

J. Macan*

Fakultet kemijskog inženjerstva i tehnologije
Sveučilište u Zagrebu, Marulićev trg 19, 10 000 Zagreb

Otvorena budućnost

I ma već nekoliko godina da se promovira dijeljenje istraživačkih podataka, bilo prilikom objave rada ili kao samostojeće skupove podataka (engl. *dataset*). Trenutačno je to dijeljenje dobrotoljno, no slutim da će u bližjoj budućnosti postati obveza, uz argument da javno financiranje povlači javnost tako stečenih rezultata. S druge strane, znanstveni časopisi objavu istraživačkih podataka mogu prigrliti kao još jedan administrativni filter koji će smanjiti pritisak na urednike i doprinijeti kvaliteti i citiranosti objavljenih radova. Na prvi pogled, tom se pristupu nema što zamjeriti. Dostupnost neobrađenih podataka olakšala bi uspoređivanje vlastitih istraživanja s objavljenima, povezivanjem rezultata više istraživanja moglo bi se doći do novih znanstvenih spoznaja, a detektiranje lažiranih ili prilagođenih rezultata bilo bi znatno lakše. Znanost financirana javnim novcem bila bi dostupna javnosti i svim znanstvenicima, što se već intenzivno potiče objavljivanjem radova u otvorenom pristupu (engl. *open access*). Nesputana znanost dovela bi do inovacija i napretka. Ima li kakav oblačak na tom svijetlom obzoru?

Povijesna i nedavna iskustva govore da nije sve tako jednostavno, kako je već vezano uz otvoreni pristup primijetio N. Raos:¹ otvorenu znanost treba znati interpretirati, te dostupnost podataka baš svakome može i hoće donijeti više štete nego koristi. Ista zamjerka vrijedi i za javno dostupne znanstvene podatke, no nije jedina. Iduća je olakšano plagiranje, tj. krađa tuđih podataka i podmetanje pod svoje, ili pak nedopušteno i neželjeno korištenje otvorenih znanstvenih podataka u komercijalne, vojne i ine svrhe. Realno, to nije toliki problem – tko je voljan činiti išta od toga, može to napraviti i na temelju publiciranih radova, a vještii prevaranti od pamtivijeka su znali izvući povjerljive informacije od svojih sugovornika.

Mislim da je najveći problem s otvoreno dostupnim podacima njihova korisnost, odnosno primjenjivost. Da bi podatci bili korisni drugim znanstvenicima, uz čiste setove podataka trebaju biti dostupni i svi podatci o načinu njihova dobivanja, tj. kontekst.² Većina znanstvenika dobar dio tog konteksta drži u glavi,³ u laboratorijskoj teci te rasuto na raznim računalima vezanima uz opremu upotrijebljenu u istraživanju. Sistematizacija tih informacija u čitljivi oblik uzela bi barem onoliko vremena i truda koliko pisanje znanstvenoga članka, ako ne i više. Dodatno je to opterećenje za znanstvenike koji su, barem kod nas, ionako preopterećeni administriranjem svake vrste. Stoga je vrlo vjerojatno da će većina otvoreno dostupnih podataka, ako će njihovo dijeljenje biti obvezatno, jednostavno biti kopije podataka primijenjenih u pisanju rada bez daljeg konteksta, zadovoljivši time formu, ali ne i cilj otvorenog pristupa. Da bi se to izbjeglo, trebat će znanstvenicima pružiti uhodanu administrativnu podršku i razrađeni pristup kako podatke pripremiti da budu i ostanu čitljivi zainteresiranoj publici. Možda bi trebalo krenuti od edukacije znanstvenika o koristi koju će oni sami imati od odgovarajuće sistematizacije istraživačkih podataka.

Naime, slijedeći primjere drugih fondova koji financiraju znanost, Hrvatska zaklada za znanost najavljuje da će idući ciklus projektnih natječaja od prijavitelja tražiti plan upravljanja podacima.⁴ Opis dostupan na stranicama HRZZ-a očekivano je birokratski pisan, dok je priručnik koji je pripremio Sveučilišni računski centar⁵ demotivirajući za nekoga tko se time do sad nije bavio, tj. potencijalnog voditelja projekta. Mislim da bi svim "dionicima" ponajprije trebalo objasniti izvorni razlog za "upravljanje" podacima, a to je da podatci budu razumljivi i iskoristivi suradnicima

na jednom istraživanju ili projektu te da se neobjavljeni rezultati rada doktoranada ne izgube njihovim odlaskom.⁶ Zagovornici upravljanja podacima i njihova javnog dijeljenja pozvat će se i na trajnost plodova istraživanja po završetku nekog projekta/doktorata, no mišljenja sam da će se vrijedni plodovi ionako objavitii, a za one druge da nije nikakva šteta što će propasti.

To me vodi na standardni knjižničarski i arhivarski problem trajnosti podataka – iako digitalna kopija može biti trajna, format u kojem se pohranjuju podatci može lako zastarjeti. Stoga bi za dugoročnu dostupnost i primjenjivost znanstvenih podataka bili potrebni standardi. Ne sumnjam da će do takvih standarda doći, ali pitam se koliko je sve to vrijedno truda u ovo doba sveopće inflacije znanstvenih radova i znanstvenih podataka. Koliko će uopće znanstvenika pregledati pojedini set podataka, ma koliko dostupan i ureden bio? Već je postalo nemoguće pratiti sve publikacije u prilično uskom istraživačkom području, kamoli proučavati njihove podatke. S druge strane, treba također održavati repozitorije podataka. Široka dostupnost podataka na internetu nas je razmazila da se sve uvijek može naći, no desetljeća iskustva pokazuju da cijeli dijelovi interneta mogu jednostavno nestati zbog manjka održavanja, pri čemu se podatci zauvijek gube ako nisu bili drugdje pohranjeni.

U konačnici, mišljenja sam da nas u budućnosti čeka mnogo birokratskog ali u krajnosti uzaludnog posla oko upravljanja podacima. Možda bude olakšan formiranjem stručnih službi koje će obavljati zamorni dio posla (izradu dokumentacije i metapodataka, osmišljavanje sheme imenovanja datoteka, olakšavanje pohrane podataka itd.). Možda bude motiviran i mrkvom uračunavanja istraživačkih podataka među publikacije za napredovanje,⁷ a ne samo batinom nužnog uvjeta za dobivanje projekta. Možda u nekom udjelu i olakša istraživanje budućoj generaciji znanstvenika. Ali tražit će neproporcionalan trud i ulaganje za ostvareni učinak. Iskreno se nadam da sam u krivu: lijepo je biti pesimist, možeš se samo ugodno iznenaditi.⁸

Literatura

1. N. Raos, Dostupnost informacija, Kem. Ind. 70 (3-4) (2021) 187–191.
2. S. Williams, Mind the gap in data reuse: Sharing data is necessary but not sufficient for future reuse, URL: <https://blogs.lse.ac.uk/impactofsocialsciences/2014/03/28/mind-the-gap-in-data-reuse/> (pristupljeno 28. 1. 2022.).
3. M. Gladwell, The Social Life of Paper, The New Yorker, 25. ožujka 2002., URL: <https://www.newyorker.com/magazine/2002/03/25/the-social-life-of-paper> (pristupljeno 28. 1. 2022.).
4. URL: <https://hrzz.hr/plan-upravljanja-istrazivackim-podacima-za-projekte-hrvatske-zaklade-za-znanost/> (pristupljeno 28. 1. 2022.).
5. D. Celjak, I. Dorotić Malič, M. Matijević, Lj. Poljak, K. Posavec, I. Turk, Istraživački podaci – što s njima?: Priručnik o upravljanju istraživačkim podacima, SRCE 2020, URL: <https://urn.nsk.hr/urn:nbn:hr:102:810090> (pristupljeno 28. 1. 2022.).
6. D. Barron, How Freely Should Scientists Share Their Data?, URL: <https://blogs.scientificamerican.com/observations/how-freely-should-scientists-share-their-data/> (pristupljeno 28. 1. 2022.).
7. S. Titz, The long march to open science, URL: <https://www.snf.ch/en/DrWpaAhhxHhZ05cf/news/news-160902-horizons-the-long-march-to-open-science> (pristupljeno 28. 1. 2022.).
8. B. Franklin, URL: <https://www.goodreads.com/quotes/566352-i-d-rather-be-a-pessimist-because-then-i-can-only> (pristupljeno 28. 1. 2022.).

* Prof. dr. sc. Jelena Macan
e-pošta: jmacan@fkit.hr