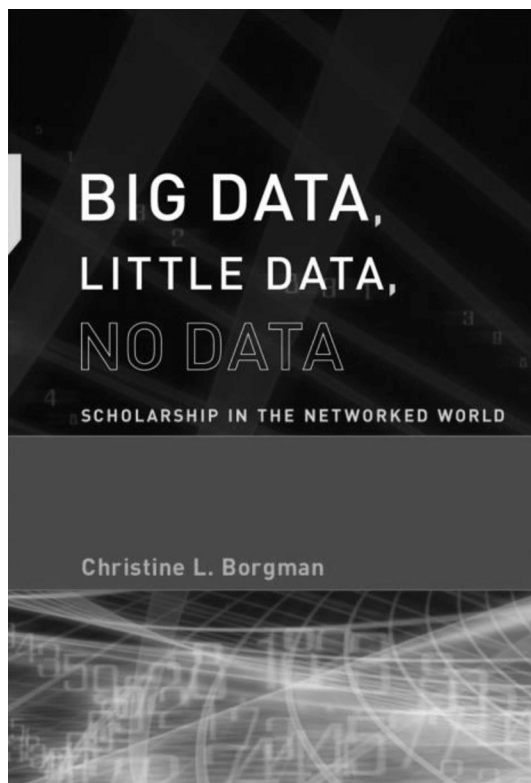


Veľké dáta vo vedeckej komunikácii v digitálnom prostredí: prečo sa treba starať o publikácie a dáta



BORGMAN, Ch. 2015. *Big Data, Little Data, No Data. Scholarship in the Networked World.* Cambridge, Mas.: MIT 2015. 383 s.

Profesorka Christine Borgmanová z UCLA je autorkou, ktorá sa dlhé roky venuje informačnému prostrediu vedy a vedeckej komunikácii v digitálnom prostredí. V najnovšej monografii prináša ucelený pohľad na problematiku veľkých vedeckých dát a súvislostí digitálnych humanitných a digitálnych sociálnych vied. Voľne nadväzuje na predchádzajúcu monografiu o vede v digitálnom veku z roku 2007 (BORGMAN, Ch., 2007. *Scholarship in the Digital Age. Information, Infrastructure and the Internet.* Cambridge: MIT Press, 2007. 336 s.). Sieťové princípy súčasnej vedeckej komunikácie v digitálnom prostredí prinášajú mnoho rôznych a nových otázok. Motiváciou je najmä hľadanie nových rámcov na prácu so záplavou dát vo „veľkých“ vedných odboroch ako astronómia, fyzika, meteorológia, genomika či makroekonomika, ale aj v „menších“ humanitných vedách.

Na úvod Borgmanová kladie niekoľko provokatívnych otázok. Súvisia s reprodukovateľnosťou a novým použitím dát, transferom poznania v rôznych kontextoch a čase, funkciami vedeckého publikovania, s rôznymi účelmi hnutí open access, open data a open source software a s budovaním tzv. znalostných infraštruktúr, ktoré sa neustále rozvíjajú. Znalostné infraštruktúry sú siete ľudí, artefaktov a inštitúcií, ktoré vytvárajú ľudské poznanie. Otázkou je aj definovanie dát vo všeobecnosti – či už ako reprezentácií objektov alebo získaných či nameraných vstupov/parametrov. Dáta sa teda vzťahujú na „entity využívané ako dôkazy javov na účely výskumu alebo vedy“ (s. 29).

Publikácia vynikajúco spracúva súčasný trend rozvoja informačnej vedy, v ktorom je vnorená úzka spolupráca informačných profesionálov s vedeckými a výskumnými pracovníkmi a tímami. Jednotlivé kapitoly po-

skytujú nielen úvod do pochopenia problematiky veľkých dát, ale aj zaujímavé prípadové štúdie. Štruktúra publikácie tiež výborne reprezentuje súčasný trend informačnej vedy, ktorá kladie viac otázok o podstate dát. Prvá časť sa zaoberá dátami a vedou, najmä v súvislosti s definovaním dát, ich rôznorodosťou a tzv. dátovou vedou. Druhá časť predstavuje prípadové štúdie z oblasti dátovej vedy z prírodných vied, sociálnych vied a humanitných vied. Tretia časť sa zaoberá stratégiou a praxou dát, najmä z hľadiska distribúcie, zdieľania a opätovného využitia dát, vlastníctva dát a ich uchovávaní a objavovania nových znalostí. Monografiu dopĺňa výborná bibliografia a register.

Termíny veľké dáta, malé dáta aj žiadne dáta sa týkajú rôznych vedných odborov a rôznej miery práce s nimi. Tieto súvislosti sa vysvetľujú veľkosťou a otvorenosťou dát a najmä princípom „dlhého chvosta“ – grafického znázornenia nerovnomernej distribúcie množstva dát v jednotlivých vedných odboroch. Veľké dáta dominujú v niekoľkých vedných odboroch najmä prírodných, sociálnych a ekonomických vedách ako astronómia, astrofyzika, genomika, ekonómia, sociálne vedy a niektoré digitálne humanitné vedy, menej dát sa spracúva v humanitných vedách. Toto pravidlo sa napríklad týka aj využívania zdrojov v pamäťových inštitúciách (20 % zdrojov má využitie 80 %).

Pokusy o nové definovanie dát, dátovej vedy, znalostnej infraštruktúry súvisia so sociálnou a technologickou zložitosťou vedeckej komunikácie. Autorka štruktúruje výklad prostredníctvom termínov ako komunita a kolaborácia, reprezentácia znalostí, otvorená veda, konvergentná komunikácia (prepojenie formálnej a neformálnej komunikácie). Hľadajú sa jednotky dát v atomizácii komunikácie, záznamy na identifikáciu dát či objektov. Myslím, že jedným z produktívnych riešení je nazeranie na tieto problémy cez optiku komplexnej a dynamicky sa meniacej ekológie znalostnej infraštruktúry. Autorka zdôrazňuje adaptáciu a evolúciu týchto ekosystémov a v tomto zmysle je podľa mňa publikácia potvrdením novej ekologickej paradigmy informačnej vedy s dôrazom na hodnoty dát, ich využívania a ich manažmentu vo vedeckej komunikácii.

Problém veľkých dát a digitálnej vedy je zložitý. Prináša mnoho nových otázok aj praktických úloh, ktoré sa doteraz v niektorých vedách neriešili a vyžadujú si profesionálny prístup informačných špecialistov. Napríklad aj zložitost' citačných prepojení a rôzne motivácie pri citovaní v publikáciách. Dáta sa často stávajú tzv. „čiernou hmotou“ oproti dobre viditeľným

publikáciám so zavedenými postupmi aj metódami citovania a vedeckej komunikácie. Diskusia o vedeckých dátach sa odohráva v niekoľkých rozmeroch vedeckej komunikácie – technologickom, ekonomickom, sociálnom/strategickom, etickom aj právnom. Dáta vytvárajú prostredie vedeckej komunikácie a ovplyvňujú manažment a rozvoj výskumu v súčasnosti aj v budúcnosti. Dáta sa však vyznačujú aj veľkou rôznorodosťou v závislosti od disciplín, metód ich získavania a analýz. Rovnaké entity možno reprezentovať rôznymi spôsobmi, aby mohli byť dôkazom o rôznych javoch. Preto kontext určuje, kedy vznikajú dáta s rôznymi pohľadmi na ich hodnotu, práva či etiku.

Jednotlivé prípadové štúdie vychádzajú z autorkinho vlastného empirického výskumu a spolupráce s rôznymi vedcami. Z oblasti prírodných vied sa uvádzajú prípady astronómie a sensorickej sieťovej vedy a technológie. Zo sociálnych vied sa opiera o výskumy sociálnych médií. Dátovú vedu v oblasti humanitných vied reprezentujú prípady výskumov klasického umenia a archeológie v kontraste s podrobnou analýzou čínskej budhistickej filológie. Z týchto prípadov sa vynára perspektíva informačnej vedy skúmajúcej požiadavky na znalostnú infraštruktúru, napätie medzi aktérmi, rôzne motivácie a postupy na zber, uchovávanie a opätovné využitie dát.

Veľmi oceňujem úvahy o etických aspektoch veľkých dát v jednotlivých vedných odboroch. Týkajú sa napríklad prístupu k prístrojom (astronómia) či citlivosti rôznych dát (navigácia, obrana). V každom odbore dáta súvisia so zdrojmi, existujúcou znalostnou infraštruktúrou, výskumnými otázkami, zaužívanými postupmi pri analýzach aj publikovaní. Uvádzajú sa aj vybrané princípy etického kódexu ACM, napríklad aj v súvislosti s dizajnom a využívaním mobilných technológií. Dynamika sociálnych médií je veľkou výzvou pre manažment výskumných dát, problémom sú nejasné hranice medzi odborními či politická citlivosť tém. V humanitných vedách dominujú viacnásobné interpretácie, rovnaký objekt možno skúmať rôznymi spôsobmi, na rôznych miestach, v rôznych časoch a z rôznych perspektív. Avšak aj nové interpretácie sa opierajú o objekty a súbory uchovávané a akumulované v čase. Inovatívne využívanie technológií, napríklad v budhistickej filológii, hrá dôležitú úlohu nielen pri budovaní znalostnej infraštruktúry, ale aj pri vývoji nových metód výskumu s úplne novými prístupmi k zdrojom.

Najkontroverznejšou časťou publikácie sú kapitoly o zdieľaní, distribúcii a opätovnom využívaní dát a o zodpovednosti a citovaní dát. Analýzy a diskusie sa týkajú princípov ponuky a dopytu pri opätovnom využívaní dát, motivácie na zdieľanie dát v širšom kontexte vedeckej komunikácie. Otázky pre informačnú vedu súvisia s tým, ako v tejto súvislosti budovať znalostnú infraštruktúru (kto, ako, kedy, prečo a s akým účinkom zdieľa dáta, kto dáta opätovne využíva, kedy, prečo a s akým účinkom). Problémom je aj vytváranie stratégií ako regulovať tieto vedecké komunikačné procesy súvisiace s manažmentom dát. K tomu patria aj strategické a ekonomické investičné aspekty digitálnych repozitárov a archívov. Týka sa to stratégií grantových agentúr, univerzít, ale aj časopisov či producentov databáz v súvislosti s podporou výskumu a inovácií. Problémom je aj identifikácia objektov, rôzne formáty digitálnych objektov a citačné správanie s využitím nových technológií. Vynára sa aj novšia teória citačného správania s dôrazom na význam prepojení a citovanie dát. Hlavnými aktérmi manažmentu dát však ostávajú dátové archívy, vydavatelia a knižnice.

V závere autorka predstavuje príklad modelu vedeckého životného cyklu založený najmä na modelovaní sensorických dát z centra CENS (Center for Embedded Networked Sensing). Tri etapy tohto cyklu obsahujú kalibráciu prístrojov a návrh experimentov (plány, laboratórne poznámky, záznamy z experimentov, technické správy ako zdroje dát), zaznamenávanie dát, ich čistenie a analýzy (toky dát, spracovanie súborov dát, modely, analýzy a výsledky), publikovanie a uchovávanie dát (rukopis, preprint, publikácia, dodatočné údaje, publikačné metadáta). V novších modeloch vedeckej komunikácie preto treba brať do úvahy zložitost sociotechnických vzťahov, napríklad autorské práva, zodpovednosť, roly aj riziká pri manažmente dát. Problémom je aj prístup k dátam a rozhodnutia o tom, čo uchovávať a prečo. Autorka uvádza mnoho ďalších zaujímavých príkladov na strategickú, etickú aj technologickú úroveň. Aj keď neexistuje všeobecná zhoda o podstate dát, isté je, že väčšina vedcov by privítala lepšie fungujúcu znalostnú infraštruktúru. To si vyžaduje aj investície do odborníkov, ktorí by prepojili politiku, praktické postupy aj technológie a udržiavali túto infraštruktúru pre vedeckú komunikáciu.

Celkovo je publikácia mimoriadnym prínosom do rozvoja informačnej vedy. Treba povedať, že autorka nie je len optimistická pri úvahách o veľkých dátach, kladie množstvo kontroverzných a provokatívnych otázok. Na druhej strane predstavuje významný trend informačnej vedy dôležitý pre rozvoj teórie aj praxe. Tu sa vynárajú nové problémy a nové profesie pri práci s dátami v nadväznosti na prácu s dokumentmi a digitálnymi objektmi. Dôležité je však, ako sa interpretuje hodnota vedeckej práce a znalostnej infraštruktúry a či sa bude investovať do neviditeľnej, ale dôležitej odbornej práce s rôznymi druhmi dát. Pretože len z existujúcich zdrojov a poznatkov možno vytvárať nové znalosti, objavy, vynálezy či inovácie.

Prof. PhDr. Jela Steinerová, PhD.

jela.steinerova@uniba.sk ■

(Katedra knižničnej a informačnej vedy, Filozofická fakulta, Univerzita Komenského)