E: Следење на податоци, следење на движењето на поединци или физички предмети кои се предмет на движење од луѓе.

F: Податоци за слики, особено воздушни и сателитски снимки, но вклучувајќи копнени видео слики.

Податоците за социјалните медиуми (категиорија D) се податоци од платформи како Facebook, Twitter, Instagram или YouTube.

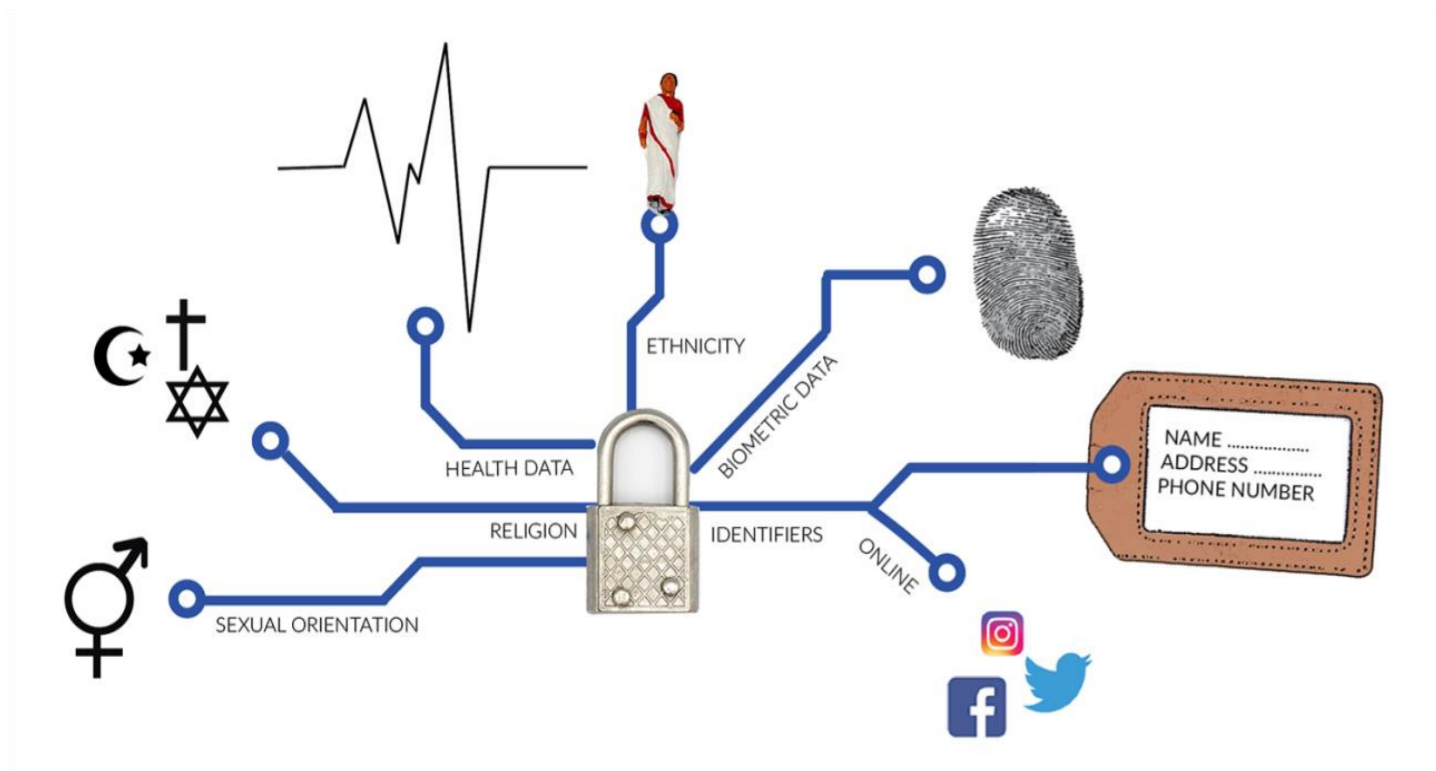
Формат

- ◇ Друг начин за кој треба да се размислува во врска со истражувачките податоци е **форматот** во кој се чуваат типовите на податоци (текстуален, нумерички, мултимедија, структуриран, софтверски код и сл.). На пример, статистичките податоци може да се чуваат во SPSS (*.sav) или STATA формат на датотеки; филмовите како *.mpg или *.avi; структурираните датотеки како *.xml или во релациска база на податоци MySQL и текстуалните датотеки како *.docx, *.pdf или *.rtf.
- ◇ **Големина и сложеност**
Големината на датотеките, како и нивната комплексност се секако важни. Управувањето со релативно мала и едноставна база на податоци, претставува различен предизвик од управувањето со големи, сложени датотеки со податоци.
- ◇ **Фаза на истражување**
Различните фази низ кои поминуваат истражувачките податоци (сирови, исчистени, обработени, анализирани податоци) вклучуваат свои предизвици за управување со истите.

1.3 Податоци во општествените науки

Во областа на општествените науки, честопати се работи со податоци кои потекнуваат од човечки учесници, што значи дека се работи за сензитивни, лични податоци кои заслужуваат посебно внимание.

- Доколку истражувачките податоците овозможуваат идентификација на одредена личност, тогаш истите се класифицираат како лични податоци. Во рамките на Општата европска регулатива за заштита на податоците (GDPR, Европската унија, 2016 година) личните податоци се дефинираат како сите информации во врска со идентификувано или препознатливо физичко лице познато како „субјект на податоци“.



- Расно и етничко потекло
- Политички убедувања
- Религиски или филозофски верувања
- Членство во синдикат
- Генетски податоци
- Биометриски податоци
- Податоци во врска со здравјето
- Податоци во врска со личниот сексуален живот или сексуалната ориентација на лицето

Тип на податоци	Квантитативни податоци	Квалитативни податоци
Општа дефиниција	Во квантитативно истражување, собраните информации се прикажани во нумеричка форма. Квантитативно истражување се користи за да се измерат однесувањето, ставовите или мислењата. Целта на квантитативните истражувања е често да се тестираат хипотези наведени на почетокот на истражувањето, да се формулираат факти и да се откријат обрасци.	Квалитативното истражување е првенствено експлоративно. Информациите кои ги собира не се во нумеричка форма. Целта на квалитативно истражување често е развивање (нови) идеи и подлабоко разбирање што не може да се постигне со нумерички оценки.
Предности на податоците	Податоците се изразени во броеви така што можат да се проценат со помош на статистички анализи.	Податоците се изразуваат на природен јазик, честопати во текстуален или визуелен приказ.
Методи за собирање на податоци	Квантитативни методи за прибирање податоци вклучуваат различни форми на истражувања - анкети преку Интернет, истражувања на хартија, мобилни истражувања и интервјуа лице в лице, телефонски интервјуа, пресретнувачи на веб-страници, експерименти и систематски набудувања. Во повеќето случаи, резултатите се генерализирани од поголема популација на примероци.	Квалитативните методи за собирање податоци вклучуваат фотографија, аудио записи, видео, неструктурирани интервјуа, полуструктурирани интервјуа, отворени прашалници, дневници, фокус групи (групни дискусии), индивидуални интервјуа и неструктурирани набљудувања. Големината на примерокот е обично помала од квантитативните примероци.
Пример за база на податоци	<p>Опис: Студија за начините на миграција на Летните олимписки игри меѓу 1948 и 2012 година. Во базата на податоци се опфатени приближно 40,000 спортисти и содржи информации за земјата што ја претставувале, како и за нивната земја на раѓање (отворен пристап, на англиски јазик).</p> <p>Reference: Jansen, J. (Erasmus University Rotterdam) (2017): Foreign-born Olympic athletes 1948 - 2012. DANS. https://doi.org/10.17026/dans-2xf-pyqp</p>	<p>Опис: Интервју со преживеан од логорот за истребување во втората светска војна Собибор (отворен пристап, на англиски јазик).</p> <p>Reference: Leydesdorff (copyright on the interview), prof. dr. S. (Universiteit van Amsterdam - dep. of Arts, Religion and Culture); Huffener (project manager), M. (Stichting Sobibor) (2012): Project 'Long shadow of Sobibor' Survivors: Interview 01 Thomas Blatt. DANS. https://doi.org/10.17026/dans-x8h-fwjg</p>

1.4 FAIR податоци

Еден од согледаните начини за постигнување на оптимална повторна употреба е податоците да се направат FAIR



Findable

To aid automatic discovery of relevant datasets, (meta)data should be easy to find by both humans and machines and be assigned a persistent identifier.

Accessible

Limitations on the use of data, and protocols for querying or copying data are made explicit for both humans and machines.

Interoperable

(Meta)data should use standardised terms (controlled vocabularies), have references to other (meta)data and be machine actionable.

Reusable

(Meta)data are sufficiently well described for both humans and computers to be able to understand them and have a clear and accessible data usage license.

Чекори кон FAIR податоци

- **Перзистентен идентификатор (ПИД од PID persistent identifier) за предметот на податоците како целина –**

Постојаните идентификатори (како DOI) спречуваат губење на врската/линкот.

Губењето на линкот е процес со кој хиперврските прекинуваат да се реферираат на оригиналниот извор низ времето, бидејќи тие се преместени или избришани. Без PID, предметот на податоци едноставно нема да може да се најде, а камоли да може повторно да се употреби на долг рок

- **Доволен сет од метаподатоци**

Доволен и стандардизиран збир/комплет на метаподатоци (што ги опишуваат податоците) ќе го зголемат пронаоѓањето, интероперабилноста и повторната употреба на истражувачките податоци. Квалитетот на описни информации во врска со податоците има големо влијание врз нивната повторна употреба.

Јасна лиценца

Истражувачите (и компјутерите) кои наоѓаат бази на податоци веднаш треба да знаат што им е дозволено да прават со нив. Да се наведат јасни права за повторна употреба е како да се има топло „Добредојдовте“ пред вратата на базата на податоци. Мотото е „отвореност доколку е можно, рестрикции ако е потребно“.

Доверлив репозиториум за податоци

кое ги има овие FAIR атрибути вградени во инфраструктурата за поднесување на податоци.

Интерес за FAIR е истражувачите да ги депонираат своите податоци заедно со целата документација потребна за нивно разбирање и повторна употреба во доверлива архива за податоци за истражување, што има експлицитна цел за зачувување на податоците и потребна експертиза за одржливо чување на податоците и задржување на нивната употребливост.

1.5 Европски диверзитет

Барања за управување со податоци во Европа

Постојат многу различни локални, национални и меѓународни шаблони/обрасци за ПУП и алатки што можат да се искористат за да се создаде ПУП за одреден истражувачки проект. Во оваа фаза најдобро е да се проверат обрасците или алатките кои што најдобро одговараат на одредената специфична ситуација.

Истражувачите можат да ги консултираат своите универзитети или оддели/институции дали имаат одреден префериран образец за изработка на ПУП. Или можеби финансиерот на истражувањето има специфични барања во врска со форматот на ПУП.

1. Data Collection and Documentation	2. Ethics, legal and security Issues	3. Data Storage and Preservation	4. Data Sharing and reuse
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> What kind of data are generated	<input type="checkbox"/> How will ethical issues be handled	<input type="checkbox"/> How are the data stored?	<input type="checkbox"/> How and where will the data be shared?
<input type="checkbox"/> How will data be generated	<input type="checkbox"/> How are the data accessed	<input type="checkbox"/> Are there back up systems	<input type="checkbox"/> How are sensitive data protected
<input type="checkbox"/> What metadata are needed	<input type="checkbox"/> Are there copyright issues	<input type="checkbox"/> How are data safely preserved	<input type="checkbox"/> How can data be accessed
	<input type="checkbox"/> Are there sensitive data		
	<input type="checkbox"/> What about intellectual property rights		

1.5 Европски диверзитет



Data Management Plan – content of the mySNF form

Question	Help text
1 Data collection and documentation	
1.1 What data will you collect, observe, generate or reuse? Questions you might want to consider: - What type, format and volume of data will you collect, observe, generate or reuse? - Which existing data (yours or third-party) will you reuse?	Briefly describe the data you will collect, observe or generate. Also mention any existing data that will be (re)used. The descriptions should include the type, format and content of each dataset. Furthermore, provide an estimation of the volume of the generated data sets. (This relates to the <i>FAIR Data Principles</i> F2, I3, R1 & R1.2)
1.2 How will the data be collected, observed or generated? Questions you might want to consider: - What standards, methodologies or quality assurance processes will you use? - How will you organize your files and handle versioning?	Explain how the data will be collected, observed or generated. Describe how you plan to control and document the consistency and quality of the collected data: calibration processes, repeated measurements, data recording standards, usage of controlled vocabularies, data entry validation, data peer review, etc. Discuss how the data management will be handled during the project, mentioning for example naming conventions, version control and folder structures. (This relates to the <i>FAIR Data Principle</i> R1)
1.3 What documentation and metadata will you provide with the data? Questions you might want to consider: - What information is required for users (computer or human) to read and interpret the data in the future? - How will you generate this documentation? - What community standards (if any) will be used to annotate the (meta)data?	Describe all types of documentation (README files, metadata, etc.) you will provide to help secondary users to understand and reuse your data. Metadata should at least include basic details allowing other users (computer or human) to find the data. This includes at least a name and a persistent identifier for each file, the name of the person who collected or contributed to the data, the date of collection and the conditions to access the data. Furthermore, the documentation may include details on the methodology used, information about the performed processing and analytical steps, variable definitions, references to vocabularies used, as well as units of measurement. Wherever possible, the documentation should follow existing community standards and guidelines. Explain how you will prepare and share this information. (This relates to the <i>FAIR Data Principles</i> I1, I2, I3, R1, R1.2 & R1.3)

2 Ethics, legal and security issues

2.1 How will ethical issues be addressed and handled? Questions you might want to consider: - What is the relevant protection standard for your data? Are you bound by a confidentiality agreement? - Do you have the necessary permission to obtain, process, preserve and share the data? Have the people whose data you are using been informed or did they give their consent? - What methods will you use to ensure the protection of personal or other sensitive data?	Ethical issues in research projects demand for an adaptation of research data management practices, e.g. how data is stored, who can access/reuse the data and how long the data is stored. Methods to manage ethical concerns may include: anonymization of data; gain approval by ethics committees; formal consent agreements. You should outline that all ethical issues in your project have been identified, including the corresponding measures in data management. (This relates to the <i>FAIR Data Principle</i> A1)
2.2 How will data access and security be managed? Questions you might want to consider: - What are the main concerns regarding data security, what are the levels of risk and what measures are in place to handle security risks? - How will you regulate data access rights/permissions to ensure the security of the data? - How will personal or other sensitive data be handled to ensure safe data storage and -transfer?	If you work with personal or other sensitive data you should outline the security measures in order to protect the data. Please list formal standards which will be adopted in your study. An example is ISO 27001-Information security management. Furthermore, describe the main processes or facilities for storage and processing of personal or other sensitive data. (This relates to the <i>FAIR Data Principle</i> A1)
2.3 How will you handle copyright and Intellectual Property Rights issues? Questions you might want to consider: - Who will be the owner of the data? - Which licenses will be applied to the data? - What restrictions apply to the reuse of third-party data?	Outline the owners of the copyright and Intellectual Property Right (IPR) of all data that will be collected and generated, including the licence(s). For consortia, an IPR ownership agreement might be necessary. You should comply with relevant funder, institutional, departmental or group policies on copyright or IPR. Furthermore, clarify what permissions are required should third-party data be reused. (This relates to the <i>FAIR Data Principles</i> I3 & R1.1)

3 Data storage and preservation

3.1 How will your data be stored and backed-up during the research?
Questions you might want to consider:
- What are your storage capacity and where will the data be stored?
- What are the back-up procedures?

Please mention what the needs are in terms of data storage and where the data will be stored. Please consider that data storage on laptops or hard drives, for example, is risky. Storage through IT teams is safer. If external services are asked for, it is important that this does not conflict with the policy of each entity involved in the project, especially concerning the issue of sensitive data. Please specify your back-up procedure (frequency of updates, responsibilities, automatic/manual process, security measures, etc.)

3.2 What is your data preservation plan?
Questions you might want to consider:
- What procedures would be used to select data to be preserved?
- What file formats will be used for preservation?

Please specify which data will be retained, shared and archived after the completion of the project and the corresponding data selection procedure (e.g. long-term value, potential value for reuse, obligations to destroy some data, etc.). Please outline a long-term preservation plan for the datasets beyond the lifetime of the project. In particular, comment on the choice of file formats and the use of community standards. (This relates to the *FAIR Data Principles* F2 & R1.3)

4. Data sharing and reuse

4.1 How and where will the data be shared?
Questions you might want to consider:
- On which repository do you plan to share your data?
- How will potential users find out about your data?

Consider how and on which repository the data will be made available. The methods applied to data sharing will depend on several factors such as the type, size, complexity and sensitivity of data. Please also consider how the reuse of your data will be valued and acknowledged by other researchers. (This relates to the *FAIR Data Principles* F1, F3, F4, A1, A1.1, A1.2 & A2)

4.2 Are there any necessary limitations to protect sensitive data?
Questions you might want to consider:
- Under which conditions will the data be made available (timing of data release, reason for delay if applicable)?

Data have to be shared as soon as possible, but at the latest at the time of publication of the respective scientific output. Restrictions may be only due to legal, ethical, copyright, confidentiality or other clauses. Consider whether a non-disclosure agreement would give sufficient protection for confidential data. (This relates to the *FAIR Data Principles* A1 & R1.1)

4.3 I will choose digital repositories that are conform to the FAIR Data Principles. [CHECK BOX]

The SNSF requires that repositories are conform to the FAIR Data Principles (Section 5 of the [guidelines for researchers](#), SNSF's explanation of the [FAIR Data Principles](#)). If there are no repositories complying with these requirements in your research field, please deposit a copy of your data on a generic platform (see [examples](#)).
If no data can be shared, this is a statement of principles.

4.4 I will choose digital repositories maintained by a non-profit organisation. [RADIO BUTTON yes/no]
→ If the answer is no: "Explain why you cannot share your data on a non-commercial digital repository."

The SNSF supports the use of non-commercial repositories for data sharing. Costs related to data upload are only covered for non-commercial repositories.

Елементи и соодветните прашања обично се вклучени во оваа првична фаза во конструирањето на ПМО.

Наслов на проектот:

Кој е насловот на проектот? Понудете краток опис

Датум и верзија на планот

- Кога е направена оваа верзија од ПУП?
- Како ги разликувате верзиите на ПУП?

Опис на проектот:

- Која е природата на проектот?
- Кое е истражувачкото прашање?
- Која е временската рамка на проектот?

Потекло на податоците:

- Каков вид на податоци ќе се користат за време на проектот?
- Доколку се користат веќе постоечки податоци: Кој е опсегот, обемот и форматот? Како се интегрирани различните извори на податоци?
- Ако собираете нови податоци, дели може да се појасни зошто тоа е потребно?

Главни истражувачи и соработници

Главни истражувачи

- Кои главни истражувачи се вклучени?
- Кои се нивните детали за контакт?

Истражувачи соработници (доколку се применети)

- Кои се нивните детали за контакт и нивните улоги во проектот?

Финансирање (ако е применето):

Доколку се додели финансирање, кој е референтниот број на доделеното финансирање?

Производител на податоци:

Која организација има административна одговорност за податоците?

Контакт со податоци за проектот:

Со кого може да се контактира во врска со проектот откако ќе заврши?

Сопственост на податоци:

- Која организација(и) ги поседува податоците?
- Ако се вклучени повеќе организации, која организација, со кои податоци располага?

Улоги

- Кој е одговорен за ажурирање на ПУП, и дали истиот се следи во извршувањето на должностите?
- Дали учесниците во проектот имаат специфични улоги?
- Која е временската рамка на проектот?

Трошоци

- Дали има трошоци што треба да се земат предвид, за купување на специфичен софтвер или хардвер?
- Дали има трошоци што треба да се земат предвид за чување и зачувување на базата?
- Дали се опфатени потенцијалните трошоци за (спремање на податоците за) архивирање?

1.6 Адаптирање на ПУП



II фаза: организирање и документирање

Елеонора Серафимовска (УКИМ-ИСППИ)

*Вебинар: Управување со истражувачки податоци од општествените науки
26 октомври 2022*

 cessda.eu

 @CESSDA_Data



Организирање и документирање

Начините на организација и документација на податоците и метаподатоците преку правилно креирање на датотеки со податоци (files), именување на датотеките и организирање и групирање на податоците во соодветни структури на папки (folders).



Организирањето и документирањето на податоците е важен процес бидејќи начинот на кој се структурираат податоците овозможува едноставно и брзо пристапување до нив, овозможува полесно да се идентификуваат информациите кои ги содржат и спречува да дојде до забуна или погрешно интерпретирање.

2.1 Дизајнирање на структурата на датотека со податоци (data file structure)

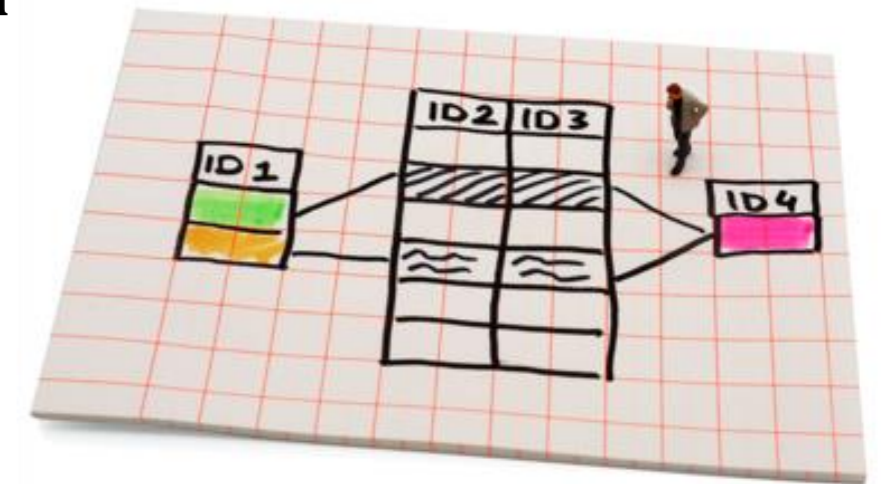
Избор на структура на датотека

При изборот на структурата на датотеката три елементи се важни и треба да бидат разгледани.

→ **каков опфат има дадената истражувачката студија** во однос на потребните податоци што треба да се прибават, бидејќи одредена истражувачка студија може да се состои од повеќе различни датотеки кои вообичаено се различно поврзани меѓу себе, а тоа ќе влијае на изборот на форматот на секоја датотека.

→ **внатрешната структура на датотеката може да биде различна и е определена од начинот на форматирање на содржината на датотеката (на пр. видови на променливи /варијабли, оригинално на англиски- variables/ и нивно организирање во датотеката).** Структурата на датотеката често дава информации и за односите меѓу различните елементи и делови од содржината на податоците.

→ **во датотеката се содржани важни информации** кои потоа се користат во креирањето на документот за метаподатоци на истражувачката студија (на пр. форматирањето на: имињата на променливите (variable name), ознаките (од англиски variable label) за променливите и ознаките за вредностите на променливата, разновидните забелешки или содржината на одредени пропратни, дополнителни променливи/варијабли). Конечно, за да се креира јасна, прецизна и точна документација на податоците, големо влијание има правилното структурирање на истражувачките податоци.



Дизајнирање на структурата на датотеки со квалитативни податоци

Квалитативните датотеки можат да содржат различни истражувачки материјали

:

**Текстови - транскрипции од интервјуа или сесии на фокус групи;
различни форми на пишани текстови како што се весници, списанија, дневници и сл.;**
фотографии;
аудио фајлови (снимени говори);
видео фајлови.



Карактеристично за квалитативните податоци е тоа што тие **не се прикажуваат во форма на променливи, бројки, матрици на податоци и др. форми на квантификација.**

Најсоодветно е **секоја индивидуална колекција на податоци да биде структурирана во одделен индивидуална датотетка** на пр. за секое поединечно интервју, или една фотографија или еден аудио запис да се креира посебна датотетка. Поединечните датотеки можат да се групираат во папки со слични датотеки.

Дизајнирање на квантитативните датотеки на податоци



Квантитативни податоци = податоците добиени од анкети

Три видови на вообичаени структури на датотеки во квантитативните општествени исражувања:

**рамна (flat) или правоаголна,
хиерархиска и
релативна**

Правоаголна датотека (flat file)	
Опис	Типови датотеки
<p>Рамните (правоаголни) датотеки со податоци се организираат во долги редови, променлива по променлива. Еден ред е посветен на еден испитаник/ субјект на набљудување и/или анализа.</p> <p>Идентификацискиот број (ID) обично доаѓа на прво место. Ако променливите се организираат колона по колона, тогаш се добива правоаголна матрица.</p>	<p>SPSS и STATA и слични софтвери и често се користат за анализа на правоаголните датотеки. Тука структурата се состои од една правоаголна матрица со податоци, придружени со варијабли (variable) и нивните називи(variable label). Во овој случај, секој запис (испитаник/субјект на анализа) вклучува иста количина на информации и има иста должина како и сите други записи во датотеката со податоци. Ако варијаблата не е применлива за одредено набљудување, таа е исполнета со празни места или вредности што недостасуваат (missing values).</p>

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	HH_ide	Numeric	8	0	identification of ...	None	None	8	Right	Scale	Input
2	member	Numeric	9	0		None	None	8	Right	Nominal	Input
3	country	Numeric	8	0		{1, cz}...	None	8	Right	Nominal	Input
4	serial	Numeric	8	0		None	None	8	Right	Nominal	Input
5	status	Numeric	8	0		{1, PARTNE...	None	8	Right	Nominal	Input
6	sex	Numeric	8	0		{1, male}...	None	8	Right	Nominal	Input
7	birth	Numeric	8	0		None	None	8	Right	Scale	Input
8	marit	Numeric	8	0		{1, single}...	None	8	Right	Nominal	Input
9	date	Numeric	8	0		None	None	8	Right	Scale	Input
10	educ	Numeric	8	0		{1, DID NOT...	None	8	Right	Nominal	Input
11	IDE_individ...	Numeric	8	0		None	None	10	Right	Scale	Input
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											
26											
27											
28											
29											

Хиерархиска датотека	
Опис	Типови датотеки
<p>Хиерархиските датотеки се состојат од записи со повисок ред и понизок ред, кои се распоредени во хиерархиска структура, т.е. неколку единици со понизок ред може да бидат поврзани со една единица со повисок ред и се содржани во истата датотека со податоци.</p>	<p>Ако во одредена база на податоци има различни нивоа на единици, датотеката со рамни податоци може да биде непрактична затоа што може да вклучува голем број празни места и да постави големи побарувања во однос на големината на датотеката. Покрај тоа, може да ја намали оперативноста и јасноста во однос на диференцијацијата на видовите единици на анализа. Апликации за бази на податоци како на пр. датабазата, MS Access или SQL, овозможуваат структурирање на истражувачките податоци по хиерархиски редослед.</p>

Релациска база на податоци	
Опис	Типови датотеки
<p>Релациската база на податоци е систем на неколку матрици на податоци и дефинирани асоцијации помеѓу нив.</p>	<p>Различни други апликации за бази на податоци, на пример, бази на податоци Д-база, MS Access или SQL, овозможуваат структурирање на истражувачките податоци по хиерархиски редослед. Истражувачите можат да ги поделат истражувачките податоци во неколку меѓусебно поврзани рамни датотеки, т.е. да ги структурираат податоците во релациска база на податоци и да ја задржат можноста за користење статистички софтвер споменат погоре.</p>

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role	
1	HH_ide	Numeric	8	2	identification of...	None	None	10	Right	Scale	Input
2	country	Numeric	6	0	{1, CZ}...	None	None	8	Right	Nominal	Input
3	serial1	Numeric	1	0	SERIAL NUMB...	None	None	8	Right	Nominal	Input
4	serial2	Numeric	1	0	SERIAL NUMB...	None	None	8	Right	Scale	Input
5	serial3	Numeric	1	0	SERIAL NUMB...	None	None	8	Right	Scale	Input
6	serial4	Numeric	1	0	SERIAL NUMB...	None	None	8	Right	Scale	Input
7	serial5	Numeric	1	0	SERIAL NUMB...	None	None	8	Right	Scale	Input
8	serial6	Numeric	1	0	SERIAL NUMB...	None	None	8	Right	Scale	Input
9	serial7	Numeric	1	0	SERIAL NUMB...	None	None	8	Right	Scale	Input
10	serial8	Numeric	1	0	SERIAL NUMB...	None	None	8	Right	Scale	Input
11	serial9	Numeric	1	0	SERIAL NUMB...	None	None	8	Right	Scale	Input
12	status1	Numeric	1	0	HOUSEHOLD ...	{1, PARTNE...	None	8	Right	Nominal	Input
13	status2	Numeric	1	0	HOUSEHOLD ...	{1, PARTNE...	None	8	Right	Nominal	Input
14	status3	Numeric	1	0	HOUSEHOLD ...	{1, PARTNE...	None	8	Right	Nominal	Input
15	status4	Numeric	1	0	HOUSEHOLD ...	{1, PARTNE...	None	8	Right	Nominal	Input
16	status5	Numeric	1	0	HOUSEHOLD ...	{1, PARTNE...	None	8	Right	Nominal	Input
17	status6	Numeric	1	0	HOUSEHOLD ...	{1, PARTNE...	None	8	Right	Nominal	Input
18	status7	Numeric	1	0	HOUSEHOLD ...	{1, PARTNE...	None	8	Right	Nominal	Input
19	status8	Numeric	1	0	HOUSEHOLD ...	{1, PARTNE...	None	8	Right	Nominal	Input
20	status9	Numeric	1	0	HOUSEHOLD ...	{1, PARTNE...	None	8	Right	Nominal	Input
21	sex1	Numeric	1	0	SEX OF THE 1...	{1, MALE}...	None	8	Right	Nominal	Input
22	sex2	Numeric	1	0	SEX OF THE 2...	{1, MALE}...	None	8	Right	Nominal	Input
23	sex3	Numeric	1	0	SEX OF THE 3...	{1, MALE}...	None	8	Right	Nominal	Input
24	sex4	Numeric	1	0	SEX OF THE 4...	{1, MALE}...	None	8	Right	Nominal	Input
25	sex5	Numeric	1	0	SEX OF THE 5...	{1, MALE}...	None	8	Right	Nominal	Input
26	sex6	Numeric	1	0	SEX OF THE 6...	{1, MALE}...	None	8	Right	Nominal	Input
27	sex7	Numeric	1	0	SEX OF THE 7...	{1, MALE}...	None	8	Right	Nominal	Input
28	sex8	Numeric	1	0	SEX OF THE 8...	{1, MALE}...	None	8	Right	Nominal	Input
29	sex9	Numeric	1	0	SEX OF THE 9...	{1, MALE}...	None	8	Right	Nominal	Input
30	birth1	Numeric	2	0	YEAR OF BIRT...	None	None	8	Right	Scale	Input

1: status1	status1	status2	status3	status4	status5	status6	status7	status8	status9	sex1	sex2	sex3	sex4	sex5	sex6	sex7	sex8	sex9
1	0	1
2	0	2	2	1
3	0	2	4	1	2	1
4	0	2
5	0	1	2	2	2	2	.	.	.	1	2	1	1	2	2	.	.	.
6	0	6	6	1	2	1
7	0	1	2	1	2	1
8	0	1	1	2
9	0	1	2	2	1	2	2	2
10	0	1	2	2	1	2	1	2
11	0	1	2	2	1	2	1	1
12	0	1	2	2	1	2	2	1
13	0	1	2	1	2	1
14	0	1	2	2	1	2	1	2
15	0	4	2	1
16	0	2
17	0	1	1	2
18	0	2	1	2
19	0	1	2	2	1	2	1	1
20	0	1	2	2	2	1	2	1	1	1
21	0	1	1	2
22	0	1	2	2	1	2	2	1
23	0	1	1	2
24	0	1	2	2	1	2	1	2
25	0	1	2	2	1	2
26	0	1
27	0	2	2	1

2.1.1 Организација на променливите

Позицијата на променливите во датотеката со податоци, нивните имиња и ознаки (label) треба да го одразуваат следново:

Имињата и ознаките на променливите

придонесуваат за структурирање на датотеката со податоци, овозможувајќи интегрирање на дел од документацијата во датотеката со податоци и помагајќи им на истражувачите да се ориентираат во структурата на збирките на податоци.

Имињата на променливите треба да бидат

кратки и треба да ги почитуваат вообичаените барања на стандарден софтвер, затоа што тие се користат како повикувачки кодови во работењето со

List of variables:

SEX
AGE

Sex of respondent
Age of respondent

MARITAL

COHAB

EDUCYRS

Marital status of respondent
Do you live together with a partner?

Education I – years (of fulltime) schooling



- Поврзаност помеѓу варијаблите

На пример:

сет на варијабли поврзани со истиот феномен (тие треба да бидат поставени заедно во базата на податоци, на пр. возраста на сите деца во едно домаќинство);

оригинални или изведени варијабли (изведените варијабли се создаваат од други варијабли, на пр. Возраста по години е повторно кодирани во пошироки категории)

- Врски до елементи на студијата и изворите на податоците

На пример, различни инструменти за мерење, различни делови на прашалникот, различни бази на податоци за извори, различни методи на набљудување итн.

- Типови променливи

- На пример, променливи за идентификација и други дополнителни променливи со различни специфични улоги, социо-демографски индикатори, генерирани променливи добиени со трансформација на оригинални информации, итн.

Именување на променливите

Базични правила за именување на променливите

‡ Започнување со буква

Не треба да се започнува со број, прашалници или извичник или посебни карактери како што се #, &, \$, @ (тие често се резервирани за специфични цели во софтверските апликации);

‡ **Имињата на променливите не можат да содржат празни места**

‡ Имињата на променливите исто така се користат како повикувачки кодови во софтверот. Поради оваа причина, променливите треба да бидат кратки и да ги почитуваат вообичаените барања на стандарден софтвер. Стандардот е да не се прават имиња на променливи подолги од осум карактери;

‡ **Да не се користат дијакритики (ознаки над или под буква) или национални специфични карактери;**

Постојат три основни пристапи за именување променливи:

‡ Користење на нумерички кодови што ја отсликуваат позицијата на променливита/варијаблата во системот (на пр. V001, V002, V003...);

‡ Користење на кодови кои се однесуваат на инструментот за истражување (на пр. Број на прашања во прашалник: Q1a, Q1b, Q2, Q3a...);

‡ Користење на менемонични имиња кои се однесуваат на содржината на променливите (на пр. BIRTH за годината на раѓање, AGE за возраста на испитаникот итн.) Зборот мнемонички значи „мемориска помош“.

Ознаки (label) на променливите

Овозможуваат краток опис на името на променливата.

Можат да бидат подолги од препорачаните осум карактери за имињата на променливата

Да се најде соодветен компромис помеѓу јасноста и големината на ознаката.

Треба да се има на ум:

-дека многу резултати од анализите се дадени во табели. Така, претерано долгите ознаки можат да резултираат во големи и непрактични табели.

-големината на ознаките може исто така да ја комплицира конверзијата на форматот.

-во некои резултати од анализите или по конверзијата на форматот, ќе се чува само дел од долгата ознака. Губењето на остатокот од ознаката на варијаблата може да ја направи истата неразбирлива.

Можат да вклучуваат **кратка** или **целосна верзија на прашањето** или **код за прашањето** во прашалникот ако имињата на променливите не се градат околу нив.

Пример:

⋮ Ознаката на променливата е прилагодена од бројот и формулацијата на прашањето од прашалникот:

„**Б10 - колку години имате?**“;

⋮ Дескриптивна ознака: „**Возраст на испитаникот**“;

⋮ Шематски: „**Испитаник: Возраст**“.

Најпосакуваниот јазик за именување на променливата е англискиот.

Ознаки за вредностите на променливите (Labels for variable values)

Променливите можат да имаат две или повеќе вредности

Понекогаш мора да се додели ознака на вредностите на променливите.

Не треба да се доделуваат ознаки на вредностите на континуирани променливи, како што се возраста (во години), висината (во метри) или тежината (во килограми), бидејќи нивните единици се општо познати.

На сите други променливи мора да се додадат.

Ова е разликата помеѓу номиналните и ординалните променливи. Номиналната променлива, како полот има две вредности, вообичаено претставени со 0 и 1 во податоците. Треба да се наведат ознаките „машки“/„женски“ на овие две вредности, така што сите истражувачи кои би можеле да ги користат податоците, би знаеле која вредност претставува кој пол. Истото важи и за редовните скали, на пример, скалата на согласување-несогласување со вредностите 1, 2, 3, 4 и 5, каде 1 претставува „целосно не се согласувам“ и 5 „целосно се согласувам“. Овие вредности мора да бидат означени за сите истражувачи и корисници на истражувањето да знаат каков степен на согласување/несогласување претставуваат броевите.

Два различни примери за именување на променливи

Име на променлива	Ознака на променлива
ISSP 2009 тематски модул варијабли	
V73	Q24a Опишете се: Јас работам напорно за да ги завршам моите дневни задачи
V74	Q24b Опишете се: Го постигнувам најдобро од моите способности
V75	Q24c Опишете се: Јас работам напорно за да ги одржувам своите перформанси на задачите
V76	Q25a Опишете се како 14,15,16 годишни: Многу се трудев да одам секој ден во школо
V77	Q25b Опишете се како 14,15,16 годишни: Го постигнувам најдоброто од моите способности
ISSP позадински променливи	
SEX	R: Пол
AGE	R: Возраст
MARITAL	R: Брачна состојба
CONAB	R: Стабилен животен партнер
EDUCYRS	R: Образование I: години на школување
DEGREE	R: Образование II: Највисоко ниво на образование
AR_DEGR	Образование специфично за земјата, Аргентина
AT_DEGR	Образование специфично за земјата, Австрија
AU_DEGR	Образование специфично за земјата, Австралија
BE_DEGR	Образование специфично за земјата, Белгија

Едноставни имиња на променливи

Првиот тематски дел од датотеката содржи едноставни имиња на варијабилните (нумерички кодови). Информациите за бројот на прашањата во заедничкиот меѓународен прашалник се вклучени во ознаките на променливите. Тоа обезбедува подобра ориентација за корисниците во датотеката со податоци. Бројките на прашања се проследени со буквално прашање, понекогаш скратено соодветно, за да останат разбирливи и да го следат правилото за кратко зачувување на ознаката на променливата.

Мнемонички имиња на променливите

Вториот дел содржи позадински променливи и користи менемонични имиња на променливите кои се однесуваат на нивната содржина. Позадинските променливи/варијабли не се директно поврзани со формулацијата на прашања во меѓународниот прашалник, туку се создадени од национални верзии на податоци. Нивните имиња се однесуваат на нивната содржина и истовремено на врските помеѓу нив (на пр., DEGREE = варијаблата за образование се претвора во меѓународно споредлива форма, XX_DEGR = варијабла за образование користејќи оригинални кодови специфични за една земја). Покрај тоа, множеството на мнемонички имиња на позадински променливи е стандардизирано во различни истражувања за ISSP, што овозможува полесно спојување на датотеките со податоци ISSP во текот на времето и изградбата на бази на податоци во временски серии.

2.3 Именување датотека и структура на папка (File naming and folder structure)

Управувањето со податоците на ефикасен и ефективен начин подразбира дека податоците треба да бидат добро организирани и документирани. Тоа подразбира дека е потребно да се направи точно и конзистентно именување на сите датотеки и понатамошно нивно правилно структурирање во една целина, поточно во соодветно именувана папка.



Стратегија на именување на датотека

Во утврдувањето на стратегијата за именување на датотеките треба да се земат предвид два важни елементи, а тоа се:

1.Името на датотеката се смета за главен идентификатор на истата

Правилното именување на датотеката овозможува полесно да се идентификува содржината, статусот и верзијата на таа датотека и помага во класификацијата и сортирањето на датотеките. Името на датотеката е посебно значајно во истражувањата каде соработуваат повеќе истражувачи или повеќе различни институции од причина што преку името на датотеката можат да се следат промените или уредувањето на истата.

2.Именувањето на датотеката треба да има конзистентност во подолг временски период и меѓу различните учесници инволвирани во истражувањето и собирањето на податоците

Именувањето на датотеките пожелно е да се практикува на систематски и конзистентен начин на сите произведени датотеки, а истово важи подеднакво и за квантитативните и за квалитативните истражувања. Ова е особено важно и претставува добра практика кога на едно истражување работат група на истражувачи.

Елементи кои треба да земат предвид кога се именува датотека:

- =Број на верзија;
- =Дата на креирање на датотеката (форматот би бил: година_месец_ден);
- =Име на креаторот;
- =Опис на содржината;
- =Име на истражувачки тим или одделение, а кои се поврзани со податоците;
- =Датум на објавување;
- =Датум на проект.

Examples ([UK Data Service, 2017b](#)) of useful file names are:

File name	Meaning
FG1_CONS_2010-02-12.rtf	Interview transcript of the first focus group with consumers, that took place on 12 February 2010.
Int024_AP_2008-06-05.doc	Interview with participant 024, interviewed by Anne Parsons on 5 June 2008.
BDHSurveyProcedures_00_04.pdf	Version 4 of the survey procedures for the British Dental Health Survey.

Најдобри практики за именување фајлови според Архивот на податоци на Обединетото Кралство:

- Да се креира кратко име кое носи значење и упатува на содржината на датотеката;
- Да се користи име на датотека кој ќе овозможи класифицирање на истата во соодветна папка;
- Да не се користат празни места, точки или специјални карактери (како на пр. & или ? или !);
- Да се користат цртичка (-) или долна црта (_) за раздвојување на деловите во името на фајлот;
- Да се избегнува долго именување на датотеката;
- Во името на датотеката, на крај да се остави слободно за место за три букви за екстензија на името заради додавање на специфичните кодови на форматот на датотеката (на пр. .doc; .xls; .mov; .tif);
- Да се додаде верзијата на името на датотеката каде е потребно.

Именување на датотеки со квалитативни податоци

При именувањето на датотеките што содржат квалитативни податоци постојат неколку важни аспекти (според Finnish Social Science Data Archive, 2016), а тоа се следниве:

Во ситуација кога има голем број датотеки кои се од различен вид, правилно е да се создаде еден посебен документ во којшто би се опишале конвенциите т.е. договорени или прифатени начини за именување на датотеките коишто се користат во истражувањето.

Името на датотеката треба да ги содржи основните информации за секоја ставка. Тоа се: индивидуално интервју, фокус група, фотографии и сл. Но, секако, основните и позадинските информации треба да бидат прикажани во посебен документ.

Конзистентност во именувањето на датотеките

Доследноста во именувањето на датотеките е од корист поради тоа што овозможува полесно идентификување на сите датотеки што се однесуваат на одреден случај (настан) (на пр. едно интервју или една фокус група). Освен тоа, името на датотеката би помогнало лесно да се поврзат сите датотеки што се однесуваат на одреден случај (настан) (на пр. аудио снимки, нивната транскрипција и фотографии кои се прибавено од интервјуираното лице).

Како добра практика се препорачува на почетокот на името на секоја датотека којашто е поврзана со ист случај (настан), т.е. во првиот дел на именувањето на датотеката, да се наведе „идентификувач на случајот (настанот)“. Во продолжение на името на датотеката треба да се содржи спецификата на датотеката, како на пр. дали е видео запис, аудио запис, транскрипција или фотографија.

Македонска верзија:

20190524_интервју1_аудио.wav

20190524_интервју1_транс.rtf

20190524_интервју1_фото.jpg

Англиска верзија:

(20190524_interview1_audio.wav)

(20190524_interview1_trans.rtf)

(20190524_interview1_image.jpg)

Структура на папка

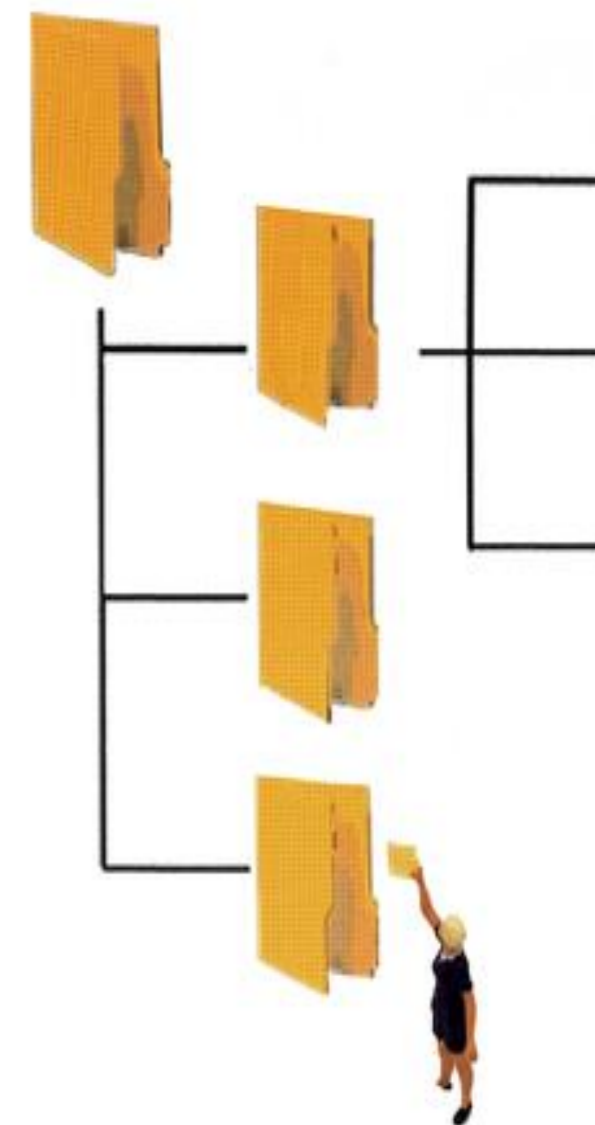
Организацијата и распоредувањето на датотеките во папки подразбира правилно структурирање на датотеките во соодветна папка.

Целта на структурирањето на папката е да се овозможи **полесно наоѓање** на датотеките и нивните верзии (доколку има повеќе верзии на една датотека), а посебно е корисно **кога постои соработка меѓу повеќе истражувачи** на еден истражувачки проект.

Начинот на кој се организираат датотеките е **условен од планирањето и организацијата** на самиот истражувачки проект или студија. Важно е да се внимава сите релевантни материјали поврзани со одредени податоци да бидат внесени во папки со податоци, вклучително и сите детални информации за начинот на собирање на податоците и процедурите за нивна обработка.

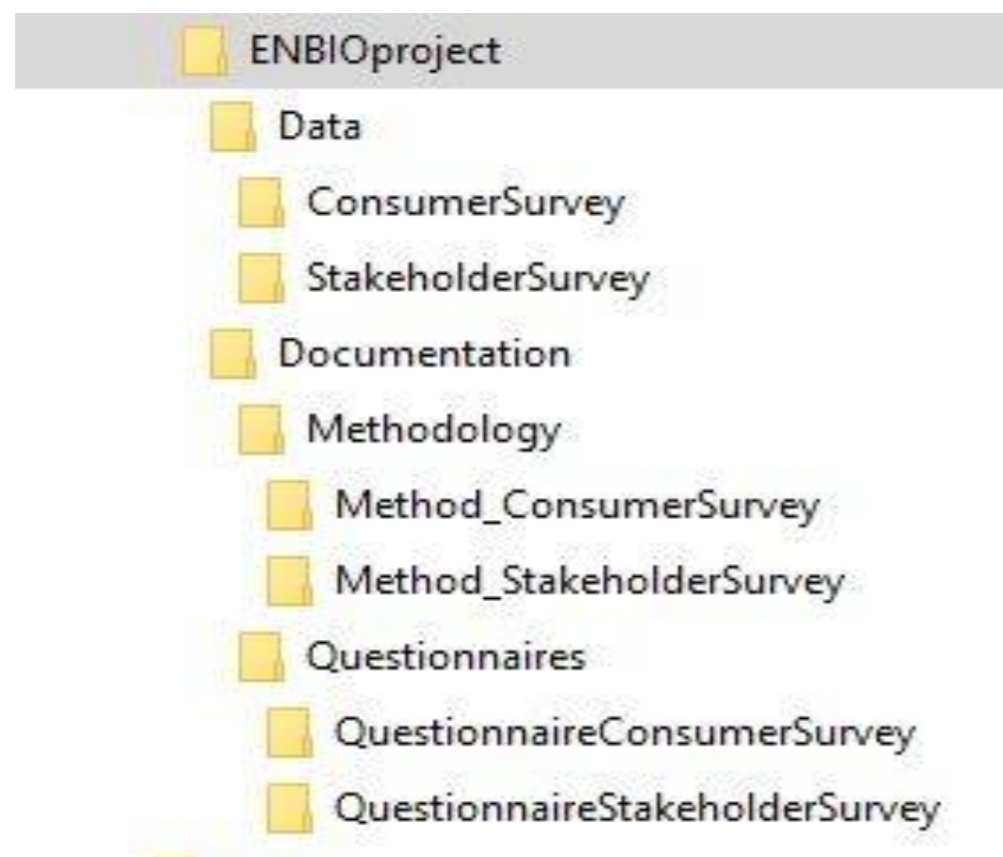
Структурирањето на датотеките се базира на хиерархискиот начин на организација на документацијата и податоците. При тоа, во зависност од потребите, може да се одбере повисока хиерархиска структура или пак порамна структура.

Во случај кога постојат неколку независни збирки на податоци, корисно е да се создадат посебни (одделни) папки на податоци за секоја збирка.

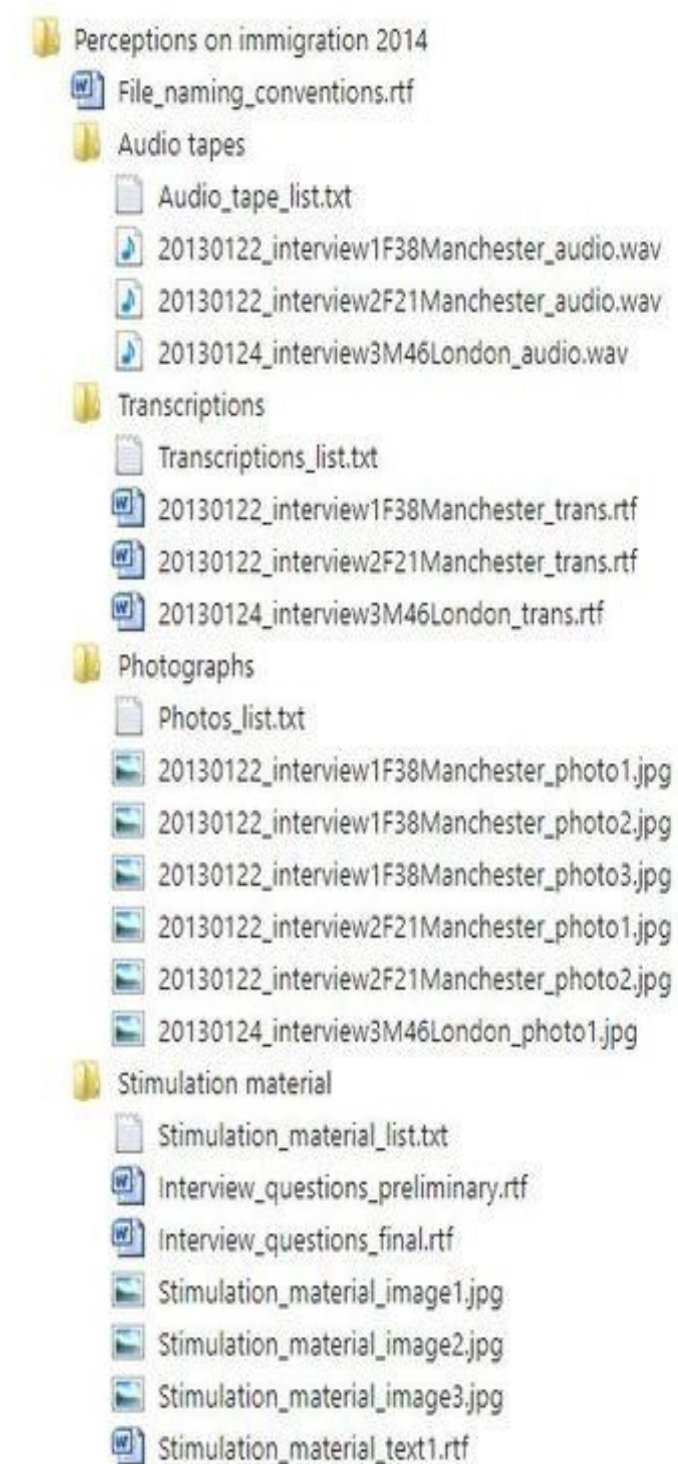


Пример за структура на папка со податоци од анкета

За оваа анкета, датотеките со податоци и документација се чуваат во посебни папки. Датотеките со податоци дополнително се организираат според типот на податоци, а потоа според истражувачката активност. Документарните досиеа се организираат и според видот на документацијата и истражувачката дејност. Тоа помага да се ограничи нивото на папки на три или четири и да нема повеќе од десет ставки на секоја листа.



Структура на папка со квалитативни податоци
Содржината на папката која се однесува на квалитативни податоци може да се состои од повеќе разновидни податоци, како што се аудиозаписи од интервјуа, транскрипти од интервјуа, стимулациски материјали кои се покажуваат на интервјуираните или пак фотографии. Во ваков случај, датотеките се материјали или документи кои се поврзани за еден случај (настан), на пр. спроведено интервју. Во примерот даден подолу, станува збор за интервјуа спроведени 22 и 24 јануари, 2013 година. Понатамошниот дел од името на датотеката ги прикажува спецификите на датотеката.



2.4 Документација и метаподатоци

Како да се започне?

Започнете рано!

Не ја оставајте документацијата за самиот крај на вашиот проект.

Запомнете да вклучите постапки за документација во планирањето за управување со податоци.

Размислете за информациите што се потребни за да ги разберете податоците. Што ќе им треба на другите истражувачи и повторни корисници за да ги разберат вашите податоци?

Создадете посебна датотека за документација за податоците што ги вклучуваат основните информации за истите.

Запомнете да ги организирате датотеките, за да постои врска помеѓу датотеката за документација и комплетите податоци.

Планирајте каде да ги складирате податоците по завршувањето на проектот. Складиштето веројатно следи специфичен стандард за метаподатоци што можете да го усвоите.

Документирајте постојано во текот на целиот проект. Документацијата за податоци дава контекстуални информации за вашата база на податоци. Ги специфицира општите цели и целите на оригиналниот проект и го открива појаснувачкиот материјал, вклучувајќи извор на податоци, методологија и процес на прибирање податоци, структура на податоци и технички информации.

Совет! Користете го англискиот јазик при документирањето.



2.4 Документацијата на ниво на проект ги објаснува целите на студијата, какви се истражувачките прашања / хипотези, кои методологии се користени, какви инструменти и мерки се користени итн.

Прашањата на кои треба да одговори вашата проектна документација се следниве:

1. За кои цели се колектирани податоците:

- Наслов на проектот;
- Поднаслов;
- Автор/и, креатор/и, на базата на податоци;
- Други соработници и нивните улоги (личност, истражувачка група или организација што учествувале во студијата и нивните улоги);
- Институцијата на автор/ите, креатор/ите;
- Финансиери;
- Броеви на грантови;
- Референци за поврзани проекти;
- Публикации од податоците.

2. Што содржи базата на податоци?

Опис на содржината на базата на податоци:

- Вид на податоци (интервјуа, слики, прашалници, итн.);
- Големина на датотеката (во бајти), формат на датотека со датотеки на податоци, и врски помеѓу датотеките;
- Опис на датотеката со податоци: верзија и издание, структура на базата на податоци, асоцијации, врски помеѓу датотеки, надворешни врски, формати, компатибилност.



3.Како се собрани податоците?

Методологијата и техниката што се користи при прибирање и креирање на податоците; Опис на сите извори од кои потекнуваат податоците (Што е предмет на проучување? На пример, периодични списанија, пакети со податоци создадени од други?), Заедно со објаснување за тоа како и зошто стигнал до сегашното место (провинција); Методи и начини на прибирање на податоци, на пример: Инструменти, хардвер и софтвер што се користат за прибирање на податоците; Методи за дигитализација или транскрипција; Протоколи за прибирање податоци; Дизајн и процедура на земање мостри; Целна популација, единици за набудување.

4.Кој ги собира податоците и кога?

Собирачите на податоци; Дата на собирање; Географска покриеност на податоците (на пр. Нација).

5.Како беа обработени податоците?

Опишете го вашиот проток на работа и специфичните алатки, инструменти, процедури, хардвер / софтвер или протоколи што можеби сте ги користеле за обработка на податоците, како што се: Уредување и прочистување на податоците, Кодирање и класификација на податоците

6.Кои можни манипулации се направени со податоците?

Опишете дали и како биле манипулирани или модифицирани податоците: Измени на податоци со текот на времето од првичното создавање и идентификувањето на различните верзии на базата на податоци; Други можни промени направени во податоците; Анонимизација; За временски серии или лонгитудинални истражувања: направени промени во методологијата, варијабилна содржина, текст со прашања, варијабилно означување, мерења или земање мостри.

7.Кои беа процедурите за обезбедување квалитет?

Опишете како е обезбеден квалитетот на податоците: Проверка на опремата и грешки во транскрипцијата; Контрола на квалитетот на материјалите; Проверка на интегритетот на податоците; Процедури за калибрација, Резолуција за фаќање на податоци и повторувања; Други постапки поврзани со квалитетот на податоците, како што се пондерирање, калибрација, причини за исчезнати вредности, проверки и корекции на транскрипциите, трансформации.

8.Како може да се пристапи до податоците?

Опишете ги условите за користење и пристапот на податоците: Каде може да се најдат податоците (во кое складиште за податоци); Постојани идентификатори; Услови за пристап како што е ембарго; Делови од податоците кои се ограничени или заштитени; Лиценца; Доверливост на податоците; Прашања околу авторски права и сопственост; Информации за цитација

Документација на ниво на податоци

Документацијата на ниво на податоци или на ниво на објекти обезбедува информации на ниво на индивидуални предмети, како што се слики или записници за интервју или променливи во базата на податоци. Можете да вметнете информации и за нивото на податоци. На пример, во интервјуа, најдобро е да се запишат контекстуалните и описните информации за секое интервју на почетокот на секоја датотека. И за квантитативните податоци променливите и имињата на вредностите на променливите можат да бидат вградени во самата датотека со податоци.

Документација на ниво на податоци, за квантитативни податоци

За квантитативен документ на податоци, се потребни следниве информации:

- Информации за датотеката со податоци

Тип на податоци, тип на датотека и формат, големина, скрипти за обработка на податоци.

- Информации за променливите во датотеката

Имињата, ознаките и описите на променливите, нивните вредности, описот на изведените променливи или, доколку е применливо, фреквенциите, основните непредвидени состојби итн. Треба да биде достапен и точниот оригинален формулар на прашањето. Ознаките на варијаблите треба:

Да бидат кратки, максимум 80 карактери;

Да се наведе единицата за мерење, кога е применлива;

Референцирање на бројот на прашања од анкетата или прашалникот, кога е применливо.



Пример за променлива и ознака на променлива :

променлива (variable): 'Q11eximp':

Ознака на променливата (variable Label) : „Q11 Колку е вежбањето важно за тебе?“

Ознаки на вредностите на променливата (Value label): 1. Сосема неважно, 2. Неважно, 3. Неутрално, 4. Важно, 5. Многу важно

Документација за квалитативни податоци на ниво на податоци

За креирање на документ со квалитативни податоци, потребно е да се наведат следниве информации:

Датотека со текстуални податоци (на пример, интервју)

Клучни информации за учесниците, како што се возраст, пол, занимање, локација, релевантни контекстуални информации;

За квалитативни збирки на податоци (на пример, збирки на слики или интервјуа) може да се направи список со податоци што даваат информации што овозможуваат идентификување и лоцирање на релевантни ставки во збирката на податоци. списокот би содржел клучни биографски карактеристики и тематски карактеристики на учесниците, како што се возраст, пол, занимање или локација, и одредени детали што се идентификатори за ставките на податоците;

Списокот на збирките на слики, треба да ги содржи главните

краткоструктури за секоја ставка, во овој случај тоа е фотографија;

Листата се креира според почетниот список со интервјуа, белешки од терен или други материјали обезбедени од депонентот на податоците.

Почетен структурен елемент на датотека од транскрипт

Датум на интервју: 08.02.2013 година [= 8 февруари 2013]

Интервјуер: Мет Милер

Псевдоним на интервјуираниот: Иан (не е вистинското име на интервјуираниот)

Занимање на интервјуираниот: новинар

Возраст на интервјуираниот: 32

Пол на интервјуираниот: машко

1	Interview videos 2012								
2	File name	Interview date	Interviewer	Interviewee's name	age	gender	occupation	Camera used for the video	Duration of the video
3	Peter_1.avi	12.4.2012	Matt Miller	Peter Herald	37	Male	Barkeeper	Panasonic HC-V10	2:45
4	Peter_2.avi	12.4.2012	Matt Miller	Peter Herald	37	Male	Barkeeper	Panasonic HC-V10	5:05
5	Lisa_1.avi	17.4.2012	Matt Miller	Lisa Smith	43	Female	Author	Canon XF305	10:12
6	Mary_1.avi	22.4.2012	Matt Miller	Mary Davies	42	Female	Teacher	Panasonic HC-V10	6:56
7	Pablo.mpg	24.4.2012	Matt Miller	Pablo Neftali	76	Male	Poet	Canon XF305	4:32



Обработка на истражувачките податоци

Јован Близнаковски / МК АИПОН, ИСППИ-УКИМ

*Вебинар: Управување со истражувачки податоци од општествените науки
26 октомври 2022*

DOI: 10.5281/zenodo.7250138

 cessda.eu

 [@CESSDA_Data](https://twitter.com/CESSDA_Data)



Licence: CC-BY 4.0

Вовед

- ◊ Под обработка на истражувачките податоци подразбираме активности на подготвување на податоците за анализа и споделување.
- ◊ Во текот на истражувачкиот проект, податоците можат да се манипулираат повеќе пати и важно е да се внимава на зачувување на нивната автентичност.
- ◊ Начинот на обработка на податоците влијае и на можностите за репликација на истражувањето.
- ◊ Обработката е клучна за целокупниот квалитет на податоците.

Вовед II (Што ќе погледнеме денес?)

- ◊ Намалување на грешките при 1) внесување на податоците и 2) кодирање (за квантитативни и квалитативни истражувања)
- ◊ Планирање на форматот на датотеки
- ◊ Осигурување на интегритетот и автентичноста на податоците
- ◊ Потреба за систематски пристап во осигурување на квалитетот за податоците



ПЛАН ЗА УПРАВУВАЊЕ СО ПОДАТОЦИ

Внесување на податоци: квантитативни истражувања

- ◊ Проверка на целосноста на податоците (дали постојат сите предвидени единици на обсервација, променливи/варијабли и сл.)
- ◊ Избегнување на рачно внесување
- ◊ Доколку внесувањето е рачно: двојно внесување и споредба
- ◊ Длабински проверки на 5-10% од единиците по случаен избор
- ◊ Логички проверки: опсег на вредности, екстремни вредности, однос помеѓу меѓусебно асоцирани променливи/варијабли, споредба со минати податоци
- ◊ Спроведување на автоматизирани проверки кога е потребно (пр. СПСС синтакси за проверки)
- ◊ Во случај на лоцирани грешки: бришење или корегирање

Внесување на податоци: квалитативни истражувања

- ◊ Спречување на грешки во транскрипцијата: квалитетни аудио датотеки
- ◊ Јасно дефинирање на методот на транскрипција (може да варира согласно дисциплината): фокус на содржината; фокус на што е кажано и како е кажано; фокус на емоционалниот и физичкиот јазик
- ◊ Рачно внесување или внесување со помош на софтвер за препознавање на говор?
- ◊ Јасно дефинирање на правилата за транскрипција (на пр. преку „шифрарник“): сите внесувачи мора да работат по исти правила
- ◊ Проверки на транскрипцијата
- ◊ Заштита на личните податоци на испитаниците (анонимизација)
- ◊ Формат на датотеки кои се погодни за анализа и за долгорочно зачувување на податоците

Кодирање на податоци: квантитативни истражувања

- ◊ Прашања од затворен тип: кодирање по однапред определен „шифрарник“ (автоматизирано во КАПИ истражувања)
- ◊ Прашања од отворен тип: независен процес на кодирање (посебна структура и усогласување на кодерите)
- ◊ Кодните шеми и нивните промени треба да бидат јасно документирани
- ◊ Кодните категории треба да бидат ексклузивни и кохерентни
- ◊ Кодирање на вредностите кои недостасуваат (missing): согласно причината зошто вредноста недостасува (пр. Б/О, одбивање, „не знам“, грешки во процесирање, неприменливо, итн)
- ◊ Јасно разликување меѓу повисоки и пониски кодни категории (пр. при кодирање на занимања)

F56	emprm14	Mother's employment status when respondent 14	F1	<ul style="list-style-type: none"> 1 Employee 2 Self-employed 3 Not working 4 Mother dead/absent 7 Refusal 8 Don't know 9 No answer 	<p>Ask all</p> <hr/> <p>Ask F57</p> <hr/> <p>Go to F58</p> <hr/> <p>Ask F57</p>
F57	occm14b	Mother's occupation when respondent 14	F2	<ul style="list-style-type: none"> 1 Professional and technical occupations 2 Higher administrator occupations 3 Clerical occupations 4 Sales occupations 5 Service occupations 6 Skilled worker 7 Semi-skilled worker 8 Unskilled worker 9 Farm worker 66 Not applicable 77 Refusal 88 Don't know 99 No answer 	<p>If codes 1, 2 or 8 at F56</p>

Пример од протоколот за податоци на ЕОИ
(ESS10 2020 Data Protocol, Edition 1.6 February 2022)

F56

emprm14

Mother's employment status when respondent 14

F1

- 1 Employee
- 2 Self-employed
- 3 Not working
- 4 Mother dead/absent
- 7 Refusal
- 8 Don't know
- 9 No answer

Ask all

Ask F57

Go to F58

Ask F57

F57

occm14b

Mother's occupation when respondent 14

F2

- 1 Professional and technical occupations
- 2 Higher administrator occupations
- 3 Clerical occupations
- 4 Sales occupations
- 5 Service occupations
- 6 Skilled worker
- 7 Semi-skilled worker
- 8 Unskilled worker
- 9 Farm worker
- 66 Not applicable
- 77 Refusal
- 88 Don't know
- 99 No answer

If codes 1, 2 or 8 at F56

F56

emprm14

Mother's employment status when respondent 14

F1

- 1 Employee
- 2 Self-employed
- 3 Not working
- 4 Mother dead/absent
- 7 Refusal
- 8 Don't know
- 9 No answer

Ask all

Ask F57

Go to F58

Ask F57

F57

occm14b

Mother's occupation when respondent 14

F2

- 1 Professional and technical occupations
- 2 Higher administrator occupations
- 3 Clerical occupations
- 4 Sales occupations
- 5 Service occupations
- 6 Skilled worker
- 7 Semi-skilled worker
- 8 Unskilled worker
- 9 Farm worker
- 66 Not applicable
- 77 Refusal
- 88 Don't know
- 99 No answer

If codes 1, 2 or 8 at F56

F56	emprm14	Mother's employment status when respondent 14	F1	<ul style="list-style-type: none"> 1 Employee 2 Self-employed 3 Not working 4 Mother dead/absent 7 Refusal 8 Don't know 9 No answer
-----	---------	---	----	--

Ask all

Ask F57

Go to F58

Ask F57

F57	occm14b	Mother's occupation when respondent 14	F2	<ul style="list-style-type: none"> 1 Professional and technical occupations 2 Higher administrator occupations 3 Clerical occupations 4 Sales occupations 5 Service occupations 6 Skilled worker 7 Semi-skilled worker 8 Unskilled worker 9 Farm worker 66 Not applicable 77 Refusal 88 Don't know 99 No answer
-----	---------	--	----	--

If codes 1, 2 or 8 at F56

F56	emprm14	Mother's employment status when respondent 14	F1	<ul style="list-style-type: none"> 1 Employee 2 Self-employed 3 Not working 4 Mother dead/absent 7 Refusal 8 Don't know 9 No answer
-----	---------	---	----	--

Ask all

Ask F57

Go to F58

Ask F57

F57	occm14b	Mother's occupation when respondent 14	F2
-----	---------	--	----

- 1 Professional and technical occupations
- 2 Higher administrator occupations
- 3 Clerical occupations
- 4 Sales occupations
- 5 Service occupations
- 6 Skilled worker
- 7 Semi-skilled worker
- 8 Unskilled worker
- 9 Farm worker
- 66 Not applicable
- 77 Refusal
- 88 Don't know
- 99 No answer

If codes 1, 2 or 8 at F56

F56	emprm14	Mother's employment status when respondent 14	F1	<ul style="list-style-type: none"> 1 Employee 2 Self-employed 3 Not working 4 Mother dead/absent 7 Refusal 8 Don't know 9 No answer
-----	---------	---	----	--

Ask all

Ask F57

Go to F58

Ask F57

F57	occm14b	Mother's occupation when respondent 14	F2	<ul style="list-style-type: none"> 1 Professional and technical occupations 2 Higher administrator occupations 3 Clerical occupations 4 Sales occupations 5 Service occupations 6 Skilled worker 7 Semi-skilled worker 8 Unskilled worker 9 Farm worker 66 Not applicable 77 Refusal 88 Don't know 99 No answer
-----	---------	--	----	--

If codes 1, 2 or 8 at F56

F56	emprm14	Mother's employment status when respondent 14	F1	<ul style="list-style-type: none"> 1 Employee 2 Self-employed 3 Not working 4 Mother dead/absent 7 Refusal 8 Don't know 9 No answer
-----	---------	---	----	--

Ask all

Ask F57

Go to F58

Ask F57

F57	occm14b	Mother's occupation when respondent 14	F2	<ul style="list-style-type: none"> 1 Professional and technical occupations 2 Higher administrator occupations 3 Clerical occupations 4 Sales occupations 5 Service occupations 6 Skilled worker 7 Semi-skilled worker 8 Unskilled worker 9 Farm worker 66 Not applicable 77 Refusal 88 Don't know 99 No answer
-----	---------	--	----	--

If codes 1, 2 or 8 at F56

MAJOR, SUB-MAJOR, MINOR AND UNIT GROUPS

1 Managers

11 Chief Executives, Senior Officials and Legislators

111 Legislators and Senior Officials

1111 Legislators

1112 Senior Government Officials

1113 Traditional Chiefs and Heads of Villages

1114 Senior Officials of Special-interest Organizations

112 Managing Directors and Chief Executives

1120 Managing Directors and Chief Executives

12 Administrative and Commercial Managers

121 Business Services and Administration Managers

1211 Finance Managers

1212 Human Resource Managers

1213 Policy and Planning Managers

1219 Business Services and Administration Managers Not Elsewhere Classified

122 Sales, Marketing and Development Managers

1221 Sales and Marketing Managers

1222 Advertising and Public Relations Managers

1223 Research and Development Managers

Пример од Меѓународната стандардизирана класификација на занимања
(INTERNATIONAL STANDARD CLASSIFICATION OF OCCUPATIONS (ISCO-08))

MAJOR, SUB-MAJOR, MINOR AND UNIT GROUPS

1 Managers

11 Chief Executives, Senior Officials and Legislators

111 Legislators and Senior Officials

1111 Legislators

1112 Senior Government Officials

1113 Traditional Chiefs and Heads of Villages

1114 Senior Officials of Special-interest Organizations

112 Managing Directors and Chief Executives

1120 Managing Directors and Chief Executives

12 Administrative and Commercial Managers

121 Business Services and Administration Managers

1211 Finance Managers

1212 Human Resource Managers

1213 Policy and Planning Managers

1219 Business Services and Administration Managers Not Elsewhere Classified

122 Sales, Marketing and Development Managers

1221 Sales and Marketing Managers

1222 Advertising and Public Relations Managers

1223 Research and Development Managers

Највисока-најопшта категорија
(една цифра)

MAJOR, SUB-MAJOR, MINOR AND UNIT GROUPS

1 Managers

11 Chief Executives, Senior Officials and Legislators

- 111 Legislators and Senior Officials
 - 1111 Legislators
 - 1112 Senior Government Officials
 - 1113 Traditional Chiefs and Heads of Villages
 - 1114 Senior Officials of Special-interest Organizations
- 112 Managing Directors and Chief Executives
 - 1120 Managing Directors and Chief Executives

12 Administrative and Commercial Managers

- 121 Business Services and Administration Managers
 - 1211 Finance Managers
 - 1212 Human Resource Managers
 - 1213 Policy and Planning Managers
 - 1219 Business Services and Administration Managers Not Elsewhere Classified
- 122 Sales, Marketing and Development Managers
 - 1221 Sales and Marketing Managers
 - 1222 Advertising and Public Relations Managers
 - 1223 Research and Development Managers

MAJOR, SUB-MAJOR, MINOR AND UNIT GROUPS

1 Managers

11 Chief Executives, Senior Officials and Legislators

111 Legislators and Senior Officials

1111 Legislators

1112 Senior Government Officials

1113 Traditional Chiefs and Heads of Villages

1114 Senior Officials of Special-interest Organizations

112 Managing Directors and Chief Executives

1120 Managing Directors and Chief Executives

12 Administrative and Commercial Managers

121 Business Services and Administration Managers

1211 Finance Managers

1212 Human Resource Managers

1213 Policy and Planning Managers

1219 Business Services and Administration Managers Not Elsewhere Classified

122 Sales, Marketing and Development Managers

1221 Sales and Marketing Managers

1222 Advertising and Public Relations Managers

1223 Research and Development Managers

MAJOR, SUB-MAJOR, MINOR AND UNIT GROUPS

1 Managers

11 Chief Executives, Senior Officials and Legislators

111 Legislators and Senior Officials

1111 Legislators

1112 Senior Government Officials

1113 Traditional Chiefs and Heads of Villages

1114 Senior Officials of Special-interest Organizations

112 Managing Directors and Chief Executives

1120 Managing Directors and Chief Executives

12 Administrative and Commercial Managers

121 Business Services and Administration Managers

1211 Finance Managers

1212 Human Resource Managers

1213 Policy and Planning Managers

1219 Business Services and Administration Managers Not Elsewhere Classified

122 Sales, Marketing and Development Managers

1221 Sales and Marketing Managers

1222 Advertising and Public Relations Managers

1223 Research and Development Managers

Најниска-најконкретна категорија
(четири цифри)

Кодирање на податоци: квантитативни истражувања II

- ◊ Променлива/варијабла за идентификација (оригинален код за секоја единица)
- ◊ Оригиналните информации треба да бидат зачувани (пр. отворени прашања, возраст, приход)
- ◊ Проверка на податоците кои се преземени вербатим со цел заштита на личните податоци на испитаниците
- ◊ Проверка преку комбинирање на промеливи/варијабли со цел заштита на личните податоци на испитаниците
- ◊ При креирање на пондери (weights): причината и калкулациите треба да бидат јасно наведени во документацијата поврзана со истражувањето
- ◊ Обучување на кодерите (анкетари) со цел намалување на нивната меѓусебна варијација: читање на прашања, заведување на одговори, итн.

Кодирање на податоци: квалитативни истражувања

- ◊ Под кодирање во квалитативни истражувања го подразбираме процесот на идентификување на концепти и нивните односи во текст, слика, фотографија... Ваквото кодирање не е само обележување туку и двострано поврзување на податоците со истражувачката идеја.
- ◊ Два основни подвида: кодирање управувано од концепт и кодирање управувано од податоци.

Сиров податок (транскрипција)	Преелиминарни кодови	Конечен код
<p>Како што приближувам кон пензионерските денови, сакам што поскоро да дојдат. Немам сè уште ни 55 но би дал сè да се пензионирам веднаш. Но, имам хипотека за исплаќање и уште за заштедување пред воопшто да помислам на пензија! Редовно играм лотарија во надеж за освојување на сите тие милиони. Сè уште немам среќа во врска со пензионирањето!</p>	<p>Наближување до пензија</p> <p>Финансиски потешкотии</p> <p>Надеж за порано пензионирање</p>	<p>АНКСИОЗНОСТ ЗА ПЕНЗИОНИРАЊЕ</p> <p>Извор: Saldaña, 2013.</p>

Кодирање на квалитативни податоци: совети

- ◊ Значењето на кодовите треба јасно да се документира: име/опис на кодот, име на кодерот, датум на кодирање и промени, дефиниција на кодот и поврзување со концепт, информации за односот на конкретниот код-концепт со други кодови-концепти кои се користат во анализата...
- ◊ Намалување на варијансата меѓу кодерите: коден дизајн и насоки („шифрарник“), проверка на квалитетот на кодирањето (од независен истражувач, меѓусебно меѓу кодерите)
- ◊ Повторно, адекватна обука за кодерите е од големо значење!

Формати на датотеки I

- ◊ Постојано се појавуваат нови софтверски решенија: не е јасно колку типовите на датотеки кои ги користиме денес ќе бидат актуелни и во иднина!
- ◊ Краткотрајна обработка на податоци
 - ◊ Изборот на формат на датотека треба да биде поврзан со софтверот кој сакаме да го користиме за анализа
 - ◊ Квантитативна анализа: MS Excel (*.xls), SPSS (*.sav, *.por), R and STATA (*.dta)
 - ◊ Квалитативна анализа: *.docx, *.rtf, *.pdf, *.mp3, *.wav, *.jpeg (NVivo, ATLAS-ti, и MAXQDA)

Формати на датотеки II

- ◊ Долготрајно зачувување на податоци
 - ◊ Квантитативни податоци: ASCII формати (едноставен текст)
 - ◊ Фиксен формат: *.csv (податоци во редови и колони)
 - ◊ Слободен формат: *.txt., *.dat, *.asc (податоци каде колоните се одделени со знак)

 baza - Notepad

File Edit Format View Help

id,age,gender,education,occupation

001,22,2,3,7

002,57,2,4,2

003,34,1,5,3



	A	B	C	D	E	F
1	id	age	gender	education	occupation	
2	1	22	2	3	7	
3	2	57	2	4	2	
4	3	34	1	5	3	
5						
6						

Формати на датотеки II

- ◊ Долготрајно зачувување на податоци
 - ◊ Квантитативни податоци: ASCII формати (едноставен текст)
 - ◊ Фиксен формат: *.csv (податоци во редови и колони)
 - ◊ Слободен формат: *.txt., *.dat, *.asc (податоци каде колоните се одделени со знак)
 - ◊ Квалитативни податоци
 - ◊ Текст: PDF/A
 - ◊ Мета-податоци: XMLP и ASCII
 - ◊ Видео: MXF (Material eXchange Format) and JPEG2000
 - ◊ Аудио: WAV
 - ◊ Фотографии, слики: JPEG

Конверзија на датотеки кон долгорочно зачувување

- ◊ Со цел долгорочно искористување и зачувување на податоците, потребно е истите да бидат „снимени“ во формати кои овозможуваат долгорочна употреба.
- ◊ Сегашните софтвери овозможуваат лесна конверзија од краткорочен кон долгорочен формат и обратно.
- ◊ Сепак, секогаш постои опасност од „губење“ на податоци при машинска конверзија:
 - ◊ Квантитативните бази можат да изгубат делови, податоци за карактерот на варијаблите, децимални броеви.
 - ◊ Текстовите можат да изгубат делови кои се уредени (пр. обележани делови, нагласени делови, заглавја, итн.).
 - ◊ Сликите, видеата и звукот можат да изгубат на квалитет.
- ◊ Следствено, податоците треба да бидат проверени по конверзијата и тоа од истражувач кој добро ги познава.

Зачувување на автентичноста

- ◊ Во текот на обработката на податоците, можат да бидат креирани неколку верзии на датотеката со податоци.
- ◊ Со цел зачувување на автентичноста, како и да се овозможи преглед на правените промени, потребно е сите индивидуални верзии внимателно да се евидентираат.
- ◊ Со цел зачувување на автентичноста, потребно е и податоците да се заштитат од неовластена модификација.

Пример на табела за контрола на верзии

Title:	Vision screening tests in Essex nurseries
File Name:	VisionScreenResults_00_05
Description:	Results data of 120 Vision Screen Tests carried out in 5 nurseries in Essex during June 2007
Created By:	Chris Wilkinson
Maintained By:	Sally Watsley
Created:	04/07/2007
Last Modified:	25/11/2007
Based on:	VisionScreenDatabaseDesign_02_00

Version	Responsible	Notes	Last amended
00_05	Sally Watsley	Version 00_03 and 00_04 compared and merged by SW	25/11/2007
00_04	Vani Yussu	Entries checked by VY, independent from SK	17/10/2007
00_03	Steve Knight	Entries checked by SK	29/07/2007
00_02	Karin Mills	Test results 81-120 entered	05/07/2007
00_01	Karin Mills	Test results 1-80 entered	04/07/2007



Зачувување на истражувачките податоци

Јован Близнаковски / МК АИПОН, ИСППИ-УКИМ

*Вебинар: Управување со истражувачки податоци од општествените науки
26 октомври 2022*


DOI: 10.5281/zenodo.7250138

 cessda.eu

 @CESSDA_Data



Licence: CC-BY 4.0

●●○○ EE 4G 11:41 

Cancel Edit post Save



Andrew Penson

 Friends of friends ▾

****Please SHARE****Lost bag containing all my research notes on laptop and paper. I can't complete my research masters degree without this - eighteen months worth of daily research. Left it on a train arriving at Charing Cross Saturday 8th July 08:50 ish which became the 09:06 to Sevenoaks. Black and grey with red "Swiss" features. I'm ruined if I don't get it back, please share as widely as possible! Thank you.
andrewpenson@hotmail.co.uk

Add to your post



Вовед

- ◇ „Без податоци – нема истражување!“
- ◇ Важно е да се преземат соодветни мерки за чување на податоците со цел да не се изгубат како и да се спречи неовластена манипулација.
- ◇ Второто е особено важно во контекст на управување со лични сензитивни податоци на испитаниците.

Вовед II (Што ќе погледнеме денес?)

- ◊ Различни решенија за складирање, нивни предности и слабости
- ◊ Креирање на стратегија за складирање, согласно карактерот на истражувачкиот проект
- ◊ Резервни копии и стратегии за справување со евентуални „катастрофи“
- ◊ Можности за заштита на податоците од неовластена манипулација



ПЛАН ЗА УПРАВУВАЊЕ СО ПОДАТОЦИ

Решенија за складирање

- ◆ Пример: терабајти на снимени видео-интервјуа и илјадници страници транскрипти во рамки на меѓународен истражувачки проект

Тип на податоци	Потреби за складирање	Решенија за складирање
Лични податоци (потребно е да се имплементираат силни мерки за заштита)	Голем капацитет за складирање Потребен е далечински пристап до видеата и транскриптите Истражувачите треба да бидат во можност истовремено да работат со податоците	Податоците се пренесуваат само во енкриптирана форма Далечинскиот пристап се овозможува преку „облак“ Главните копии од видеата и транскриптите се енкриптирани и „снимени“ на „облак“, преносен хард-диск и усб читач Резервните копии (хард-диск, усб читач) се лоцирани на различни и безбедни локации Формуларите за согласност и клучевите за енкрипција се чуваат во посебен сеф

Решенија за складирање: преносни уреди

Предности	Слабости	Мерки на претпазливост за чувствителни лични податоци
<p>Овозможуваат лесен пренос на податоците без користење на интернет</p> <p>Можат да бидат особено корисни при работа на терен</p> <p>Ниски трошоци</p>	<p>Лесно се губат и оштетуваат (непотребен безбедносен ризик)</p> <p>Недоволно сигурно за долгорочно чување на главните копии</p> <p>Можни конфузии околу верзијата на податоците</p>	<p>Енкриптирање и заштита со силни лозинки</p>

Решенија за складирање: „облак“

Предности	Слабости	Мерки на претпазливост за чувствителни лични податоци
<p>Автоматски резервни копии</p> <p>Автоматско впишување на различни верзии на податоци</p>	<p>Можни проблеми со безбедност на податоците</p> <p>Недоволна контрола во однос на складирањето на податоците и резервните копии</p> <p>Бесплатните даватели на „облак“ услуги можат да бараат права на користење на содржината која вие ја управувате за свои цели</p> <p>Податоците може да се изгубат ако вашата „облак“ сметка е суспендирана или доколку давателот на услуга престане со работа</p>	<p>Енкриптирање на податоците (особено важно за да не се дојде до конфликт со регулативата за заштита на податоци)</p>

Решенија за складирање: локално складирање

Предности	Слабости	Мерки на претпазливост за чувствителни лични податоци
<p>Целосна контрола над податоците</p> <p>Полесно за заштита од неавторизиран пристап</p>	<p>Складирањето на само еден уред ги прави податоците „ранливи“</p> <p>Пристапот е ограничен на лицето кое го користи уредот</p>	<p>Заштита на уредот со силна лозинка и енкрипција на хард-дискот</p>

Решенија за складирање: вмрежени дискови

Предности	Слабости	Мерки на претпазливост за чувствителни лични податоци
<p>Податоците се централно складирани</p> <p>Споделен пристап до одредена мера</p> <p>Резервните копии се централно управувани и можат да бидат автоматизирани</p>	<p>Посериозни мерки се потребни за спречување на неовластен пристап или случајно бришење/манипулација на податоците</p> <p>Пристапот за надворешни соработници може да биде отежнат</p> <p>Повисоки трошоци</p>	<p>Комбинација на мерки (физички, силни лозинки, енкрипција)</p>

Решенија за складирање: различни медиуми

Медиум	Предност	Слабост
CD, DVD (оптички)	Лесен пренос Ниски трошоци	Лесно се оштетуваат Лесно се губат Чести грешки во читање и запишување Краток рок на траење Мал капацитет
Хард-диск (HDD) (магнетен)	Пониски трошоци во споредба со SSD дискови Висок капацитет на складирање	Можност за физичка деградација Лесно се оштетува
Преносни SSD дискови	Цврсти Со поголемо траење	Податоците тешко можат да се „спасат“ доколку дискот се расипе Повисоки трошоци Помал капацитет од HDD

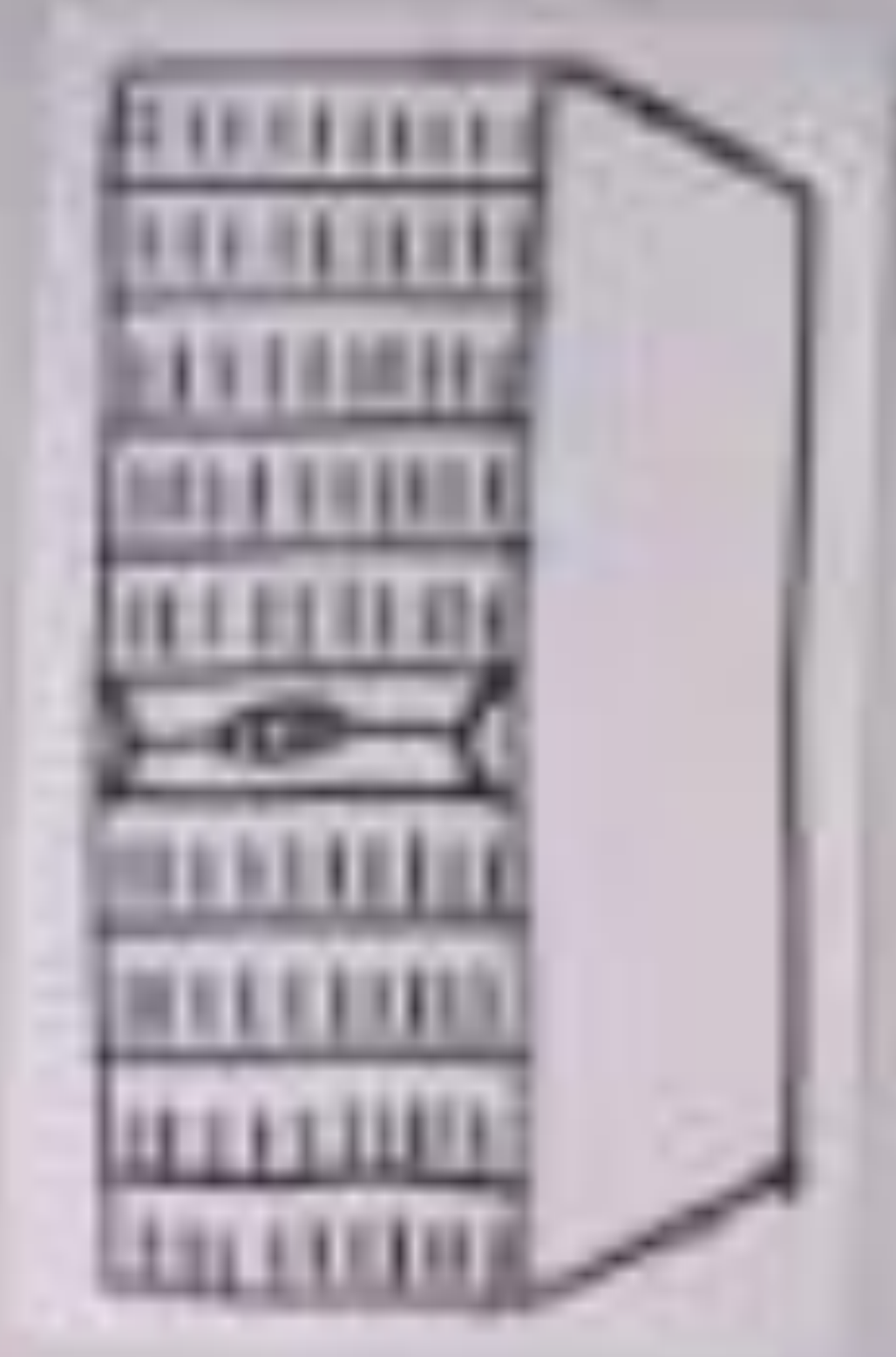
Стратегија за складирање (совети)

- ◊ Стратегијата треба да биде прилагодена на видот на истражувачкиот проект
 - ◊ Какви податоци користиме? Колкав капацитет е потребен?
 - ◊ Кој треба да има пристап?
 - ◊ Дали податоците содржат чувствителни лични податоци?
- ◊ Користете барем два различни медиума за складирање.
- ◊ Периодично менувајте ги медиумите за складирање (2-5 години).
- ◊ Спроведувајте периодични проверки на интегритетот на податоците.

◊ Извор: UK Data Service (2017)

Резервни копии

- ◇ Креирање на барем три резервни копии
 - ◇ Користење на барем два различни медиуми
 - ◇ Чување на различните медиуми на различни локации
 - ◇ Редовни проверки на интегритетот на датотеките
-
- ◇ Стратегија за резервни копии во 10 чекори: CESSDA Data Management Expert Guide (стр. 93-94)



Безбедност

- ◇ „Силни“ лозинки: минимум 15 карактери, различни видови карактери (букви, бројки, знаци)
- ◇ Безбедно чување на лозинките: сеф, избегнување да се користат лозинките на отворени wifi мрежи
- ◇ Енкрипција: процес на енкридање на дигиталните информации на начин на кој само авторизирани лица ќе имаат пристап (клучеви)
- ◇ Особено важно за податоци кои се закачуваат на „облак“, но и за податоци кои се пренесуваат на преносни медиуми за чување на податоци



Прашања? Мислења?

 cessda.eu

 @CESSDA_Data

  Licence: CC-BY 4.0



Етички правила Заштита на правата на испитаниците

Анета Цекиќ / МК АИПОН, ИСППИ-УКИМ

*Вебинар: Управување со истражувачки податоци од општествените науки,
26 Октомври 2022*

DOI: 10.5281/zenodo.7250138

 cessda.eu

 [@CESSDA_Data](https://twitter.com/CESSDA_Data)



Licence: CC-BY 4.0

Што опфаќа овој дел:

- ◊ Правни и етички аспекти при собирање и обработка на податоци
- ◊ Влијанието на GDPR врз собирањето/обработката на податоци
- ◊ Мерки кои истражувачите може да ги преземат за да се осигураат дека ги почитуваат правните стандарди и етичноста во научно-истражувачката работа
- ◊ Како да се пополни Планот за управување со податоци во делот кој се однесува на Заштита

Етички стандарди

1. Почитување на научните стандарди
2. Согласност / почитување на законите
3. Да не се нанесе социјална или лична штета на испитаниците / учесниците во истражувањето

Професионални и етички упатства (RESPECT проект)

1. Почитување на научните стандарди

- ◊ Истражувачите треба да да го земат предвид целокупниот доказен материјал и да ги презентираат истражувачките резултати без изоставање на делови, без погрешна интерпретација или измама / доведување во заблуда.
- ◊ При планирање на истражувањето и формулирање на истражувачките прашања, дизајнот на истражувачките методи и инструменти да не упатуваат или пристрасно да ги насочуваат резултатите (со изборот на прашањата или истражувачки постапки)

2. Согласност со законите

- ◊ Истражувачите треба да се осигураат дека ги имаат предвид сите национални закони или меѓународни правила кои може да се применливи на нивниот истражувачки проект.
- ◊ При меѓународни проекти- повеќе национални законодавства.
- ◊ Особено внимание се посветува на **заштитата на личните податоци и правата од интелектуална сопственост (авторските права).**

3. Да не се нанесе социјална или лична штета

- ◆ Истражувачите треба да се стремат да избегнат или да ја минимизираат евентуална штета врз одредени групи или поединци. Истражувањето треба да се дизајнира одговорно, земајќи ги целосно во предвид карактеристиките на учесниците.
- ◆ Да се осигура дека нивното учество е **доброволно** и се базира на **информирана согласност**.

Архивирање и објавување на лични податоци

- ◊ Доколку собирате податоци кои овозможуваат да се идентификуваат лица, тогаш тие податоци се сметаат за лични податоци.
- ◊ Општата регулатива за заштита на податоци- General Data Protection Regulation (GDPR, European Union, 2016a) личните податоци ги дефинира како: која било информација која се однесува на идентификувано лице или физичко лице кое може да се идентификува, познато како „субјект на податоци“.
- ◊ Физичко лице кое може да се идентификува е лице кое може да се идентификува или директно или индиректно преку реферирање на некој идентификатор како *име, идентификациски број, податоци за локација, онлајн идентификатори* и преку еден или повеќе *фактори специфични за физичките, психолошките, генетски, ментални, економски, културни или социјални обележја/идентитети* на тоа физичко лице. Личните податоци може да вклучуваат голем број информации, како имиња, адреси, телефонски броеви или IP адреси.

Чувствителни лични податоци

Одредени лични податоци се сметаат за особено чувствителни и бараат специфична заштита кога откриваат информации кои може да создадат важни ризици за фундаменталните права и слободи на лицата. Тоа се, на пример, религиозната припадност, сексуалната ориентација или расното или етничкото потекло.

Во рамки на GDPR следните категории се дефинирани како **„посебни категории на лични податоци“**:

- » Расно или етничко потекло;
- » Политички мислења;
- » Религиозни или верски убедувања;
- » Членство во синдикати;
- » Генетски податоци;
- » Биометриски податоци;
- » Здравствени податоци;
- » Податоци за сексуалниот живот или сексуалната ориентација.

Процес на етичка проценка

- ◆ Принципите на добрата истражувачка пракса упатуваат на тоа да се земат во предвид последиците од Вашето истражување и интересите на учесниците.
- ◆ Етичкото оценување преку Комитети за истражувачка етика (КИЕ) е потребно секогаш кога се собираат (чувствителни) лични податоци.
- ◆ Улогата на КИЕ е да се заштити безбедноста, правата и добросостојбата на учесниците во истражувањето и да се промовира етички издржано истражување.

Етичка само-проценка

Прашање 1: **Цели и методи на проектот**

Дали целите и методите на проектот се во спротивност со општо прифатените вредности? Дали спроведувањето на проектот вклучува ризик или повреда на луѓето, животните, или природата во степен кој не е занемарлив? Ако е така, дали личностите кои се вклучени се свесни за ризикот?

Прашање 2: **Истражувањето вклучува лица кои може да се идентификуваат**

Дали Вашето истражување предвидува собирање и обработка на лични податоци? Доколку да, дали ќе се обезбеди информирана согласност од учесниците? Дали личните податоци ќе бидат доволно анонимизирани со цел да се обезбеди адекватна заштита на приватноста?

Прашање 3: **Укажувачи (Whistle-blowing)**

Доколку член на тимот почне да се сомнева во етичките аспекти на проектот, дали ќе му/и биде дозволено да ги претстави своите грижи на независно консултативно тело? Дали оваа можност се знае однапред?

Етички упатства при истражување, Национален етички комитет на Норвешка (n.d.)

Етичка проценка

Задолжителна во Хоризонт 2020 и Хоризонт Европа- преку само-проценка и евалуација од страна на експерти (степенот на евалуација зависи од природата на проектот)

Разлики меѓу земјите- битни за меѓународни истражувања

Обработка на лични податоци

- ◊ Од 25ти мај 2018, Општата регулатива за заштита на податоци (General Data Protection Regulation (GDPR, European Union, 2016a)) се применува на сите истражувачи во Европската економска област кои собираат лични податоци на граѓаните на ЕУ.
- ◊ GDPR се однесува на податоци за живи лица. Податоци кои не се сметаат за лични податоци не спаѓаат под овие закони, иако сепак може да има етички причини овие информации да се заштитат.

6 начела за заштита на лични податоци (GDPR)

I. Обработувајте ги податоците во согласност со законите, фер и транспарентно

Учесникот е информиран за тоа како ќе се постапува со податоците и обработката на податоците ќе треба да биде направена на тој начин.

II. Држете се до почетната намена/цел

Податоците треба да се собираат за конкретна, експлицитна и легитимна цел и да не се обработуваат дополнително на начин кој не е компатибилен со таа цел.

III. Минимизирајте ја големината на податоците

Личните податоци кои ќе се собираат треба да бидат адекватни, релевантни и ограничени на она што е потребно.

IV. Запазете ја точноста

Личните податоци треба да бидат точни, и онаму каде што е неопходно, да се обновуваат. Треба да се преземат сите разумни чекори за да се осигура дека неточните податоци ќе се избришат или корегираат без одложување.

V. Отстранете ги податоците кои не се користат

Личните податоци треба да се чуваат во форма која овозможува идентификување на субјектите на податоците не подолго од што е потребно за целите за кои личните податоци се обработуваат.

VI. Обезбедете интегритет и доверливост на податоците

Личните податоци се обработуваат на начин кој обезбедува адекватна безбедност на личните податоци, вклучувајќи заштита против неавторизирана или незаконска обработка, како и заштита од случајно губење, уништување или штета, користејќи соодветни технички или организациски мерки.

Исклучок за (научни) истражувања

- ◊ Обработка со цел архивирање во јавен интерес, научни или историски истражувачки цели или статистички цели, ќе биде предмет на соодветна заштита, во согласност со оваа Регулатива, со цел заштита на правата и слободите на субјектите на податоци. Тие мерки треба да осигурат дека се преземени технички и организациски мерки особено со цел да се обезбеди почитување на принципот на минимизирање на податоците. Тие мерки може да вклучат псевдонимизација доколку оваа цел може да биде исполнета на тој начин. Таму каде што овие мерки може да бидат обезбедени со понатамошна обработка која не овозможува или повеќе нема да овозможи идентификација на субјектите на податоци, тие цели може да се исполнат на тој начин. | **General Data Protection Regulation, Член 89**
- ◊ Во пракса- **принцип II и V се помалку стриктни**. Понатамошна обработка на лични податоци со цел архивирање, научни или историски истражувачки цели и статистички цели не се смета за некомпатибилно со почетната цел за собирање на податоците, дури и кога оваа цел не е експлицитно наведена порано.
- ◊ Исто така, личните податоци може да се чуваат подолго за истите цели. Во сите случаи, треба да бидат преземени соодветни технички и организациски мерки за да се заштитат правата и слободите на учесниците во истражувањата, како што се минимизирање на податоците и псевдонимизација.

Правна основа за обработка на ЛП

GDPR признава 6 основи за собирање/обработка на лични податоци:

- » **согласност на субјектите на податоци**
- » неопходно е за да се изврши некаков договор
- » законска обврска за контролорот на податоци
- » неопходно е за да се заштитат виталните интереси на субјектите на податоци
- » **се спроведува во јавен интерес или преку овластен орган на државната власт (јавна задача)**
- » **легитимен интерес кој се спроведува од контролорот на податоци.**

За секој истражувачки проект, доколку се собираат и обработуваат лични податоци, треба да се одлучи за **најсоодветната правна основа**, истата да се забележи (и потоа не треба да се менува).

GDPR во пракса

Кога ќе го стартувате Вашиот проект:

- » утврдете кој ќе биде контролор на податоците (веројатно Вашата институција);
- » одлучете кој првен основ за обработка на податоци ќе го примените;
- » доколку партнерите во проектот треба да пристапат до личните податоци, обезбедете договори со нив;
- » размислете дали е потребна Проценка за заштита на податоците;
- » информирајте ги учесниците во истражувањето како нивните лични податоци ќе бидат користени, зачувани, обработени, пренесени, кој е контролорот на податоците (со контакт информации), правната основа и целта на обработката, периодот на задржување и нивните права; ова може да биде направено преку леток со информации или преку веб страница;
- » размислете каде безбедно да ги складирате личните податоци;
- » минимизирајте ги личните податоци кои ќе ги собирате и псевдонимизирајте каде е потребно.

Организациски подготовки за заштита на податоците

Правила и упатства во рамки на организацијата и тимот во врска со:

- » Лозинки: минимум барања за јачина на лозинките; управување/безбедно чување на лозинки.
- » Енкрипција: кои типови на податоци се енкритираат и за која цел, со кои алатки?
- » Безбедно пренесување и транспорт на податоци
- » Безбедно уништување податоци.

Ограничете го пристапот до чувствителни податоци:

Да се утврди кој треба да има пристап до кој вид на податоци. Ретрикции при пристап или доколку некој го напушти тимот.

Создадете свесност и одржувајте ја комуникацијата во тимот:

Често се случуваат грешки поради недоволна свесност за потенцијалните проблеми или закани. Дали сите во тимот ги знаат овие информации? Да се обезбеди дека сите во тимот се вклучени во дискусии на овие теми.

Безбедно бришење на податоци

- ◊ Физичко уништување на медиумот (сечење на диск)
- ◊ Користење на софтвери за сигурносно бришење (на хард дискови)
- ◊ За SSD (solid state drives) и USB (flash drives) само физичкото уништување е 100% гаранција дека податоците нема да може да бидат повратени
- ◊ Консултација со ИТ лица

Консултација на релевантни правила

- ◊ **Закон за заштита на лични податоци**- какви се националните законски одредби во врска со оваа материја?
- ◊ **Агенција за заштита на лични податоци** – консултација во случај да не сте сигурни какви мерки треба да се применат
- ◊ Консултација со тела за етика во научните истражувања

Мерки за етика и заштита на личните податоци

- ◆ Анкетни истражувања- брошура/леток со информации, анонимизација
- ◆ Квалитативни истражувања- формулар за согласност, псевдонимизација

Брошура/леток со информации

Елементи на добар леток со информации:

- » Цел на истражувањето;
- » Што сè вклучува учеството во истражувањето;
- » Придобивки и ризици од учеството во истражувањето;
- » Детали за истражувањето, на пример: финансиери, институција, име на проектот, контакт информации од истражувачите и како да се поднесе жалба;
- » Процедурата за повлекување од истражувањето;
- » Планираната употреба на податоците за време на истражувањето, дисеминацијата, чувањето, објавувањето и архивирањето на податоците;
- » Стратегиите со кои се обезбедува етичка употреба на податоците;
- » Процедурите за заштита на личните податоци, обезбедување доверливост и анонимизација на податоците, особено во врска со архивирањето на податоците, споделувањето и повторната употреба.

Информирана согласност

- Се дава на почетокот на истражувањето (но и гранулирана (за идни истражувања и реупотреба) и ретроспективна согласност).
- Начин преку кој се задоволуваат етичките и правните обврски кон испитаниците.
- Се зголемува вредноста на истражувањето.

Со цел добивање на согласност, истражувачите треба:

- » Да ги информираат учесниците за целите на истражувањето;
- » Да се им објасни на испитаниците што ќе се случи со нивниот придонес (вклучувајќи идно архивирање и споделување на податоците);
- » Да се наведат чекорите кои ќе бидат преземени за да се заштити нивната анонимност и доверливост;
- » Да се наведе дека имаат право да се повлечат од истражувањето. Согласноста треба да се даде слободно, информирано, недвосмислено, специфично и со јасна афирмативна акција која означува согласност за обработка на личните податоци.

Анонимизација

- ◊ Да се пристапи кон анонимизација во контекст на целиот проект
- ◊ Како да се употреби заедно со информираната согласност и контролата за пристап до податоците.

Личните податоци може да се откријат преку две категории идентификатори:

- ◊ **Директни идентификатори** се на пример името на испитаниците, адресата или телефонот кои специфично ги идентификуваат;
- ◊ **Индиректните идентификатори** се оние кои кога ќе се постават заедно со други информации исто така може да го откријат испитаникот/ката, преку крос-референцирање на занимањето, плата, возраст или локација.

Анонимизација наспроти псевдонимизација

- ◊ Анонимизацијата неповратно ги уништува сите можности да се идентификува субјектот на податоците.
- ◊ Псевдонимизацијата, барем теоретски, овозможува ре-идентификување на субјектите на податоци преку дополнителни информации.
- ◊ Анонимни податоци се таков вид на податоци преку кои не може да се идентификуваат лицата во датасетот на ниту еден начин. Ниту преку директни идентификатори како име или број на социјално осигурување, ниту индиректно преку дополнителни променливи/варијабли, листа на имиња, формула за енкрипција или преку кодна листа.
- ◊ Дефиниција за псевдонимизација на GDPR “обработка на личните податоци на начин на кој личните податоци нема да може да му се препишат на специфичен субјект на податоците без употреба на дополнителни информации”. Да се pseudonymise dataset “дополнителните информации мора да се чуваат одделно и се предмет на технички и организациски мерки кои обезбедуваат дека личните податоци нема да му се препишат на идентификувано лице или физичко лице кое може да биде идентификувано”. Директните идентификатори се чуваат одделно и безбедно од обработените податоци.

Квантитативни податоци

- ◊ Отстранување или агрегација на променливи/варијабли, или намалување на прецизноста или деталноста на варијаблата (ако е текстуална);
- ◊ Агрегација или намалување на прецизноста на променливата/варијаблата, како на пример возраст или место на живеење;
- ◊ Генерализација на значењето на детална текстуална променлива/варијабла преку замена на потенцијално чувствителен текст со погенерален текст;
- ◊ Ограничување на високиот и нискиот опсег на променливата/варијаблата, со цел да се скријат исклучоците (outliers) доколку вредностите за некои лица се невообичаени или нетипични во рамки на истражуваната група.

Квалитативни податоци

- ◊ Користење на псевдоними или генерички дескриптори за да се едитираат податоците, со цел да се избегне бришење на информации;
- ◊ Планирање на анонимизацијата во текот на транскрипцијата или иницијалното пишување на проектот;
- ◊ Користење на псевдоними или замени кои се конзистентни во рамки на истражувачкиот тим или проект. На пример, користење на истите псевдоними во публикации или последователно истражување;
- ◊ Внимателно користење на 'search and replace' техники со цел да не се направат ненамерни промени или да не се прескокнат зборови каде има грешка во пишување;
- ◊ Јасно идентификување на замените во текстот, на пример со [загради] или користејќи XML tags како `<seg>word to be anonymised</seg>`;
- ◊ Креирање на „клуч за де-анонимизација“ за сите замени, агрегации или отстранети зборови кој се чува безбедно и одделно од анонимизираните датотеки со податоци.



Consent Form for [name of project]

Please tick the appropriate boxes

Yes No

Taking Part

I have read and understood the project information sheet dated DD/MM/YYYY. Yes No

I have been given the opportunity to ask questions about the project. Yes No

I agree to take part in the project. Taking part in the project will include being interviewed and recorded (audio or video).¹ Yes No

I understand that my taking part is voluntary; I can withdraw from the study at any time and I do not have to give any reasons for why I no longer want to take part. Yes No

Use of the information I provide for this project only

I understand my personal details such as phone number and address will not be revealed to people outside the project. Yes No

I understand that my words may be quoted in publications, reports, web pages, and other research outputs. Yes No

Please choose **one** of the following two options:

I would like my real name used in the above

I would not like my real name to be used in the above.

Use of the information I provide beyond this project

I agree for the data I provide to be archived at the UK Data Archive.² Yes No

I understand that other authenticated researchers will have access to this data only if they agree to preserve the confidentiality of the information as requested in this form. Yes No

I understand that other authenticated researchers may use my words in publications, reports, web pages, and other research outputs, only if they agree to preserve the confidentiality of the information as requested in this form. Yes No

So we can use the information you provide legally

I agree to assign the copyright I hold in any materials related to this project to [name of researcher]. Yes No

Name of participant [printed] Signature _____ Date _____

Researcher [printed] Signature _____ Date _____

Project contact details for further information: Names, phone, email addresses, etc.

Notes:

- Other forms of participation can be listed.
- More detail can be provided here so that decisions can be made separately about audio, video, transcripts, etc.

This work is licenced under the Creative Commons Attribution-Non-Commercial-Share Alike 2.0 UK: England & Wales License. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/2.0/uk/>



TRANSCRIPT SYMBOLS

INT	Interviewer
RESP	Respondent
[?]	Unintelligible
[]	Edited to maintain anonymity [1- Added to clarify anonymisation of transcript]

Mr Tom Jeavons **[2 - Delete and replace with [This gentleman]]**, aged 63, **[3 - Delete]** was suffering from metastatic cancer resulting from a primary site in the bladder **[4 - Delete]**. His wife, Sue **[5 - Delete]** (58), **[6 - Delete]** had been his main carer for many months as he struggled with severe pain, anxiety and other symptoms. Eventually, she received support from the hospice at home team, based at their nearby hospice – St Barbara. **[7 - Delete]** 11 days before his death, he was admitted to their inpatient unit, where he died. The case was identified by the staff there as a “critical case”, involving palliative sedation and the difficulties staff experienced in controlling his complex symptoms. Other interviews carried out were with the hospice consultant, Dr Jane O’Connor **[8 - Delete]** and three nurses: Elaine McDonald, Claire Smith and Mark Ferguson **[9 - Delete]**. Mr and Mrs Jeavons’ **[10 - Delete and replace with [The couple’s]]** GP, Dr Paul Hyde, **[11 - Delete]** was also interviewed which added a different medical perspective, making this an unusual case.

Central themes in all of the interviews were his intractable and distressing symptoms and the repeated requests from Mr Jeavons **[12 - Delete and replace with [the patient]]** for euthanasia. His wife mentions earlier discussions with Mr Jeavons **[13 - Delete and replace with [her husband]]** about the possibility of going to a Dignitas clinic, but he was already too ill to travel. She also expresses how concerned she was about what Mr Jeavons’s **[14 - Delete and replace with [his]]** adult children might witness when he was dying in the hospice.

INT: So, really, it’s as I said to you: I want you to tell me what you can remember about Mr Jeavons’ **[15 - Delete and replace with [your husband’s]]** care in the last week of his life ... or about Mr Jeavons **[16 - Delete and replace with [your husband]]** in the last week of his life.

RESP: Yeah, erm, 11 days, Tom **[17 - Delete and replace with [he]]** was in St Barbara’s Hospice **[18 - Delete and replace with [the hospice]]** for the last 11 days of his life so...

Авторски права

- ♦ Авторското право е меѓународно призната форма на интелектуална сопственост, кое е резултат на оригинална работа, како што е истражување. Не треба да е регистрирано за да се однесува на одреден производ од работата (дело).
- ♦ Авторските права на „производи“ од истражувања вклучуваат табели со податоци (и други форми на оригинално селектирани или организирани податоци), публикации, извештаи и компјутерски програми. Авторските права не ги покриваат фактите, идеите или концептите, туку само специфичниот начин на кои тие се презентирани.
- ♦ Правото му припаѓа на авторот/ката на делото или на нивната релевантна институција-универзитетите имаат различни пристапи во врска со ова.
- ♦ Дело заштитено со авторство вообичаено не може да се објави, репродуцира, адаптира или преведе без согласност од неговиот сопственик.

Најбитни прашања за истражувачите во врска со АП

- ◆ Доколку сакате да употребите нечии податоци или планирате да ги архивирате или споделите вашите, треба да се запрашате **кој е носител на авторските права** на датасетот.
- ◆ **Дали смеете да ги користите** и на каков начин? Дали смеете да ги архивирате и да ги објавите во репозиториум за податоци?
- ◆ Како одговарате на прашањето за тоа кој е носителот на авторските права?
Дали сте тоа Вие, Вашиот работодавец, архивот на податоци, колеги истражувачи? Одговорот зависи од повеќе фактори како на пример: кој придонел кон креирањето на податоците, дали се користени податоци од други датасетови, што е наведено во договорот за вработување на истражувачите и сл.

Препреки за транс-европско споделување податоци

Различни пристапи помеѓу земјите

Иако најголем дел од истражувачките податоци нема да ги исполнат условите за заштита бидејќи не се сметаат за „дела“ (се однесуваат пред се на факти), сепак разликите меѓу земјите може да се проблематични. Германија, Данска и Холандија имаат повисоки стандарди за оригиналност, Обединетото кралство или ниски стандарди (вештина, способност за просудување и работа) со што истражувачките податоци лесно добиваат целосна заштита од авторски права.

Разлики во носителите на авторските права

Најчесто самите истражувачи, освен Данска и ОК каде носител е работодавецот.

Лиценците како решение

За да се осигура дека истражувачките податоци може да се споделуваат и слободно да се употребат повторно секогаш треба да се прибават лиценци од потенцијалниот носител на правата. Со соодветната лиценца истражувачите може да го отстранат барањето за право на ИС кои може да се примеливи врз податоците кои ги генерираат во текот на јавно финансираното истражување.

План за управување со податоци

Типови податоци

- ◊ Дали користите лични податоци и дали Вашите податоци бараат некоја посебна заштита?
- ◊ Како (чувствителните) лични податоци ќе бидат заштитени за време на проектот?

Етичка проценка (ако е потребно)

- ◊ Дали за Вашиот проект е потребно одобрување од етички комитет (национален/локален)?
- ◊ Ако да, дали сте подготвени за овој процес?
- ◊ Како можните етички прашања ќе бидат земени предвид, и кои правила за однесување ќе се следат?

Информирана согласност (доколку е применливо)

- ◊ Дали е потребна информирана согласност за Вашиот проект? Доколку е така, како ќе биде обезбедена?
- ◊ Дали ќе добиете пишана согласност од испитаниците за споделување на податоците и надвор од Вашето истражување?
- ◊ Дали разговарате за архивирање и споделување на податоците со испитаниците од кои собираете податоци?
- ◊ Дали адекватно ја документирате согласноста, како што бара GDPR?
- ◊ Како се организирани и зачувани документите за согласност?

План за управување со податоци

◊ Заштита на учесниците

- » Како ќе ја обезбедите приватноста на учесниците во истражувањето?
- » Дали го почитувате законодавството во областа на заштита на податоците?
- » Дали е потребно да ги анонимизирате податоците, на пример, да се отстранат информациите преку кои може да се идентификува или личните податоци, за време на истражувањето или при подготовката за споделување?

◊ Интелектуална сопственост

- » Дали треба да бидат разгледани аспекти во врска со авторските права или правата од интелектуална сопственост?
- » Дали определите кој е носител на авторските права на Вашите податоци? Дали можеби се работи за заедничка сопственост?
- » Дали е потребна дозвола за собирање или повторна употреба на податоците?
- » Дало овие права ќе бидат пренесени на друга организација со цел архивирање?

Прашања или коментари?

anetac@isppi.ukim.edu.mk

 cessda.eu

 @CESSDA_Data

  Licence: CC-BY 4.0



6. Архивирање & Публикување

Елеонора Серафимовска, УКИМ, ИСППИ, Скопје

 cessda.eu

 @CESSDA_Data

  Licence: CC-BY 4.0

CESSDA Training Team (2017 - 2022). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC. Retrieved from <https://dmeg.cessda.eu/>

Податоците со висок квалитет имаат голем потенцијал повторно да се користат на многу начини. Архивирањето и правилното објавување на вашите податоци е сржта за правење на вашите податоци **FAIR**.

Потребно е да се донесе информирана одлука за тоа каде да ги архивирате и објавите вашите податоци на таков начин што другите ќе можат правилно да пристапат, разберат, користат и цитираат.

Архивирање на податоци за идна референца

Архивирањето на податоците од истражувањето е за складирање и зачувување на податоците од истражувањето на долг рок. Кога ги архивирате вашите податоци, бидете сигурни дека подоцна ќе можете да ги читате и пристапите до нив.

Потоа, можете да дозволите и пристап од други за целите на проверка кога ќе пристигне такво барање. Во сите случаи, треба да ги чувате вашите податоци безбедно, во соодветен формат на датотека, со соодветна документација.

Објавување податоци за повторна употреба

За да ги направите вашите податоци за цели надвор од онаа за која сте ги собрале, треба да ги објавите вашите податоци. Објавувањето на вашите податоци е чин на јавно откривање на податоците од истражувањето што сте ги собрале, правејќи можност лесно да се најдат, да бидат достапни и повторно да се користат.



Причини за архивирање и публикување

Инсентивите што ги мотивираат истражувачите да ги архивираат и објавуваат своите истражувачки податоци (Van den Eynden & Bishop, 2014; Hahnel et al., 2017) спаѓаат во четири главни категории:

- ◊ **Кариерни придобивки**
Објавувањето на податоците може да доведе до зголемена видливост, повторна употреба и цитирање, а со тоа и препознавање на научната работа.
- ◊ **Научен напредок**
Архивирањето и објавувањето на податоците има директни придобивки за самото истражување, за дисциплината и за науката воопшто, со тоа што овозможува нови соработки, нови употреби на податоци и воспоставување врски со следната генерација истражувачи.
- ◊ **Норми**
Нормите на проектот, истражувачката група и/или дисциплина може да одредат дали истражувачот е склон да ги објавува своите/нејзините податоци. Генерално, отвореноста на податоците од истражувањето е во срцето на научната етика и тоа треба да стане норма/стандард
- ◊ **Надворешни мотиватори**
Надворешните двигатели како што се политиките за управување со истражувачки податоци од финансиери и издавачи на истражување имаат значително влијание врз архивирањето и објавувањето на податоците

Рути за објавување податоци



За базата на податоци да се смета како публикација, треба да следи сличен процес на објавување како статија (Brase et al., 2009) и треба да биде:

Правилно документирана со метаподатоци;

Прегледана за квалитет;

Да може да се пребарува и открива во каталози (или бази на податоци);

Да може да се цитира во публикации.

Авторите на извештајот од Knowledge Exchange (Knowlegde Exchange, 2013) го дефинираат овој тип на објавување податоци како „**Објавување со големо П**“ и го споредуваат со „објавување со мало п, што значи дека истражувачите ги објавуваат своите датотеки со податоци на веб-локација некаде . Објавувањето со мало „п“ значи дека нема гаранции дека податоците ќе бидат таму по некое време или дека датотеките нема да се оштетат.

Пет рути за објавување



Во списание кое објавува сетови (бази) на податоци

Институционален репозиториум на податоци

Репозиториум за општа намена ([Zenodo](#), [Figshare](#) or [Harvard Dataverse](#))

Специфичен репозиториум за податоци за одредена област/научна сфера

Доверлив репозиториум за податоци специфични за одредена област/научна сфера

Избор на репозиториум (архива/центар за податоци) за истражувачки податоци

Ги има многу со разни намени / *Whyte (2015)*.

Понекогаш складиштето е веќе одредено од вашиот финансиер или друга надворешна страна.

Објавување со архивите на CESSDA

CESSDA – Consortium of European Social Science Data Archives

CESSDA is a European research infrastructure consortium (ERIC) with some 20 European data archives that work together to improve access to social science research data. SND has been appointed by the Swedish Research Council to represent Sweden in this collaboration.



За висококвалитетни податоци со потенцијал за повторна употреба, ви препорачуваме да обезбедите долгорочен пристап со објавување на вашите податоци во доверливо складиште/репозиториум, како многу од архивите на CESSDA.

Архивите на CESSDA имаат за цел да ги направат податоците од истражувањето достапни со што е можно помалку ограничувања, а во исто време да ги заштитат (чувствителните) лични податоци од несоодветен пристап.

Публикување со архивите на CESSDA

Придобивки од архивите на CESSDA

Објавувањето на вашата база на податоци во архивата CESSDA има голема предност што има стручна помош на дофат. Експертите за управување со податоци за истражување на CESSDA можат да ви помогнат да ја зголемите разбирливоста, видливоста, пронаоѓањето, повторното користење, долговечноста и севкупниот квалитет на вашите збирки податоци на следниве начини:

CESSDA – Consortium of European Social Science Data Archives

CESSDA is a European research infrastructure consortium (ERIC) with some 20 European data archives that work together to improve access to social science research data. SND has been appointed by the Swedish Research Council to represent Sweden in this collaboration.



Публикување со архивите на CESSDA



Разбирливост: Во архивата CESSDA, можете да ги депонирате вашите податоци со помош на експерт за податоци. Овој експерт ќе ве советува кои информации се потребни за да ги разберете вашите податоци

Пронаоѓање и видливост: Кога ги објавувате вашите податоци во архива CESSDA, вашите податоци стануваат повидливи на неколку начини: Со примена на постојан идентификатор на вашите збирки на податоци, вашите податоци секогаш **може да се најдат и цитираат**.

Пристапност и повторна употреба: Со комбинација на лиценцирање податоци и категории за пристап архивите на податоци CESSDA можат да го контролираат точното ниво на пристап и дозволената повторна употреба. На овој начин, можете да го направите оптималниот избор за да го подобрите потенцијалот за повторна употреба на вашите податоци од истражувањето, а истовремено да ги заштитите идентитетите на вашите учесници.

Долговечност: Тие даваат совети за најдобрите формати на датотеки за долгорочно зачувување;

Тие нудат експертиза и услуги за конвертирање на податоци во нови формати

Тие додаваат вредност на податоците, на пример со нова функционалност за пребарување на податоците.

Квалитет: Во неколку архиви на CESSDA, експерт ќе го прегледа квалитетот на вашите податоци судејќи на пр. содржината на студијата, методологијата, релевантноста, правната конзистентност и документацијата на материјалите.

Цитирање на вашите податоци



За добиените истражувачки податоци да бидат единствено на нивните создавачи, се препорачуваат два типа на идентификатори:

Постојан идентификатор (PID persistent identifier) на вашата база на податоци

Објавувањето на сетовите на податоци станува сè поважно како цитирачки придонес во истражувањето. За да станете цитирани, треба да бидете сигурни дека вашите збирки на податоци добиваат единствен, постојан идентификатор. Дигиталниот идентификатор (DOI) е добро познат идентификатор во академската заедница. Имањето PID е важен аспект за да се осигурате дека вашите податоци ги исполнуваат F (Пронаоѓање) и A (Пристапност) во управувањето со податоци FAIR.

Постојан идентификатор на авторот

За да ги направите резултатите од вашето истражување уште поповрзани, можете да го креирате вашиот личен постојан идентификатор на авторот. ORCID iD обезбедува таков постојан дигитален идентификатор, разликувајќи ве од секој друг соработник и поддржувајќи автоматизирани врски меѓу сите ваши професионални активности. Со креирање и користење на ORCID iD ќе можете да ја презентирате целата ваша - растечка - работа преку еден канал.

Лиценцирање на вашите податоци

Доколку ги објавите вашите податоци во репозиториум за податоци по ваш избор, на вашите податоци ќе се примени договор за лиценца. Договорот за лиценца е правен договор помеѓу создавачот/депонентот на сетот податоци и репозиториумот на податоци, што означува што му е дозволено на корисникот да прави со податоците. Важен аспект е да се осигурате дека вашите податоци го исполнуваат R (повторливото користење) во FAIR управувањето со податоци.

За повторна употреба што е можно поверојатно, ве советуваме да изберете лиценца која:

- ◆ Ги прави достапни податоците за најшироката можна публика;
- ◆ Го прави најширокиот можен опсег на употреба на вашите податоци



Лиценцирање на вашите податоци

За лиценците за Creative Commons

Главните атрибути на користење на лиценци Creative Commons (2017) за лиценцирање на податоци, збирки на податоци и бази на податоци (Корн и Опенхајм, 2011) се:

- ♦ Леснотијата на користење на лиценците;
- ♦ Широкото усвојување на лиценците;
- ♦ Нивната флексибилност;
- ♦ Нивната достапност во форми читливи од луѓе и машински читливи што им овозможуваат на истражувачите и на компјутерите веднаш да знаат што им е дозволено да прават со вашите податоци;
- ♦ Шанса вашите податоци да се користат повторно.

Licence	Can I copy & redistribute the work?	Is it required to attribute the author?	Can I use the work commercially?	Am I allowed to adapt the work?	Can I change the licence when redistributing?
CC0	Y	N	Y	Y	Y
CC BY	Y	Y	Y	Y	Y
CC BY-SA	Y	Y	Y	Y	N
CC BY-ND	Y	Y	Y	N	Y
CC BY-NC	Y	Y	N	Y	Y
CC BY-NC-SA	Y	Y	N	Y	N
CC BY-NC-ND	Y	Y	N	N	Y

Категории на пристап

- ◆ Објавувањето податоци во репозиториум за податоци не ги прави автоматски отворено достапни. (Чувствителните) лични податоци сè уште може да се заштитат со ограничување на пристапот до податоците. Контролите за пристап може да дозволат контрола до поединечно ниво на датотека, што значи дека мешаните нивоа на контрола на пристап може да се применат на собирање податоци.
- ◆ Многу репозиториуми на податоци работат со тристепен пристап до податоци:



Категории на пристап

Слободен пристап

Податоци до кои може да пристапи секој корисник без разлика дали е регистриран или не. Податоците во оваа категорија не треба да содржат лични информации освен ако не се даде информирана согласност.

Пристап за регистрирани корисници (заштитен)

Податоци до кои се достапни само корисници кои се регистрирале во архивата. Овие податоци не содржат директни идентификатори, но може да постои ризик од обелоденување преку поврзување на индиректни идентификатори.

Ограничен пристап

Пристапот е ограничен и може да се дозволи само на барање. Оваа категорија за пристап е наменета за најчувствителните податоци што може да содржат разоткривачки информации.

Ограничениот пристап бара долгорочна посветеност на истражувачот или лицето одговорно за податоците да се справи со претстојните барања за дозвола.

Ембарго

Освен што нудат можност за ограничен пристап „засекогаш“, повеќето репозиториуми/складишта на податоци ви дозволуваат да поставите привремено ембарго на вашите податоци. За време на периодот на ембаргото, се објавува само описот на базата на податоци. Самите податоци ќе станат достапни во отворен пристап по одреден временски период.



Промоција на вашите податоци

Размислете за промовирање повторна употреба на вашите податоци на еден од следниве начини:

- ◊ Изберете отворен пристап
- ◊ Лиценцирајте ги вашите податоци
- ◊ Секогаш наведувајте ги вашите податоци
- ◊ Публикувајте во списание за податоци
- ◊ Подучувајте со вашата база на податоци
- ◊ Изберете репозиториум за податоци што ги промовира вашите податоци
- ◊ Зголемете го влијанието на вашите податоци со алтметрика





Прашања





Пронаоѓање на истражувачки податоци

Јован Близнаковски / МК АИПОН, ИСППИ-УКИМ

*Вебинар: Управување со истражувачки податоци од општествените науки
26 октомври 2022*

DOI: 10.5281/zenodo.7250138

 cessda.eu

 [@CESSDA_Data](https://twitter.com/CESSDA_Data)



Licence: CC-BY 4.0

Вовед

- ◊ Користењето на веќе собраните истражувачки податоци има долга традиција во општествените науки.
- ◊ Податоците се реупотребуваат на други истражувачки прашања или со цел репликација на оригиналните студии.
- ◊ Особено во анкетните истражувања: т.н. „омнибус“ податоци се собираат и се прават јавно достапни за понатамошна употреба (на пр. Европско општествено истражување; Компаративна студија на изборните системи; Студија за европски вредности).
- ◊ Пронајдените податоци можат да се користат и за други цели: споредување на податоци, споредување на истражувачки дизајни, зајакнување на истражувачките стандарди и стимулирање на иновација.

Вовед II (Што ќе погледнеме денес?)

- ◊ Процесот на пронаоѓање на истражувачки податоци
- ◊ Различни ресурси за пронаоѓање на податоци
- ◊ Архивите на податоци како ресурс за пронаоѓање
- ◊ Пристап, користење и цитирање на податоци

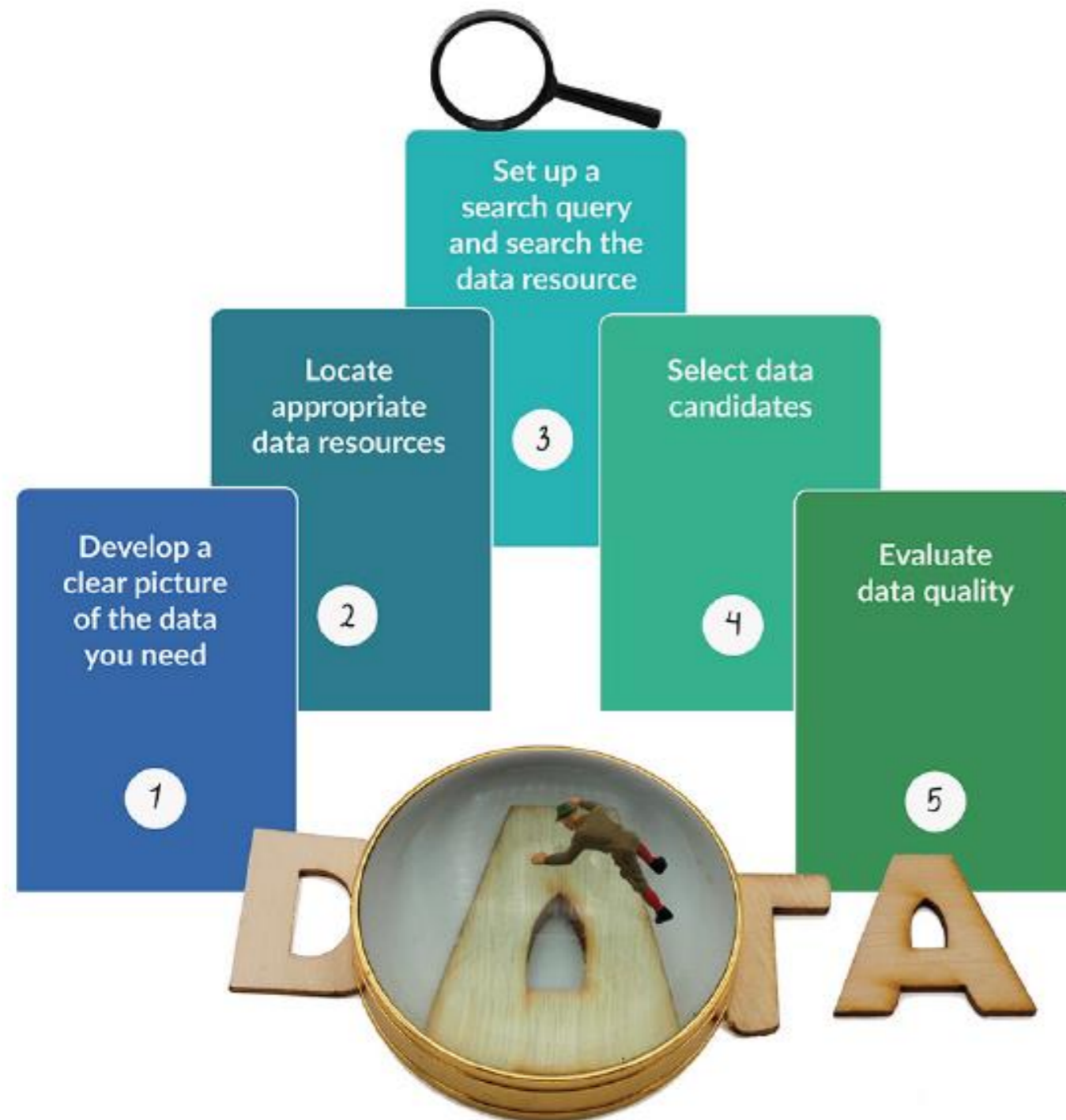


ПЛАН ЗА УПРАВУВАЊЕ СО ПОДАТОЦИ

Пример за пронаоѓање

- ◇ Истражувачот X е заинтересиран за компаративни податоци за изборно однесување. **Што може да направи?**
- ◇ Преглед на литературата (можен исход: лоцирање на достапни податоци)
- ◇ Преглед на достапните податоци во соодветна меѓународна програма за истражување на јавното мислење (на пр. Компаративната студија на изборните системи)
- ◇ Преглед на достапните податоци во архиви на податоци (на пр. GESIS, FORS...)
- ◇ Преглед на податоците достапни на интернет (на пр. GOOGLE)

Чекори во пронаоѓањето на податоците



- ◊ 1: Јасна претстава за тоа какви податоци се потребни
- ◊ 2: Идентификување на соодветни ресурси
- ◊ 3: Пребарување (на соодветен начин)
- ◊ 4: Селектирање на податоци кои би можеле да се искористат
- ◊ 5: Евалуација на квалитетот на податоците

1. Какви податоци се потребни?

- ◊ Која е тематската област на студијата?
- ◊ Кое е истражувачкото прашање?
- ◊ Со кои научни концепти сакате да работите?
- ◊ Како сакате тие концепти да бидат операционализирани?
- ◊ Каква е теоријата на која се потпира проектот?
- ◊ Каков е типот на студијата? (оригинална, репликација, настава)
- ◊ Какви се специфичните карактеристики на податоците кои ви се потребни? (тип на податоци, целна популација, географски карактеристики, временска рамка)
- ◊ Други карактеристики? (формат на датотеките, отворени податоци)

2. Кои ресурси се достапни?

- ◆ Регистри на репозиториуми на податоци (re3data.org; OpenAIRE Explorer; OpenDOAR; FAIRsharing)
- ◆ Интернет пребарувачи и агрегатори на мета-податоци (Google; Google Scholar; Google Dataset Search; DataCite; DataSearch)
- ◆ Каталог на податоци (CESSDA Data Catalogue)
- ◆ Научно списание посветено на податоци (Research Data Journal for the Humanities and Social Sciences)

3. Како да се уреди пребарувањето?

- ◆ Запознавање со структурата на конкретниот ресурс на податоци; регистрација; запознавање со можностите за пребарување
- ◆ Користење на соодветни концепти во соодветни полиња (различни полиња: question text, keywords, итн.)
- ◆ Во случај на „неубедливи“ пронајдоци: проширување или намалување на опсегот/термини кои се пребаруваат

4. Како да се изберат податоците кои би можеле да се користат (кандидати)?

- ◊ Проценка на релевантноста на податоците согласно истражувачкото прашање, вклучително и користење на соодветни концепти/индикатори

5. Како да се процени квалитетот на податоците?

- ◊ Преглед на документацијата
 - ◊ На ниво на целокупен проект (цели, истражувачки прашања, хипотези, методологија, инструменти, мерки, итн.)
 - ◊ На ниво на податоци (на пр. документација за спроведено теренско репрезентативно истражување)

Корисни архиви на податоци во општествените науки

- ◆ CESSDA архиви и CESSDA каталог на податоци
- ◆ Национални архиви кои не се членки на CESSDA (пр. Русија, Италија, Романија) и архиви вон Европа
- ◆ EUROSTAT – статистичката канцеларија на ЕУ
- ◆ European Union Open Data Portal: отворени податоци од ЕУ институциите
- ◆ Лонгитудинални истражувачки проекти (ЕОИ; ЕВС; итн...)
- ◆ OECD iLibrary; UNdata; UNESCO Institute for Statistics (податоци за Целите за одржлив развој); UNICEF (податоци за млади и жени); World Bank Open Data (податоци за развој)

Пристап и користење

- ◇ Проверете ги условите за пристап и користење
 - ◇ Отворени податоци
 - ◇ Отворени податоци по регистрација
 - ◇ Отворени податоци под специфични услови (дозвола, само за научни цели, само за некомерцијални цели, доверливи податоци)
 - ◇ Дозволен пристап само до мета-податоците
 - ◇ Податоци под „ембарго“
- ◇ Земете ги предвид и:
 - ◇ Евентуалните трошоци за пристап
 - ◇ Форматот на податочните датотеки

Различни начини на пристап и користење

- ◇ Директно симнување
- ◇ Онлајн анализа (не е потребен посебен софтвер)
- ◇ Дозволен пристап само во посебна безбедна соба
- ◇ Безбеден систем за далечниски пристап

Цитирање на податоците

- ◊ Податоците секогаш треба да бидат цитирани
- ◊ Доколку е достапно: користење на постојани идентификатори (пр. DOI – Digital Object Identifier)
- ◊ Примери:
 - ◊ Автор(и) (Година на публикација). Наслов. Верзија. Издавач. Тип на ресурс. Идентификатор
 - ◊ Webb, P., Scarrow, S., Poguntke, T. (2017). Political Party Database, 2011-2014. [data collection]. UK Data Service. SN: 8265, <http://doi.org/10.5255/UKDA-SN-8265-1>
 - ◊ European Foundation for the Improvement of Living and Working Conditions, (2017). European Working Conditions Survey, 2015. 4th Edition. UK Data Service. [data collection]. <http://doi.org/10.5255/UKDA-SN-8098-4>



Прашања? Мислења?

 cessda.eu

 @CESSDA_Data

  Licence: CC-BY 4.0