**2. Ненормативни източници**

***2.1.  Добри практики***

Данните, които сваляме от интернет, съдържат полезна информация, която често е скрита сред други данни – шум, обаче за да достигнем то нея първо трябва да се приведат във вид, удобен за анализиране. Това се постига чрез редица от стъпки, които трябва да се следват, за да се извлече максимума от сета от данни :

* Прочистване на данните – в случаите, когато има излишък от данни, които не представляват интерес на проучването, когато има наличие на дублиращи се или липсващи данни, или когато има данни с екстремални стойности, които не отразяват реалността.
* Трансформация на данни – при моделиране/визуализиране на данните, за представяне на резултата от анализите пред заинтересованите лица

Често резултатите от анализите трябва да се представят пред някаква аудитория. Един от подходите за поднасяне на тази информация по подходящ начин е „Storytelling“, чиито най-важни концепции и препоръки са :

* Обвързване на данните с история – повишаване на ефективността при запомняне на множество данни
* Използване на визуализации за представяне на резултатите от анализа – чрез различни видове диаграми
* Опознаване на публиката - научна общност или нетехнически лица
* Предоставяне на контекст – съпоставка с предишен резултат на анализ
* Сравняване на резултатите (през 3 от годините)

Има няколко концепции за придържане към добрите практики при работа с данни:

* “Tidy Data” – използва се за анализиране на една величина. Всички основни манипулации на данни като филриране, трансформиране, агрегиране и моделиране са значително опростени.
* “Untidy Data” – по-добър от “Tidy Data” при репрезентиране на обобщения, връзки и модели, но е проблематичен при бъдещи манипулации на данните
* Корелационнен анализ – анализира се силата на връзките между величините съставящи този сет от данни. Т.е чрез тази техника може да се определи дали наистина съществува връзка между две величини.
  1. ***Технически стандарти***

Въпреки свързването на множество сетове от данни е изискване, регламентът не дава насоки за това какво представлява ефективна техника за идентифициране на критерии, по които да се свържат данните, или за филтрирането им .

* 1. ***Технологии***
  + Microsoft Excel - програма за електронни таблици, част от пакета MS office, създаден от Microsoft. Excel позволява бързо и лесно да се обработват и съхраняват списъци от данни, да се извършват изчисления, да се създават отчети и диаграми.
  + Конвертор от csv към excel чрез Mozilla - <https://convertio.co/csv-xlsx/>
  1. ***Общоприети лицензи***
  + Microsoft Office Standard 2019
  + Mozilla Public License v2.0

Източници:

* Maggie Castle and Jaimee LaTorra**,** 8 Data Storytelling Concepts (with Examples!) - <https://twooctobers.com/blog/8-data-storytelling-concepts-with-examples/>
* Guide to Best Health Care Data Management Practices for Data Analysis –
* <https://www.prometheusresearch.com/guide-to-best-health-care-data-management-practices-for-data-analysis/>