



Open Data Science

Etiske overvejelser

Stephanie Rose Acampado Soelmark, PZG932

Etiske overvejelser

Open science omhandler at tilgængeligøre data for alle. Dette indebærer både kode, analyse, design, resultater med mere. Deling af data er til fordel for det videnskabelige samfund, man skaber hermed et samarbejde hvor alle kan komme med input og eller bygge videre på denne data for dermed at komme med nye fund. Et nyt syn på data kan ligeledes fører til identificering af fejl, således at man hurtigst muligt kan rette op på disse. Vælger man derimod at holde sin data privat er der større risiko for fejl og ensidige resultater. Dermed er open science en måde hvorpå man i samarbejde med sine medmennesker kan opnå de bedst mulige resultater (Hampton, 2015, s. 2-6).

Når man arbejder med data bør man være hensynsfuld, da denne data kan have indflydelse på mange individer. Således bør man følge visse etiske overvejelser når man håndtere data. I denne sammenhæng har ACM (Association for Computing Machinery) lavet et sæt etiske retningslinjer kaldet "The Code", disse anbefales at anvende ved bearbejdning af data (ACM, 2020).

Gennem dette kursusforløb har vi afleveret tre portfolio opgaver hvor vi har bearbejdet forskellige data. Første portfolio omhandler et Titanic datasæt. Dette indebærer information om 887 passagerer herunder navn, overlevelse, klasse med mere, det vil sige personlige oplysninger. Ifølge The Code bør man ikke anvende sådanne oplysninger uden den vedrørendes samtykke. Da vi har fået givet dette datasæt af vores underviser, vides det ikke om dette et blevet gjort. Fra dataen udledes det at størstedelen af disse passagerer døde med Titanic, dog er der stadig en mulighed for at de overlevende er blevet spurgt om samtykke. Dermed omhandler størstedelen af den pågældende data information om omkomne mennesker, som sandsynligvis stadig har pårørende der er i live. Dermed er måden hvorpå man behandler denne data vigtig, da man skal respektere og forhindre negative konsekvenser for de omkomne og deres pårørende (ACM, 2020).

Anden portfolio omhandler den engelske avis The Guardian. Hertil anvendte vi en API for at kunne få adgang til avisens data. Dermed havde vi adgang til alle avisens artikler, dette er en masse information man kan manipulere og udstille som ønsket. Følger man The Code er det dog ens ansvar at bearbejde sådanne data uden af det skaber negative konsekvenser for andre. Man skal

være forsigtig og ansvarlig med information, samt anvende det på sådan en måde at det bidrager til samfundet (ACM, 2020). I min gruppe har vi i denne sammenhæng valgt at fokusere på artikler omhandlende Boris Johnson. Intentionen her var at finde ud af i hvilke emner Johnson bliver diskuteret, dermed søgte vi blot at finde The Guardian's fremvisning og fortolkning af denne. Altså behandlede vi subjektivt fremstillede data.

Tredje portfolio omhandler Statens Museum for Kunst (SMK), der gennem deres projekt SMK Open har tilgængeliggjort alt deres data, det vil sige information om kunstværker på museet, herunder kunstnernavn, ophavs år, materialer med mere (SMK Open, u.å.). Dette tilgængeligøres ved brug af SMK's API (SMK's API (beta-version), u.å.). Denne har vi anvendt og dermed søgt at få adgang til information om statuetter på museet. I forhold til The Code er ophavsret en etisk retningslinje, man bør anerkende skabere af ideer og respektere deres ophavsret (ACM, 2020). Ved at gøre al sin data tilgængelig bør SMK dermed også anerkende kunsterne ved at tilgængelig gøre deres navne og udgivelsesår såvel som deres kunst. Dette er dog ikke tilfældet da vi i vores datasæt ikke fandt kunstnernavn og år ved alle værker. Derudover stod visse værker som "ikke offentlige", dermed er ikke alle værker tilgængelige. Det vil sige at SMK ikke har offentliggjort alle værker og ikke al nødvendig information, hvilket strider imod The Code. Dog er SMK Open ikke et færdiggjort projekt, dermed kan dette nås endnu.

Data er skabt af mennesker, derfor kan man aldrig være sikker på at opnå en objektiv sandhed. Man bør ikke udelukkende basere sine konklusioner på data, da denne er manipuleret af førnævnte. Derfor bør man analysere og fortolke hvorfra den pågældende data stammer. Et eksempel er The Guardian's, dennes data er opsat af menneskerne bag The Guardian og dermed fremstillet i tonen af avisen. Et andet eksempel er SMK Open hvor vi i vores datasæt fandt at der angiveligt skulle være en specifik periode hvorfra det højeste antal værker stammer fra, dog kan dette også skyldes at denne tidsperiode er hvor værkerne først blev digitaliseret, dermed er det ikke deres ophavs år men deres digitaliserings år.

Litteraturliste

ACM. (2020). ACM Code of Ethics and Professional Conduct. Lokaliseret d. 3 januar 2020 på:

<https://www.acm.org/code-of-ethics>

Hampton, S. E., Anderson, S. S., Bagby, S. C., Gries, C., Han, X., Hart, E. M., ... & Mudge, J. (2015). The Tao of open science for ecology. *Ecosphere*, 6(7), 1-13.

SMK (u.å.). SMK's API (beta-version). Lokaliseret d. 28 november 2019 på:

<https://www.smk.dk/article/smk-api/>

SMK (u.å.). SMK Open. Lokaliseret d. 28 november 2019 på: <https://www.smk.dk/article/smk-open/>