

Z ANALIZO PODATKOVNIH BAZ DO TRAJNIH INFORMACIJ O ZNANOSTI

ODKRIVANJE RAZISKOVALNEGA POLJA

Helmut M. Artus

GESIS-IZ
Sozialwissenschaften, Bonn

Kontaktni naslov:
helmut.artus@gesis.org

Izvleček

Podatkovne baze so več kot le registri bibliografskih informacij, saj vsebujejo tudi informacije o zgodovini raziskav. Raziskave podatkovnih baz zajemajo pet sestavin: upravljanje, uporabo, analizo, zgodovino/sociologijo in metodologijo/epistemologijo. Razlikujemo strukturalni in sociološki pogled na podatkovne baze. Primer razvoja podatkovne baze SOLIS ilustrira pomen zgodovine podatkovne baze. Sociologija podatkovne baze nam odkriva družbeno pogojenost podatkov, pri čemer so vplivni tudi odnosi v samih raziskovalnih organizacijah. Več od statistične pove epistemološka analiza podatkovnih baz, ki lahko razkrije, da sploh ne gre za analizo neposredne znanosti, ampak za analizo že presejanih in indeksiranih podatkov iz drugih podatkovnih baz. Na ta način se "mehki" podatki pretvarjajo v "trde", ki so v resnici umetna dejstva (artefakti). Zato je preiskovanje podatkovnih baz brez intelektualnega preverjanja tvegano. Raziskovalci podatkovnih baz morajo delovati interdisciplinarno in poleg "domače" discipline obvladovati tudi informatiko podatkovnih baz, sicer v njih ne bodo mogli razločiti trajnih informacij.

Ključne besede

podatkovna baza, analiza podatkovnih baz, zgodovina podatkovnih baz, sociologija podatkovnih baz, SOLIS

Abstract

Databases are more than just registers providing bibliographic information because of the research history they contain. Database research involves five areas: management, usage, analysis, history/sociology and methodology /epistemology. A distinction is made between a structural and sociological aspect of databases. The example of SOLIS database development illustrates the significance of database history. Sociology reveals to what extent data is socially conditioned, whereby the influence of the relations within research organisations is of importance as well. An epistemological database analysis tells more than a statistical analysis as it may reveal that not direct science but rather pre-screened and indexed data retrieved from other databases is analysed. In this way "soft" data is becoming "hard" data or, actually, artefacts. Consequently, database search without intellectual control involves risks, and if the intellectual control is not exerted in database analysis either, this can only result in artefacts. Database researchers must have an interdisciplinary approach and they must be familiar not only with their 'own' discipline but also with database informatics, or they will not be able to distinguish between permanent and other information within databases.

Keywords

database, database analysis, history of databases, sociology of databases, SOLIS

UVODNA OPOMBA

Da bi bile stvari od samega začetka jasne: moj prispevek je glas **za** raziskovanje podatkovnih baz in **ne proti**, čeprav bo včasih tako videti. Toda podpirati nekaj ne pomeni nujno odpovedati se vsakršni kritiki ali dvomom. Moja kritika želi biti pozitivna in konstruktivna.

UVOD

Ne glede na to, ali v znanstveno literaturo verjamemo ali ne, je dejstvo, da aktualna raba te literature že nekaj časa upada (z izjemo t. i. citatne klasike). Kot logično posledico lahko predpostavimo tudi "zastarelost" informacij, vsebovanih v podatkovnih bazah.

Podatkovne baze so seveda veliko več kot le registri bibliografskih informacij, saj so obenem tudi sistemizirane in pametno strukturirane zakladnice informacij o zgodovini znanosti: o inovacijah in spremembah, o razvoju in usmeritvah, pa tudi o socialnih strukturah, sociokognitivnih mrežah znanstvenikov itd. Uporaba takšnih informacij za analitične namene prispeva k trajnosti podatkovnih baz.

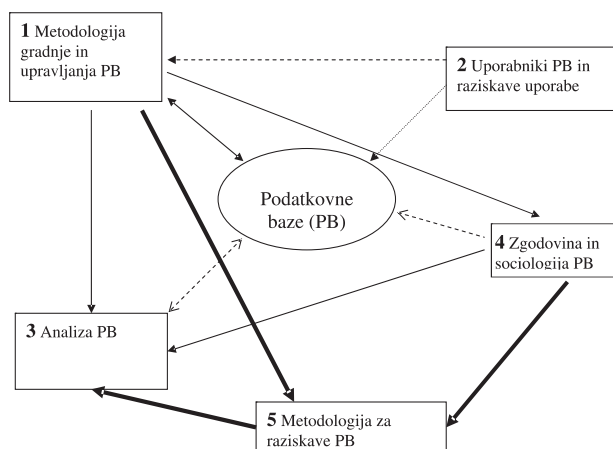
Da bi prišli do veljavnih podatkov, nam ne zadoščajo preproste statistične obdelave. Raziskave podatkovnih baz, kot jih vidimo, so kompleksni in predvsem interdisciplinarni projekti. Obstaja pet različnih raziskovalnih področij, ki jih uvrščamo pod pojem “raziskave podatkovnih baz”, kar opisujemo v tem prispevku.

KAJ SO RAZISKAVE PODATKOVNIH BAZ? KRATEK PREGLED

Kot smo že omenili, raziskave podatkovnih baz zajemajo pet različnih sestavin ali raziskovalnih področij (gl. sliko 1):

- Prva sestavina se nanaša na metodologijo gradnje in na upravljanje podatkovnih baz. Menimo, da je to najmanj pomemben del, ki zajema vsakodnevno opravilo slehernega človeka v vseh inštitutih, ki imajo opravka s podatkovnimi bazami. Nadaljnje razlage sploh niso potrebne.
- Druga sestavina je tradicionalno polje, povezano z javnimi podatkovnimi bazami. Čeprav ga običajno označujemo kot raziskave uporabnikov, zajema tudi raziskave o uporabnosti podatkovnih baz, običajni rabi in včasih tudi o njihovi koristnosti. Tudi tukaj gre za trivialnost. Treba pa je omeniti, da je kakovost kot takšna nekaj več od uporabnosti. Dobrodošlo je tesnejše sodelovanje z empirično sociologijo.
- Tretja in morda najbolj znana sestavina se imenuje analiza podatkovnih baz, kar pomeni, da vzamemo bibliografske podatke¹ kot podatkovno bazo za kvantitativne analize znanstvenega polja z namenom, da bi odkrili njegovo kognitivno in družbeno strukturo, predstavljeno v realnem času. Vprašanja, ki jih lahko obravnavamo v okviru analiz podatkovnih baz, so:

- kognitivna struktura poddisciplin,²
- raziskovalne mreže,³
- kronološki razvoj znanstvenih konceptov,⁴
- spremembe običajev znanstvenega publiciranja, kot jih kažejo komparativne analize tiskanih in online časopisov,⁵
- povezovanje podatkov iz empiričnih raziskav s pomočjo anket s podatki iz podatkovnih baz,⁶
- tematska struktura univerzitetnega poučevanja,⁷
- prenos znanja iz raziskav v univerzitetno poučevanje s povezovanjem različnih podatkovnih baz.⁸



Slika 1: Sestavine raziskovanja podatkovnih baz

- Četrto sestavino, ki jo postavljamo v jedro tega prispevka, tvorita dve področji: **zgodovina** podatkovnih baz z vsemi instrumenti indeksiranja ter pravili, zunanjimi vplivi ipd.; in **sociologija** podatkovnih baz, nanašajoča se na pravila in načine izvajanja, kar vključuje družbene strukture in družbene aktivnosti, družbene spremembe in družbene konflikte, konformizem in odklone ter še mnogo drugega.
- Peta in zadnja sestavina je metodologija analiziranja podatkovnih baz, ki – do določene mere – izhaja kot metodološka konsekvence iz dognanj zgodovine in sociologije podatkovnih baz v povezavi z vedenjem o metodologiji gradnje in upravljanja podatkovnih baz.

V tem prispevku se osredotočamo na četrto in peto sestavino, tj. na zgodovino in sociologijo ter tudi na metodologijo.

KAJ JE PODATKOVNA BAZA? DVE RAZLIČNI PERSPEKTIVI

Če definiramo koncept “podatkovne baze”, običajno uporabimo naslednji izraz: “Računalniška podatkovna baza je strukturirana zbirka zapisov ali podatkov, ki je shranjena v računalniškem sistemu.”⁹

To nedvomno drži, vendar je nedvomno to le ena plat medalje. Lahko podamo tudi povsem drugačne definicije, ki se glasijo približno takole: “Podatkovna baza je strukturirana zbirka podatkov, ki jo praviloma zgradi večje ali manjše število ljudi v različnih pogojih dela; pogosto ti pogoji niso prav nič idealni, ustrezni ali vsaj normalni.” ipd.

Doslej razen nas ni še nihče podal takšne definicije. Toda če se pogovarjate z ljudmi, ki so vključeni v gradnjo po-

datkovnih baz, boste še in še poslušali zgodbe o tej človeški plati podatkovnih baz, še in še se bodo pritoževali in dajali kritične pripombe.

Očitno se pogledi teoretikov in praktikov na podatkovne baze razlikujejo. V naslednji tabeli skušamo predstaviti grob in seveda nepopoln pregled teh dveh perspektiv na podatkovne baze. Gre za dva povsem različna pogleda:

- Prvi, ki ga imenujemo strukturalistični pogled, obravnava podatkovne baze kot formalno strukturirana polja z vsebinami, kreiranimi po natančnih pravilih. Pri tem pogledu je samoumevno, da so podatki nesporna in nedvoumna osnova za kvantitativne statistične analize.
- Drugi, ki ga imenujemo sociološki pogled, se ozira na ljudi, ki izvajajo ta pravila in ustvarjajo podatkovne baze, ter upošteva tudi družbene procese ustvarjanja teh podatkovnih baz in še pogoje, pod katerimi se vse to dogaja in pod katerimi vsi ti ljudje delajo. Pri tem pogledu so podatkovne baze kot človekove stvaritve nepopolne, zato niso samoumevna osnova za kvantitativne analize.

Strukturalistični pogled	Sociološki pogled
Cilj (kar naj bi bilo)	Izvedba (kar dejansko je)
Idealne razmere za delo po pravilih indeksiranja in z drugimi instrumenti	Realno stanje; morebitno pomanjkanje kompetenc v nekaterih disciplinah
Nehistorično – neupoštevanje sprememb in razvoja	Historično – “filozofija” ali “ideologija” PB, politični vplivi, razvoj in spremembe vsebin, pravil indeksiranja in instrumentov – priročnikov, tezavrov, klasifikacijskih in drugih deskriptorjev ter pravil njihove uporabe in kontrole; zaposleni in njihova fluktuacija itd.
Deterministična pravila	Uporaba pravil kot izrazito družben, psihološki in kognitiven proces razumevanja (vprašljivost pravil in dokumentov samih), interpretacije (pravil in dokumentov), selekcija in iskanje (deskriptorji vseh vrst) in odločanje (Je moj deskriptor ustrezen? Je zadosten? Morda preveč odvisen od okoliščin?), kar vse pomeni, da je indeksiranje izrazito kompleksen, od naključja odvisen in do določene mere tvegan proces.
Enoznačnost formalnih struktur	Dejanska večznačnost kljub formalizacijam

Strukturalistični pogled	Sociološki pogled
Družbena normiranost vseh aktivnosti	Dejanska aktivnost, ki vključuje odklone, konflikte, rivalstvo ipd.
Idealni delovni pogoji	Dejanske delovne razmere s časovnimi pritiski, konkurenčnimi prioritetami itd.
Predpostavljeni idealni igralci	Realni (nepopolni) igralci Ni enega samega igralca, ampak jih je mnogo z različnimi mnenji in kvalifikacijami, različnimi interesi in motivi, različnimi budžeti, različno stopnjo pripadnosti (glede na različne inštitute) Ne gre le za številčnost igralcev z vsemi njihovimi razlikami, ampak mnogi niti ne delajo istočasno, pa so vseeno v interakciji (in vplivajo druga na drugega: tekmujejo med seboj ter prihajajo v rivalske in konfliktno odnose itd.) Pripadajo različnim institucijam s posebnimi interesi, politikami, strankami, prioritetami ipd. (v primeru SOLIS in SOFIS je bilo skozi celo obdobje zajetih okoli 20 sodelujočih inštitutov, praviloma pa vsaj 10 naenkrat).

Tabela 1: Dvojna narava podatkovnih baz

Kot smo že omenili, seznam s tem še zdaleč ni izčrpan.

Zaradi tega je treba pred uporabo podatkovnih baz kot osnove za kvantitativne analize te iste podatkovne baze podvreči raziskovanju kot objekte, da bi se izognili tveganju za napačne povezave in sklepanja. Takšno grobo primerjanje jasno pokaže vsaj tri posledice:

- **Prva posledica** je, da potrebujemo **zgodovino podatkovne baze** (ki jo je treba vedno brati kot zgodovino vsake posebne podatkovne baze). Za sleherno podatkovno bazo, ki jo uporabljamo kot osnovo za kvantitativne ali statistične analize, moramo sestaviti kar se da podrobno zgodovino njenega razvoja in vseh pomembnejših sprememb bodisi v pogledu vsebine bodisi v pogledu klasifikacijskih kod ali metodoloških deskriptorjev, tezavrov in pravil obnašanja.
- **Druga posledica** je, da potrebujemo **sociologijo podatkovne baze**, da bi lahko razumeli vplive, ki obnem formirajo in deformirajo socialni sistem, kot pod-

poro gradnji podatkovne baze in družbeno akcijo vseh ljudi, udeleženih pri kreiranju podatkovne baze, kot tudi pravil, ki jih pri tem uporabljajo. Ta sociologija podatkovne baze naj bi bila splošna teorija in obenem njena aplikacija na primerih posebnih podatkovnih baz.

- Je pa še **tretja posledica**, ki zadeva **metodologijo in epistemologijo**, določeneje t. i. problem empirične baze (nem. *Basisproblem*). Vprašanje je, ali lahko podatke, ki so bili ustvarjeni in urejeni za povsem drugačen namen, brez vsakršne dodatne utemeljitve uporabimo kot osnovo za nekakšne empirične analize.

ZGODOVINA PODATKOVNE BAZE

Kot smo že prej omenili, potrebujemo kar podrobno zgodovino razvoja in vseh pomembnejših sprememb podatkovne baze, bodisi njene vsebine, pravil indeksiranja ali vsakvrstnih deskriptorjev.

Vzemimo za primer disciplinarno in bibliografsko vsebino podatkovne baze SOLIS. Nastala je v zgodnjih osemdesetih letih z uradnim in političnim¹⁰ dogovorom, da bo pokrivala vso sociološko literaturo v nemško govorečih državah, tj. v obeh Nemčijah, ki sta v tistem času še obstajali, ter v Avstriji in Švici. Dve leti kasneje je nastala retrospektivna dokumentacija, ki je segala do leta 1972, vendar ni vključevala knjig in sive literature, ki jo je SOLIS sicer dokumentiral od samega začetka. Vključeni so bili časopisni članki in tudi ti ne iz vseh časopisov, pač pa le iz najpomembnejših. Približno v istem obdobju so bile narejene še dodatne bibliografske obdelave, ki so segle do leta 1945, toda brez izvlečkov in le z zelo rudimentarnim indeksiranjem, kar kazi sleherne kvantitativne analize.¹¹

Na ta način so se v treh letih zgodile nič manj kot tri pomembne spremembe glede vsebin, ki jih bibliografija upošteva:

- Vsa sociološka literatura – tj. časopisni članki, članki v antologijah, knjige in siva literatura (neobjavljena raziskovalna poročila) – z izvlečki in podrobnimi opisi, začenši z letom 1980.
- Zgolj članki iz pomembnih socioloških časopisov z izvlečki in podrobnimi opisi za obdobje 1972–1979.
- Knjige, časopisni članki in članki v antologijah brez izvlečkov in s skromnim indeksiranjem za obdobje 1945–1971.

Samo po sebi je umevno, da imajo tolikšne spremembe velikanski vpliv na sleherno kvantitativno analizo podatkovnih baz. Zgodile pa so se še druge pomembne spremembe: Leta 1983 se je vsebina dramatično razširila. Namesto ene same discipline – sociologije – je morala

podatkovna baza SOLIS pokriti literaturo celotnega družboslovja v nemško govorečih državah.

To je bila že četrta pomembna sprememba glede vsebine v treh letih. Vendar je to le majhen delček “zgodovine vsebine”.

Ko je dobil GESIS-IZ naročilo, da pokrije družboslovje v celoti, so bili politični pogoji takšni, da ni smel sam dokumentirati vseh disciplin, pač pa v tesnem pogodbenem sodelovanju s specializiranimi inštituti. Zato je bila podatkovna baza SOLIS nekaj časa kooperativni izdelek najmanj dvanajstih inštitutov.

Potreben pa je bil še trinajsti inštitut, ki pa je hotel biti popolnoma neodvisen in je zavračal pogoje kooperacije. Ob tem je bilo izrecno prepovedano, da bi GESIS-IZ sam dokumentiral neko določeno disciplino – šlo je za politične vede (politologijo). Deset let je trajalo vojno stanje, ki se je končalo s celo serijo tožb in doseglo vrhunec z grožnjo ustavne obtožbe, ki se na srečo ni uresničila. Na neki točki tega procesa je GESIS-IZ prevzel okoli 3.600 dokumentov, ki so bili nekaj mesecev kasneje (maja 1992) sodno izločeni iz podatkovne baze. Neka druga razsodba jih je kakšna tri leta kasneje spet vrnila v SOLIS. V vsem tem obdobju je GESIS-IZ lahko dokumentiral le majhen del objavljene politološke literature, in še to le iz splošnih družboslovnih časopisov, ne pa iz politoloških časopisov.

Vidimo, da vsaj za obdobje dvanajstih ali nekaj več let delež politoloških publikacij v SOLIS nikakor ne more biti indikator znanstvene produkcije te discipline (kar se ugotovi z analizo podatkovne baze), pač pa kvečjemu indikator za mrežo medinstitucionalnega sodelovanja ali nesodelovanja v projektu podatkovne baze SOLIS. Leta kasneje je GESIS-IZ sklenil drugo pogodbo o sodelovanju z nekim inštitutom, ki je sicer zelo primeren, a ima izrazite lastne interese, ki načenjajo dokumentalistično filozofijo SOLIS.

Primer političnih ved je bil najbolj izrazit konflikt ob vseh zaprekah in motnjah v sistemu kooperacije, ki ima ne-hierarhično strukturo. GESIS-IZ je bil v precepu: ni mu bilo dopuščeno dokumentiranje znanstvenih disciplin ali njihovih delov, ki so jih dokumentirali že drugi inštituti, obenem pa ni imel pravice uveljavljati splošna pravila, neobhodna za skupno gradnjo podatkovne baze. Kot nujna posledica razvoja in sprememb vsebine, je bilo potrebno dopolnjevanje ali diferenciacija in specifikacija dobrega dela instrumentov za indeksiranje, tezavrov z vsemi njihovimi kontrolnimi termini vred, celotnih seznamov kod za klasifikacijo in metodoloških deskriptorjev.

Vse to je le en segment zgodovine naše podatkovne baze. Kot kaže primer, je pisna zgodovina podatkovne baze

lahko obenem zgodovina političnega vpliva, institucionalnega sodelovanja in konfliktov ter zgodovina inštituta, odgovornega za podatkovno bazo kot celoto v smislu koordinacije med različnimi inštituti in njihovimi interesi. Med drugim je to tudi zgodovina institucionalne filozofije in filozofije podatkovne baze; socialnih, političnih in finančnih vplivov; števila, kvalifikacij in specializacij osebja podatkovne baze, njihove fluktuacije, tehnične opremljenosti itd. Na koncu, čeprav ne najmanj pomembno, je to lahko zgodovina tehnologije.

Kvantitativne podatkovne baze brez poznavanja vsega tega zgodovinskega ozadja ne vzdržijo tveganja možnih napak ali deformacij in lahko hitro postanejo raziskovalni artefakti.¹²

SOCIOLOGIJA PODATKOVNE BAZE

Takšne zgodovine podatkovnih baz so nujne, vendar niso zadostne. Dopolniti jih mora sociologija podatkovne baze.¹³

Povsem neobdelano raziskovalno polje so analize podatkovnih baz ne zgolj kot formalnih struktur, grajenih na podlagi obveznih priročnikov, tezavrov, klasifikacijskih seznamov, metod itd., ampak tudi kot rezultat aktivnosti in interakcij množic ljudi ter kooperacije ducatov inštitutov z različnimi interesi ali filozofijami, ki včasih povzročajo bolj ali manj resne konflikte, o čemer smo pisali v prejšnjem poglavju.

Rekonstrukcije in empirične analize takšnih kompleksnih sociokognitivnih sistemov morajo biti predmet sociologije podatkovnih baz, ki v povratnem smislu odkrivajo družbeno pogojenost podatkov, pridobljenih iz podatkovnih baz. Zapomniti si moramo, da je to ključna točka sleherne analize podatkovne baze – namreč analiziranje družbene pogojenosti podatka.¹⁴

Vplivi, odločilni za formacijo in deformacijo družbenega sistema, izhajajo iz kooperacije z drugimi inštituti, ki prispevajo posebne dele, nanašajoče se na znanstvene discipline ali poddiscipline, k skupni podatkovni bazi. V obratni smeri postavljajo ti inštituti zahteve, ki morda niso kompatibilne s splošno filozofijo in usmeritvijo projekta podatkovne baze, kar smo videli prej.

Morda najzanimivejše in najpomembnejše teme takšne sociologije podatkovnih baz so sistematične analize in teoretične razlage družbenih akcij znotraj inštitutov, ki kreirajo določeno podatkovno bazo, kot tudi odnosov med njimi.

Kot empirična veda se sociologija ukvarja s pravili delovanja, njena bistvena vsebina pa je delovanje samo,

seveda v primerjavi in razlikovanju pravil od norm, ki opredeljujejo to delovanje. Zaradi tega sociologijo zanima normativna skladnost ali odstopanje delovanja kot tudi družbeni pogoji tega odstopanja, odsotnosti norm ali anomalije.

Vprašanje donosa med normami in pravili na eni strani in njihovo dejansko uporabo na drugi strani se ne omejuje zgolj na področje "normalnega" družbenega delovanja, pač pa zajema tudi dokumentalistično dejavnost samo. V tem so zajeti tudi pogoji, determinante in možnosti družbene in dokumentalistične dejavnosti kot tudi kontrolni mehanizmi, vključno s sankcioniranjem in načini ter stopnjo institucionalizacije.

Tukaj se oblikujejo naslednje tematike:

- Socialna struktura oddelkov, odgovornih za gradnjo podatkovnih baz, vključujoč kreiranje, razvoj, nadzor in ažuriranje priročnikov, pravil indeksiranja in drugih instrumentov.
- Kontrola kakovosti, povratne informacije, komunikacije in interakcije – stopnja, oblika in kakovost sodelovanja med kolegi.
- Socialni nadzor, sankcije, konflikti in upravljanje s konflikti, bodisi družbenimi bodisi kognitivnimi konflikti pri interpretiranju in aplikaciji pravil, norm in instrumentov. V primeru zunanjih sodelavcev je vprašanje nadzora, kritike in usposabljanja na delovnem mestu še težje kot v primeru nadzora znotraj hiše.
- Delovne razmere, vključujoč vprašanje, ali je dovolj kolegov, ki so ne le kvalificirani na sploh, ampak imajo tudi nujne specializacije za točno tista predmetna področja, za katera odgovarjajo. Tukaj je tudi mesto za vprašanje o pridobivanju in fluktuaciji takšnih specialistov.

A tudi v primeru, ko je inštitut dosegel visoko stopnjo kvalificiranosti in specializiranosti, nadzora in kognitivnega skladja ter uspeva vse to institucionalizirati, ostaja odprto še eno pomembno vprašanje. Do katere mere lahko delovanje – družbeno kot tudi dokumentalistično delovanje – uravnavamo z normami in pravili? V večini primerov smo zadovoljni, če dejanska aktivnost ni v nasprotju s pravili. Naj to pojasnimo s čisto enostavnim primerom:

- "V primeru, da se zasveti rdeča luč, pritisni tipko A!" To je deterministično pravilo, ki na nedvoumen način povezuje nedvoumen dogodek (prižgana rdeča luč) z zahtevano nedvoumno reakcijo "pritisni tipko A!"
- "Opiši najpomembnejše vsebine določene publikacije!" pa je nedeterministična zahteva. Gre bolj za usmerjevalno idejo za proces indeksiranja kot pa za navodilo za konkretno aktivnost. Nekdo bo razumel njen pomen bolj načelno, ne bo pa mogoče na tej

podlagi izpeljati konkretne ali nedvoumne aktivnosti v zvezi z indeksiranjem.

Kadar je deterministična akcija le iluzija, ne moremo pričakovati enotnega indeksiranja, ampak se bomo morali zadovoljiti že s tem, da bo pretežni del primerov zajet v "horizontu soglasja" skupine za indeksiranje v določenem inštitutu. To pomeni, da tudi analitik podatkovne baze ne sme pričakovati, da bodo ti podatki generirani po jasnih determinističnih pravilih, ampak lahko v najboljšem primeru pričakuje le približne podatke. Neizogibno je razhajanje in razpršenost podatkov.

Morebiti naletimo na problem, ki ga imenujemo odnos negotovosti. Znanost teži k natančnosti, ki pa je pri analizah podatkovnih baz težko dosegljiva. Bolj ko so natančna in diferencirana pravila indeksiranja ter instrumenti, težje in bolj zapleteno jih je uporabljati, višjo kvalifikacijo in specializacijo potrebujemo in več časa gre za dokumentiranje. Z drugimi besedami – če smo pošteni (in realni, kar je isto) – večja ko je natančnost, večja je možnost napak in nevarnost, da nastane raziskovalni artefakt.

Vsa ta dejstva in dejavniki, ki imajo opraviti z zgodovino in sociologijo podatkovnih baz, njihova pravila in instrumenti ter njihove aplikacije imajo posledice za vse analize podatkovnih baz. Nekatere od teh posledic nastopijo takoj, nekatere se pokažejo le v kontekstu posebnih analiz. Vsekakor je treba upoštevati oboje. Povsem jasno moramo razumeti, da vsi ti pojavi niso niti pomanjkljivosti, niti napake podatkovnih baz in njihovih družbenih in organizacijskih osnov. Katera koli družbena akcija se dogaja, se lahko dogodi le pod družbenimi pogoji in z družbenimi posledicami. Človekova dejavnost brez človeka ni možna. Torej so tudi vse posledice človeške (in družbene), kar pomeni, da so nujne, neizogibne in neizbežne. Vedno, ko sodelujejo ljudje ali inštituti, so takšni pojavi in problemi vsakdanja stvar. S tem se je pač treba sprijazniti. Zato je njihovo obravnavanje pri analizah podatkovnih baz in metodologijah teh analiz neizogibno.

EPISTEMOLOŠKO VPRAŠANJE

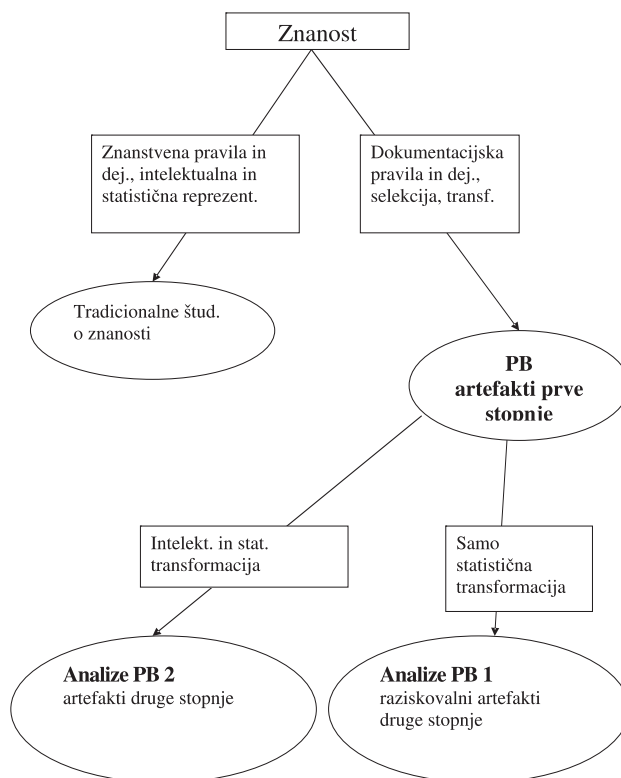
Izbrano in zelo potrebno raziskovalno polje zahteva še veliko več pozornosti in zanimanja, kot ga je bilo deležno doslej. Če sploh na kaj, se raziskovalci osredotočajo kvečjemu na metode in metodologije analiziranja podatkovnih baz, kar pomeni večinoma na uporabo statističnih orodij.

Resnično zanimiva (in pomembna ter komplicirana) vprašanja pa so tista, ki se nanašajo na generiranje podatkov, njihovo veljavnost, metodološke predpostavke veljavnosti in njihovo primerjavo z družbeno stvarnostjo.

Kot smo povedali že prej, se vprašanje glasi: Ali podatki, ki so ustvarjeni in urejeni za povsem drug namen, lahko brez dodatnih utemeljitev služijo kot podlaga za nekakšno empirično analizo?

Problem je, da se analize podatkovnih baz pretvarjajo, da so analize znanosti, čeprav izhajajo iz podatkovnih baz, ki so same artefakti in predstavljajo znanost le pod specifičnimi kriteriji in za specifične namene. Podatki, vsebovani v podatkovnih bazah, so šli skozi proces selekcije in transformacije glede na pravila indeksiranja. Takšna podatkovna baza torej ni resnična slika predmeta – v tem primeru znanosti – ampak je artefakt. Ali bolj natančno: je artefakt prve stopnje.

V postopku analiziranja podatkovnih baz je ta artefakt prve stopnje predmet nadaljnjih procesov selekcije, redukcije in transformacije. Vseeno je, ali so podatki v podatkovni bazi kvalitativni ali kvantitativni – pretvorjeni so v zbir strukturiranih in na koncu digitaliziranih podatkov, iz katerih se ustvari novi artefakt druge stopnje.



Slika 2: Fakti, artefakti in raziskovalni artefakti

Gre za sporno točko obravnave, ali je ta transformacija metodološki ali epistemološki problem, če upoštevamo le posledice. Problem ni ta, da imamo opraviti z artefaktom, ampak da gre obenem za raziskovalni artefakt. Transformacija podatkov podatkovne baze v podatke analiz podatkovnih baz v nobenem primeru ni zgolj formalna transformacija podatkov brez sprememb ali brez vpliva

na vsebino in pomen. Dejansko s to transformacijo "meki" podatki postanejo "trda" dejstva, ki pa so na žalost le umetna dejstva (artefakti).

Da bi to razumeli, moramo imeti vpogled v proces preiskovanja podatkovnih baz, čemu so podatkovne baze sploh namenjene. Ko vnesemo deskriptorje in uporabimo ustrezno preiskovalno logiko, dobimo številne dokumente, ki vsi izpolnjujejo naše iskalne kriterije. Kar sledi, je intelektualno pregledovanje in selekcija resnično relevantnih zadetkov, kar pomeni, da so zgolj v teh primerih uporabljeni deskriptorji šele uspeli opisati jedro publikacije.

Ko izvajamo preiskovanje podatkovnih baz, še vedno pričakujemo, da so tam pomembne razlike glede relevantnosti med deskriptorji, uporabljenimi v dokumentu. Če pa iste deskriptorje uporabimo v analizi podatkovne baze, bodo vsi takoj pridobili enak pomen, ki v odnosu do posameznega dokumenta povzroči enostranost. Površna vsebina postane enako pomembna kot njeno bistvo.

To ni edini vir zastranitev. Enak učinek se pokaže pri klasifikaciji in metodoloških deskriptorjih, pa tudi pri deskriptorjih, ki se sploh ne nanašajo na vsebine, ampak na formalne kategorije, kot so viri (npr. časopisi) ali tipi dokumentov (npr. članek, knjiga, raziskovalno poročilo itd.).

Problem torej ni to, da je podatek iz podatkovne baze vzeti kot osnova za kvantitativno analizo, ampak način, kako je to storjeno.¹⁵ Preiskovanje podatkovnih baz brez intelektualnega preverjanja (nadzora, selekcije) je zelo problematično in ne dosega sodobnih zahtev. Analize podatkovnih baz se z odrekanjem intelektualnemu preverjanju izpostavljajo tveganju, da bodo zgrešile znanstveno realnost in pridelale artefakt. Z drugimi besedami: pot od mehkih dejstev do trdih artefaktov je zelo kratka (angl. *from weak facts to hard arte-facts*).

PRVI "VELIKI PROJEKT" RAZISKOVANJA PODATKOVNIH BAZ

Naša argumentacija temelji na številnih znanstvenih in dokumentalističnih izkušnjah, na teoretičnih (tudi metateoretičnih), pa tudi na empiričnih in praktičnih izkušnjah. Prepričani smo, da je tudi logično pravilna.

Drugo vprašanje je, ali so tudi posledice pravilne in nedvoumne, o čemer ni mogoče odločiti teoretično ali z logično argumentacijo, pač pa le empirično. Morda pa vsi omenjeni učinki negirajo drug drugega, ali pa so tako postranski, da je sleherno razpravljanje o njih brez pomena? Najbolj verjeten rezultat bi lahko bil, da je kakovost formalne transformacije v enem delu sprejemljiva, v drugih delih pa podatki potrebujejo večjo ali manjšo intelektualno predelavo. Toda, kot sem re-

kel: to je vprašanje, o katerem se je mogoče odločiti le empirično.

Na ta način bi moral biti prvi veliki projekt ali delovno naročilo pri raziskovanju podatkovnih baz razjasnitev naslednjih problemskih sklopov:

1. Potrebujemo ločeno, neodvisno in veljavno empirično anketo, ki se nanaša na natančno iste segmente znanosti, kot jih pokriva analiza podatkovnih baz. Ta anketa mora biti izvedena na teoretičnem in metodološkem nivoju sodobnega empiričnega družboslovja in mora zajeti tudi ustrezne primerjave med empiričnimi podatki iz analiz podatkovnih baz. Pri tem mora biti dosledno onemogočena sleherna cirkularnost.
2. Posamična anketa in primerjava nista dovolj. Kot vemo iz preiskovanja podatkovnih baz, veljavnost deskriptorjev močno korelira s posebnimi predmetnimi vsebinami in poddisciplinami. Empirična anketa mora upoštevati te razlike, če naj bodo njeni rezultati res veljavni; biti morajo skratka utemeljitev za nadaljnje postopke preiskovanja podatkovnih baz.

Omenjena razjasnitev bo seveda zelo trdo delo, a je edina alternativa postopku, ki smo ga poimenovali "analiza podatkovnih baz 2" (slika 2).

KVALIFIKACIJA, ZNANSTVENA INTERAKCIJA IN TRANSDISCIPLINARNA KOOPERACIJA

Prej podana razlaga nam pove, da je raziskava podatkovne baze uspešna le v primeru, če je konceptualizirana kot interdisciplinarno podjetje. Kot je razvidno s slike 1, sociologija podatkovnih baz izrazito reagira na metode in metodologijo gradnje podatkovnih baz kot tudi na analize podatkovnih baz. Bližnji odnos med sociologijo podatkovnih baz in med raziskavami uporabnikov o uporabi in koristnosti je tudi očitna. To se nanaša tako na zastavljanje vprašanj, metod in na teoretične pristope.

Ugotavljamo, da morajo imeti raziskovalci, ki delajo na področju raziskovanja podatkovnih baz, podvojeno kvalifikacijo v svoji "domači disciplini" (npr. sociologiji, psihologiji, statistiki, informatiki itd.) in v gradnji podatkovnih baz. Slednja kvalifikacija zahteva več let profesionalnega delovanja na področju gradnje podatkovnih baz in njihovega upravljanja. Le tesna povezava med različnimi sektorji in poglobljeno razumevanje vseh drugih aspektov jamči zanesljivo in veljavno raziskovanje podatkovnih baz in lahko prispeva k trajnosti informacij, vsebovanih v njih.

Reference 1

Izbrane publikacije, ki se nanašajo na eno ali več podatkovnih baz GESIS-IZ in uporabljajo njihove podatke za kvantitativne analize.

Uporabljene baze podatkov so:

SOLIS Družboslovna literatura v nemško govorečih državah
SOFIS (prej FORIS) Družboslovni raziskovalni projekti v nemško govorečih državah

LEHRE Univerzitetni programi v nemškem družboslovju (podatkovna baza je bila ukinjena)

ZEITSCHRIFTEN Družboslovni časopisi v nemško govorečih državah

- [1] **Artus, Helmut M.:** Graue Literatur. Zum informellen Kommunikationssystem der Sozialwissenschaften: Abschlußbericht. Bonn, 1992, 342 pp., 101 tables.
- [2] **Artus, Helmut M.:** The transition from 'grey' to 'white' literature: a study in the communication and publication behaviour of social scientists. pp. 126–144.
V: Kühnhold, W. W.; W. P. Kirchner (eds.): Publications as an integral part of scientific research. Proceedings of the 5th International Conference of Scientific Editors, Hamburg, 1987 June 14–19, Hamburg 1988 (mikrofilm).
Primer kombinacije anketnih podatkov (zadevajočih komunikacijske in publikacijske procese) v povezavi s podatkovno bazo, zadevajočo isti raziskovalni projekt po formalnih kriterijih (število raziskovalcev, trajanje projekta, naročena raziskava, število publikacij itd.)
- [3] **Artus, Helmut M.:** Soziologielehre: eine räumliche und thematische Analyse. pp. 59–86.
V: Artus, Helmut M.; Matthias Herfurth (eds.): Soziologielehre in Deutschland: Lehre, Studium, beruflicher Verbleib, Lehrangebot, Studien- und Prüfungsordnungen. Opladen: Leske&Budrich, 1996.
Primer tradicionalne kvantitativne analize podatkovne baze LEHRE.
- [4] **Artus, Helmut M.:** Old WWine in New Bottles? Developments in electronic information and communication: structural change and functional inertia. pp. 9–16. V: The Grey Journal: an International Journal on Grey Literature 1, No. 1, 2005.
Primer tradicionalne kvantitativne analize podatkovne baze ZEITSCHRIFTEN.
- [5] **Best, Heinrich; Renate Ohly:** Entwicklungstendenzen der deutschsprachigen Soziologie im Spiegel ihrer führenden Fachzeitschriften – Ergebnisse einer Korrespondenzanalyse. pp. 575–592.
V: Best, Heinrich; Brigitte Endres-Niggemeyer; Matthias Herfurth; H. Peter Ohly (eds.): Informations- und Wissensverarbeitung in den Sozialwissenschaften: Beiträge zur Umsetzung neuer Informationstechnologien. Opladen: Westdeutscher Verlag, 1994, 623 pp.
- [6] **Best, Heinrich; Renate Ohly:** From paradigms to eclecticism: thematic profiles of German language core sociology journals 1984–1991. pp. 95–113. V: Bulletin de Méthodologie Sociologique (BMS) März (1994) 42.
- [7] **Binder, Gisbert; Matthias Herfurth:** Quantitative analysis of bibliographic databases. pp. 33–43.
V: Saelen, Kirsti T.; Arnaud Marks; György Rozsa; Karl A. Stroetmann; Vladimir Vinogradov; Jiri Zahradil; Manfred Krause; Harald Koch; Liparit Kiuzadjan (eds.): The role of social science information in knowledge creation: integrative aspects in information, communication and knowledge; proceedings of the Vth ECSSID General Conference, Berlin, GDR, January 22–24 1989; Vol. 2., Wien, 1990, 206 strani.
- [8] **Binder, Gisbert; Matthias Stahl:** Sozialforschung 1990 in den fünf neuen Bundesländern: Ergebnisse einer scientometrischen Analyse der Projektdatenbank FORIS. pp. 561–570.
V: Neubauer, Wolfram; Karl-Heinz Meier (eds.): Deutscher Dokumentartag 1992: Technik und Information. Markt, Medien und Methoden. Frankfurt am M.: Deutsche Gesellschaft für Dokumentation 1993 (DGD-Schrift (DOK-5) 1/93).
Primer tradicionalne kvantitativne analize podatkovne baze, zadevajoče tematiko kot tudi formalne vidike.
- [9] **Binder, Gisbert; Matthias Stahl:** Der thematische Zusammenhang von Forschung und Lehre: eine scientometrische Analyse. pp. 593–612.
V: Best, Heinrich; Brigitte Endres-Niggemeyer; Matthias Herfurth; H. Peter Ohly (eds.): Informations- und Wissensverarbeitung in den Sozialwissenschaften: Beiträge zur Umsetzung neuer Informationstechnologien. Opladen: Westdeutscher Verlag, 1994, 623 pp.
- [10] **Binder, Gisbert; Matthias Stahl:** Der Forschungsbezug von Lehrveranstaltungen in den Sozialwissenschaften. pp. 87–106.
V: Artus, Helmut M.; Matthias Herfurth (eds.): Soziologielehre in Deutschland: Lehre, Studium, beruflicher Verbleib, Lehrangebot, Studien- und Prüfungsordnungen. Opladen: Leske&Budrich, 1996.
Primer za kombinacijo kvantitativne analize in tematske primerjave podatkov, vzetih iz različnih podatkovnih baz, v tem primeru iz SOLIS in LEHRE.
- [11] **Güdler, Jürgen:** Kooperationsnetzwerke in der Forschung. Entstehung, Struktur und Wirkung am Beispiel der Soziologie. Bonn: Informationszentrum Sozialwissenschaften 2003, 238 pp. (Forschungsberichte, Band 5).
- [12] **Güdler, Jürgen:** Dynamik der Medienforschung. Eine scientometrische Analyse auf der Grundlage sozialwissenschaftlicher Fachdatenbanken. Bonn: Informationszentrum Sozialwissenschaften 1996, 136 pp. (Forschungsberichte, Band 1).
Primer analize mreže, temelječ na podatkih o podatkovni bazi, v tem primeru na kooperativnih strukturah in raziskovalnih mrežah.
- [13] **Mutschke, Peter:** Autorennetzwerke: Verfahren der Netzwerkanalyse als Mehrwertdienste für Informationssysteme. April 2004, 49 pp. (IZ-Arbeitsbericht Nr. 32).
Primer analize mreže, temelječe na podatkih o podatkovni bazi in njeni uporabnosti za nadgradnjo informacijskega servisa v družboslovju.
- [14] **Renner, Ilona:** Soziale Kohärenz und Innovativität. Struktureffekte zur Akzeptanz neuer Themen in sozialwissenschaftlichen

Forschungsfeldern, in: Kölner Zeitschrift für Soziologie und Sozialpsychologie, 49, 1997, pp. 74–97.

Reference 2

Druga objavljena in neobjavljena literatura o materiji tega prispevka.

- [15] **Artus, Helmut M.:** Zu einer Soziologie der Datenbank (Diskussionspapier), IZ Sozialwissenschaften; Bonn, Sept. 1992, 21 pp. (neobjavljeno).
- [16] **Artus, Helmut M.:** Science Indicators Derived from Databases: the Case of the Social Sciences. pp. 297–311. V: Scientometrics, Vol. 37, 1996, No. 2.
- [17] **Sahner, Heinz:** Theorie und Forschung. Zur paradigmatischen Struktur der westdeutschen Soziologie und zu ihrem Einfluß auf die Forschung Theorie und Forschung. Zur paradigmatischen Struktur der westdeutschen Soziologie und zu ihrem Einfluß auf die Forschung. (=Beiträge zur sozialwissenschaftlichen Forschung; Bd. 34), Opladen: Westdeutscher Verlag, 1982, 338 pp. Primer popolnoma neuporabne bibliografske podatkovne baze, posebej kreirane za posamičen raziskovalni projekt avtorja.

- 12 Gl. tudi Artus (1992) v seznamu Reference 2 na koncu tega prispevka.
- 13 Omenimo naj, da se lahko zgodovina podatkovne baze nanaša tako na individualno podatkovno bazo, ki je v obravnavi, kot tudi na podatkovne baze kot takšne na sploh, medtem ko se sociologija podatkovne baze omejuje le na splošne vidike. Oboje pa mora biti striktno empirično, sicer je neuporabno.
- 14 Glede analiz procesov represije teh družbenih pogojenosti gl. Artus (1996) v seznamu Reference 2.
- 15 Razprava o t. i. procesu produkcije podatkov v sedemdesetih in osemdesetih letih – običajno zadeva podatke iz javne uprave ali zgodovinskih virov – je razkrila eno stran uporabnosti določenih podatkov za določene namene pod določenimi pogoji; toda na drugi strani je bila preveč površna, da bi jo upoštevali v tem prispevku.

Iz angleščine prevedel Franci Pivec.

Opombe

- 1 V tem prispevku se dotikamo le bibliografskih podatkovnih baz. Več razlogov bi lahko navedli za raziskovanje projektних podatkovnih baz; v manjši meri pa tudi za nekatere druge tipe podatkovnih baz, zadevajočih raziskovalne institute. Iz določenih razlogov je ta prispevek v celoti osredotočen na podatkovne baze v GESIS, kot so SOLIS (družboslovna literatura v nemško govorečih državah), SOFIS (prej FORIS, družboslovni raziskovalni projekti v nemško govorečih državah) in ZEITSCHRIFT-EN (nemški družboslovni časopisi), ter na primere njihove rabe v kvantitativnih analizah. To lahko pojasni, da je večina primerov precej starih (vendar ne zastarelih), saj so bile raziskave podatkovnih baz v GESIS-IZ opuščene.
- 2 Gl. npr. Binder in Herfurth (1990) in Binder in Stahl (1993) ter Best in R. Ohly (1994) v seznamu Reference 1 kot tudi Sahner (1982) v seznamu Reference 2 tega članka.
- 3 Gl. npr. Güdler (1996); Mutschke (2004).
- 4 Gl. npr. Renner (1997).
- 5 Gl. npr. Artus (2005).
- 6 Gl. npr. Artus (1988, 1992).
- 7 Gl. npr. Artus (1996).
- 8 Gl. npr. Binder in Stahl (1994, 1996).
- 9 Članek "Podatkovna baza" je objavljen na Wikipedii na spletnem naslovu: <http://en.wikipedia.org/wiki/Database>.
- 10 Za razumevanje tega poglavja moramo upoštevati, da se je vse to dogajalo v okviru javnega financiranja. Vsi zajeti instituti so delovali pod enakimi pogoji, z izjemo naših partnerjev v Avstriji in Švici.
- 11 Kasnejši bibliografski podatki prihajajo iz tiskane bibliografije, ki temelji na anketi (!): vsi nemški sociologi so bili pozvani, da poročajo o svojih publikacijah.