

Podatkovna otvorenost u svjetlu digitalne revolucije

Josipa Kern, Jadranka Mustajbegović, Biserka Bergman Marković, Inge Heim, Silvije Vuletić

Akademija medicinskih znanosti Hrvatske, Zagreb, Hrvatska

Početakom 2016. godine Akademija medicinskih znanosti Hrvatske dobila je informaciju o Sporazumu četiriju međunarodnih znanstvenih organizacija (Međunarodni savjet za znanost, Međuakademijsko partnerstvo, Svjetska akademija znanosti i Međunarodni savjet za društvene znanosti) s molbom da ga podrži. Sporazum definira podatkovnu otvorenost kao koncept, kao ideju prema kojoj izvjesni podatci trebaju biti dostupni za uporabu i ponovljeno objavljivanje bez ikakvih ograničenja poput autorskih prava, patenata ili nekih drugih mehanizama kontrole. Otvoreni pristup se odnosi u prvom redu na podatke iz znanstvenih istraživanja koja su financirana javnim novcem, posebice istraživanja od posebne važnosti za aktualne globalne izazove. Otvorenost pristupa podacima smatra se temeljnim preduvjetom za rigorozno provođenje znanstvenih istraživanja i maksimiziranje javne koristi od aktualne digitalne revolucije.

Sporazum predlaže 12 načela kao vodič u provođenju podatkovne otvorenosti, usmjerenih na uloge znanstvenika, izdavača, knjižnica i drugih zainteresiranih strana, kao i na tehničke zahtjeve na otvorenost podataka, te procjenjuje granice otvorenosti.

Ključne riječi: podatkovna otvorenost; otvorena znanost

Adresa za dopisivanje: Josipa Kern
Akademija medicinskih znanosti Hrvatske
10000 Zagreb
Praška 2/III
E-pošta: josipa.kern@gmail.com

Uvod

Osnova globalne politike u znanosti je razvoj i izgradnja infrastrukture kao preduvjeta osiguranja integriteta i vjerodostojnosti znanosti što je posebno zahtjevan i osjetljiv izazov za biomedicinu i zdravstvo. Suvremena digitalna revolucija je moćno sredstvo koje može pomoći u razvoju i izgradnji takve infrastrukture.

Izazovi digitalne revolucije

Digitalna revolucija pojam je koji označava brzi razvoj informacijskih i komunikacijskih tehnologija i inovacija u digitalnim sustavima koji iz temelja mijenja način mišljenja ljudi, njihova ponašanja, komuniciranja, rada i privređivanja (1). Dvadeset i prvo stoljeće nudi mnoštvo izazova znanstvenicima širom

svijeta. To vrijedi za sva znanstvena područja pa i za biomedicinu. Eksplozija kapaciteta za prikupljanje, pohranjivanje, obradu i prijenos podataka i informacija nikada nije bila snažnija. Zahvaljujući mnoštvu senzora i uređaja rađaju se podatci na različitim mjestima i u različitim uvjetima. Oblici su im različiti a formati njihova bilježenja variraju od brojki do tekstova, do snimki zvuka, biomedicinskih signala i slika (u miru ili u pokretu), a sve to u digitalnom obliku.

Samo prikupljanje podataka postaje sve jednostavnije jer su uređaji direktno povezivi s računalima, sve češće i na daljinu, putem interneta. Govori se o IoT (engl. *Internet of Things*), o nosivim uređajima (engl. *Wearable devices*), ili o dijagnostičkim uređajima za kreiranje kompleksnih informacija (slika, videa)

poput radioloških ili laboratorijskih uređaja potpomognutih informacijskim tehnologijama. Tu su, naravno, i aparati za digitalno fotografiranje (npr. za potrebe dermatologije), ali i vrlo kompleksni uređaji poput robota koji ne samo da prikupljaju podatke nego ih (putem umjetne inteligencije) i upotrebljavaju kako bi obavili radnju radi koje su i načinjeni.

Podataka je mnogo - što s njima?

Danas se u svijetu govori o *podatkovnoj masi* (engl. *big data*). Za razliku od podatkovnog mnoštva (engl. *massive data*) koje označava *isključivo* veliku količinu podataka (najčešće kao proizvod raznih uređaja koji provode mnoga mjerenja, u pravilu, longitudinalno, dakle, tijekom vremena) podatkovnu masu karakterizira osim količine (engl. *volume*), različitost izvora iz kojih dolaze i formata njihova bilježenja (engl. *variety*), kao i brzina kojom podatci priteču, proizvode se ili se mijenjaju (engl. *velocity*) (2). Neki autori dodaju i četvrtu dimenziju podatkovne mase - istinitost ili pouzdanost podataka (engl. *veracity*). Doduše, istinitost ili pouzdanost nije intrinzično svojstvo podatkovne mase. Istinitost podataka da se u njih može pouzdati i da im se može vjerovati je nešto što se želi postići. No, u stvarnosti se ne može uvijek izbjeći tzv. smetnje ili šumove u podacima, ili pak pristranost. O važnosti podatkovne mase za napredak medicinske znanosti i prakse - od personalizirane medicine do prevencije na populacijskoj razini - govore danas mnogi, primjerice, stručnjaci na Stanfordskoj petoj godišnjoj konferenciji o podatkovnoj masivnosti u biomedicini (3).

Pohrana podataka danas više nije bauk. Sve češće se pojavljuju nove vrste medija za pohranu velikih količina podataka, a i cijena im postaje sve prihvatljivija. Osim toga, nude se mogućnosti pohrane podataka u oblaku (engl. *cloud*). Javljaju se tvrtke koje se bave iznajmljivanjem prostora za pohranu i, istovremeno, zaštitom i sigurnošću pohranjenih podataka

(primjerice, *Google Health* ili *Microsoft HealthVault*).

Obrada podataka danas sve više počiva na primjeni umjetne inteligencije, na strojnom učenju iz prikupljenih podataka koji kontinuirano priteču, na analizi i predviđanju ishoda te procjeni rizika pojedinih ishoda.

Obradom i analizom podataka dolazi se do novih znanja, a informacije o novinama u nekom području objavljuju se u znanstvenim i stručnim časopisima i služe drugim znanstvenicima kao baza za daljnja istraživanja. Naime, biti „znanstveno pismen“ je pojam koji podrazumijeva „uporabu postojećih znanja i otkrivanje novih“. Otvoreni časopisi (engl. *Open journal*) učinili su iskorak koji omogućava širokoj publici, pa i znanstvenicima iz manje razvijenih sredina da putem interneta pristupe postojećim znanjima.

Osnovu istraživanja u biomedicini čine podatci prikupljeni od pacijenata. Podatke koje prikupljaju zdravstveni profesionalci pri obradi pacijenta pohranjuju se digitalno, u obliku elektroničkog zdravstvenog (medicinskog) zapisa. Međutim, ponekad je korisno da i sam pacijent prikuplja podatke u digitalnom obliku i time stvara svoj elektronički osobni zdravstveni zapis (npr. putem nekog nosivog uređaja poput holtera EKG-a, ili KMAT – kontinuirano mjerenje arterijskog tlaka, ili CPAP-a, ili bilježenjem, što je, primjerice, osoba radila u određeno vrijeme dok je nosila holter). Pritom je važno da podatci o pacijentu koji dolaze iz različitih izvora budu povezivi i dostupni kada se za to ukaže potreba.

Koncept podatkovne otvorenosti - otvorena znanost

Potreba za nadogradnjom postojećih znanja navodi na ideju da se izvorni podatci (oni, na temelju kojih se došlo do novih znanja) učine dostupnima i drugim znanstvenicima. Primjerice, željelo bi se uz objavljeni znanstveni rad objaviti i podatke

na temelju kojih se došlo do zaključaka koje rad navodi. Isto tako mnogi se podaci mogu naći na internetu, oni su dostupni, dakle otvoreni, primjerice putem društvenih mreža i foruma ili bibliografskih baza podataka. I takve se podatke može iskoristiti radi dobivanja novih znanja, npr. predviđanje epidemija na temelju podataka koji cirkuliraju u društvenim mrežama (4).

Osiguravanje dostupnosti izvornih podataka vodi na novi koncept u globalnoj znanosti, na tzv. *podatkovnu otvorenost* (engl. *Open data concept*) i *otvorenu znanost* (engl. *Open science*).

Inicijativa je došla od četiriju vodećih organizacija koje predstavljaju tzv. globalnu znanost: Međunarodni savjet za znanost (*International Council for Science – ICSU*), Među-akademijsko partnerstvo (*Inter Academy Partnership - IAP*), Svjetska akademija znanosti (*The World Academy of Science - TWAS*) i Međunarodni savjet za društvene znanosti (*The International Social Science Council – ISSC*).

U proljeće 2016. godine Akademija medicinskih znanosti Hrvatske dobila je informaciju o sporazumu tih četiriju organizacija vezanom uz inicijativu *Open Data in a Big Data World* (ODBDW) s molbom da ga Akademija podrži (5). „Podržati sporazum“ značilo je popuniti upitnik (6) u kojemu je trebalo navesti što je Akademija dosada napravila u vezi s principima navedenim u Sporazumu i što bi u budućnosti mogla i planirati provesti.

Cilj ovog rada je koncept podatkovne otvorenosti približiti dionicima u biomedicinskim znanostima kako bi se poduzele radnje sa svrhom daljnjeg razvijanja globalnog koncepta uključivanjem hrvatskih biomedicinskih znanstvenika u aktivnosti vezane uz otvorenu znanost.

Što sadrži Sporazum?

Sporazum definira podatkovnu otvorenost kao koncept, kao ideju prema kojoj

izvjesni podatci trebaju biti dostupni za uporabu i ponovljeno objavljivanje bez ikakvih ograničenja poput autorskih prava, patenata ili nekih drugih mehanizmima kontrole (7), a podrazumijeva otvoreni pristup koji se odnosi u prvom redu na *podatke iz znanstvenih istraživanja koja su financirana javnim novcem*, posebice istraživanja od posebne važnosti za aktualne globalne izazove. Otvorenost pristupa podacima smatra se temeljnim preduvjetom za rigorozno provođenje znanstvenih istraživanja i maksimiziranje javne koristi od aktualne podatkovne revolucije.

Podatkovna otvorenost je rastući trend. Bez obzira radi li se o podatkovnoj masi ili ne, prioritet je osigurati integritet i znanstvenu vjerodostojnost. Podatkovna otvorenost je preduvjet za pouzdan način stjecanja znanja (5) jer dozvoljava proces samokorekcije (engl. *self-correction*) već samim time što objavljene rezultate može provjeriti svatko pozivajući se na podatke iz kojih su rezultati proizašli. Otvorenost podataka se odnosi jednako na tehnički aspekt kao što je to i s pravnim. Tehnički, otvoreni, dakle, vidljivi i uporabljivi podatci mogu se obraditi i na druge načine osim onih koje su primijenili istraživači koji su rezultate i objavili i, potencijalno, drukčije interpretirati. Naime, kaže se da je „sunčeva svjetlost najbolje dezinfekcijsko sredstvo“ – dakle, u znanosti je nužna transparentnost. Kad se govori o otvorenosti, misli se na *inteligentnu otvorenost*.

Inteligentna otvorenost podrazumijeva da:

- podatke se može naći putem weba, tj. otkriti da uopće postoje (engl. *discoverable*)
- može im se pristupiti elektronički, s računala, putem interneta (engl. *accessible*)
- mora biti dovoljno informacija o relevantnosti podataka za specifično istraživanje, tj. objašnjenje ili opis

okolnosti pod kojima i radi čega su se podatci prikupili (engl. *intelligible*),

- mora ih se moći procijeniti s obzirom na kompetentnost onoga tko ih je proizveo i njegov mogući financijski interes (engl. *assessable*)
- moraju postojati meta-podatci (podatci o podacima), tj. šifre i način šifriranja, način na koji su pojedini podatci izvedeni tj. izračunati, tehničke karakteristike računalne tehnologije. (engl. *usable*).

Da bi se postigla inteligentna otvorenost, na primjer u hrvatskim javnozdravstvenim istraživanjima, posebno među autorima epidemioloških i kliničko epidemioloških istraživanja kardiovaskularnih i bihevioralnih faktora rizika, znanstvena tematika koja je u nas vrlo dobro pokrivena objavljivanjem u časopisima, ponajprije bi trebalo razviti zajedničku motiviranost autora da u svoju istraživačku metodologiju rutinski ugrade sve dimenzije otvorenosti kada pripremaju svoje projekte. To zahtijeva da se svi principi inteligentne otvorenosti uključe u sva javna znanstvena istraživanja kao kriteriji prihvaćanja projekata, konkretno kao kriterije prihvaćanja projekata u Hrvatskoj zakladi za znanost.

Principi podatkovne otvorenosti

Principi podatkovne otvorenosti (5) uključuju tvrdnje koje idu u prilog otvorenosti ali i one koje izražavaju ograničenja kao i praktične aspekte realizacije podatkovne otvorenosti. Principi se mogu sažeti u tri cjeline:

1. Odgovornost znanstvenika, istraživačkih institucija i sveučilišta, izdavača, agencija koje financiraju istraživanja, strukovnih asocijacija, društava i akademija, te knjižnica, arhiva i repozitorija,
2. Granice otvorenosti – ograničenja vezana uz privatnost, sigurnost, komercijalnu uporabu i sl. Naime, iako su podatci otvoreni, oni posjeduju dozvole

za otvorenost (legalna otvorenost), radi poštivanja intelektualnog vlasništva i sprječava treće osobe od korištenja i redistribucije podataka bez izričite dozvole (8).

3. Podatkovna otvorenost u praksi koja se temelji na citiranju izvora podataka koje je prikupio netko drugi, na maksimalnoj interoperabilnosti, nerestriktivnoj ponovnoj uporabi podataka, i povezivosti (koliko god je to moguće) s drugim podacima.

Zaključno

Biomedicina i zdravstvo su nesumnjivo područja koja pripadaju znanostima koje obiluju podacima (engl. *data-intensive science*). Premda velikim svojim dijelom primjenjuje *pristup vođen znanjem* (engl. *knowledge-driven approach*) gdje se postavljaju hipoteze vezane uz postojeća znanja i onda provjeravaju na prikupljenim podacima, biomedicina sve više usvaja *pristup vođen podacima* (engl. *data-driven approach*). Pristup vođen podacima podrazumijeva podatke kao ishodište, digitalne podatke dobro dokumentirane, primjereno zaštićene, pouzdano čuvane - podatke koje se može integrirati i analizirati primjenom raznih metoda poput statističkih metoda, vizualizacije, umjetne inteligencije, modeliranja i simulacije. Dobro definirani postupci analize takvih podataka omogućavaju prepoznavanje obrazaca ili klastera, generiranje hipoteza, kao i predviđanja određenih događanja.

Dokument *Open Data in a Big Data World* mogao bi biti podloga za rasprave u smjeru politike podatkovne otvorenosti u području biomedicine i zdravstva u Hrvatskoj. Kako može pomoći AMZH? U suradnji s Pravnom, Tehničkom i Šumarskom akademijom trebala bi osmisliti kako promovirati načela Sporazuma, kako privući pažnju znanstvenika i znanstvenih institucija te ukazati na važnost podatkovne otvorenosti. Jedan od načina bi svakako bio prijevod Sporazuma i upoznavanje članova akademija i

znanstvenih institucija s njegovim sadržajem, zatim organiziranje rasprave i upoznati širu znanstvenu javnost sa zaključcima, čime bi se Hrvatska mogla pridružiti svjetskim trendovima u globalnoj znanosti.

Literatura

1. What is the Digital Revolution . Available at: https://www.itu.int/net/wsis/basic/faqs_answer.asp?lang=en&faq_id=42 , pristup 9. prosinca 2017.
2. Beyond Volume, Variety and Velocity is the Issue of Big Data Veracity. Available at: <https://insidebigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/> , pristup 9. prosinca 2017.
3. Dusheck J. Experts: Integrating diverse kinds of data key to precision health. Inside Stanford Medicine 2017;9(11):4. Available at: <http://sgcc.stanford.edu/content/dam/sm-news/images/2017/06/061217.pdf> , pristup 9. prosinca 2017.
4. Eysenbach G. Infodemiology and Infveillance: Framework for an Emerging Set of Public Health Informatics Methods to Analyze Search, Communication and Publication Behavior on the Internet. J Med Internet Res. 2009 Jan-Mar; 11(1): e11. doi: 10.2196/jmir.1157.
5. Open Data in a Big Data World. An international Accord. Available at: <https://www.icsu.org/publications/open-data-in-a-big-data-world> , pristup 9. prosinca 2017.
6. Endorse the international accord: Open Data in a Big Data World . Available at: <https://docs.google.com/forms/d/e/1FAIpQLSdVzLmyLnwKIORYebAHgQngHpyRYoeUinSQSdDDI-EL6WUj2A/viewform> , pristup 9. prosinca 2017.
7. The Free Dictionaire by Farlex. Available at: <https://encyclopedia.thefreedictionary.com/open+data> , pristup 9. prosinca 2017.
8. The Open Data Handbook. Kako učiniti podatke otvorenima. Available at: <http://opendatahandbook.org/guide/hr/how-to-open-up-data/> , pristup 9. prosinca 2017.

Summary

Open Data in the Light of Digital Revolution

Josipa Kern, Jadranka Mustajbegović, Biserka Bergman Marković, Inge Heim, Silvije Vuletić
Croatian Academy of Medical Sciences, Zagreb, Croatia

In 2016, the Croatian Academy of Medical Sciences received the information on agreement *An international accord* of four international science organizations (International Council for Science – ICSU, Inter-Academy Partnership - IAP, The World Academy of Science, and The International Social Science Council - ISSC) with an appeal to support it. This agreement defines the *open data concept* that certain data should be free for use and re-use without any limitations such as copyrights, patents or some other control mechanisms. Primarily, the open access refers to data from scientific research funded by public funds, especially research being of particular relevance to global challenges. The open data access is considered to be a fundamental prerequisite for rigorous implementation of scientific research and maximizing public benefit from the current digital revolution.

The accord proposes 12 principles to guide the practice and practitioners of open data, focused on the roles played by scientists, publishers, libraries and other stakeholders, and on technical requirements for open data. It also assesses the “boundaries of openness”.

Keywords: open data; open science