

Lielo un atvērto datu tehnoloģijas: viedvalsts attīstības virzītājspēks

JĀNIS GRABIS, Rīgas Tehniskās universitātes Informācijas tehnoloģijas institūta
direktors

Ievads

Viens no 20. gadsimta 90. gadu aktuālākajiem vārdiem informācijas tehnoloģijā bija biznesa procesu reinženierija, kas bija iespējama, pateicoties jaunākajām tehnoloģijām – personālajiem datoriem, internetam un daudzlietotāju sistēmām. Pašreizējais modes kļedziens ir digitālā transformācija, kas nozīmē plašu informācijas tehnoloģiju ienākšanu visās dzīves jomās. Piemēram, Latvijas Universitātē ir rektora vietniece digitālās sabiedrības jautājumos, un Rīgas Tehniskajā universitātē digitalizācija ir kļuvusi par vienu no universitātes attīstības stratēģijas horizontālajām prioritātēm. Digitālās transformācijas pazīstamākās tehnoloģijas ir mākoņdatošana, blokķēdes, lietiskais internets, paplašinātā realitāte un, protams, arī lielo datu tehnoloģijas.

Digitālā transformācija daudziem uzņēmumiem ir ļāvusi gūt milzīgu peļņu, bet sabiedrībai kopumā tā rada arī vairākus jaunus izaicinājumus. Roboti draud aizstāt cilvēkus viņu darba vietās. Pēc vienas prognozes, 73% pavāru vairs nebūs nepieciešami¹, bet citi mierina, ka viņiem būs iespējas veikt augstākas pievienotās vērtības darbus. Atkarība no informācijas tehnoloģijām rada kibernetikas draudus. Viens no blokķēžu izmantošanas iemesliem ir savstarpējās uzticēšanās veicināšana dalītās sistēmās, bet 2017. gada jūlijā “Ethereum” tīklam tika nozagtas monētas 13,5 miljonu ASV dolāru vērtībā.² Mūsu ikdienā izmantotie informācijas tehnoloģijas risinājumi, piemēram, tīmekļa meklēšanas programmas, rekomendāciju sistēmas un sociālo tīklu kvalitāte, ir tieši atkarīgi no mūsu vēlmes dalīties ar saviem personīgajiem datiem. Daloties ar datiem, mēs radām drošības draudus, piemēram, sociālajos tīklos izplatītas ir identitātes zādzības.³ Tomēr arī sabiedrība kopumā var daudz iegūt no jauno tehnoloģiju izmantošanas. Ļoti plaši

¹ Holley P. 2018 The Boston restaurant where robots have replaced the chefs, The Washington Post, <https://www.washingtonpost.com/news/innovations/wp/2018/05/17/will-robots-replace-chefs-at-this-new-boston-restaurant-they-already-have6>

² Browne R. 2018 Cryptocurrency start-up suffers ‘security breach’, theft of \$13.5 million worth of digital tokens, CNBC, <https://www.cnbc.com/2018/07/10/bancor-security-breach-13point5-million-worth-of-cryptocurrency-stolen.html>

³ Evans T. 2016 Sharp rise in identity fraud as scammers use Facebook and other social media sites to hunt for information, The Telegraph, <https://www.telegraph.co.uk/money/consumer-affairs/sharp-rise-in-identity-fraud-as-scammers-use-facebook-and-other/>

izplatīta ir sociālo tīklu saziņas iespēju un attēlu apstrādes izmantošana, lai iegūtu informāciju, sniedzot palīdzību ārkārtas situācijās.

Racionāla un efektīva digitālo tehnoloģiju izmantošana paver iespējas padarīt efektīvākās arī valsts pārvaldes funkcijas un risināt sabiedriski nozīmīgas problēmas. IV Pasaules latviešu zinātnieku kongresā tika izvirzīta tēze par nepieciešamību Latvijai kļūt par viedvalsti. Čārlzs Levi un Deivids Vongs 2014. gadā teica, ka vieda sabiedrība veiksmīgi izmanto digitālās tehnoloģijas un tīklā savienotas viedierīces, lai uzlabotu cilvēku dzīvi.⁴ Citi autori uzskata, ka viedvalsts interpretācijās tiek pārspīlēta tehnoloģiju nozīme. Vispārīgāka viedvalsts definīcija ir “vieda sabiedrība ir tāda, kas rada un izmanto zināšanas, lai būtu sekmīgāka” (Šarls Līdbīters). Šajā rakstā viedvalsts interpretēta šaurākā, datu pieejamības kontekstā. Viedvalstī katram tās pilsonim ir pieejama informācija un spējas (prasmes un tehnoloģijas), kas nepieciešamas, lai varētu pamatoti pieņemt viņam svarīgus lēmumus. Valsts un sabiedrības uzdevums ir padarīt šos informācijas resursus plaši pieejamus, un katra indivīda personīgā izvēle ir to izmantošana vai neizmantošana.

Lielo datu tehnoloģijas

Potenciāli pieejamo un analizējamo datu apjoms un veids strauji pieaug. Šo datu apstrādei izmantotās metodes un tehnoloģijas visas kopumā tiek dēvētas par lielo datu tehnoloģijām. Tās ir paredzētas liela apjoma, strauji mainīgu dažādu veidu datu apstrādei. Tās apvieno statistikas, mākslīgā intelekta, datubāzu, programmēšanas un datorarhitektūras jomas.

Lielos datu mūsdienās rada plašs datu devēju tīkls. Katru dienu pasaule radīto datu apjoms palielinās par 2,5 kvintiljoniem baitu.⁵ Dati tiek radīti uzņēmumos, sensoru tīklos, e-komercijas vietnēs, sociālajos tīklos, kiberfizikālajās sistēmās un daudzās citās sistēmās. Autonomās automašīnas dienā apstrādā aptuveni 4000 GB datu dienā, un šāda datu apjomu apstrādei ir nepieciešams 5G mobilās datu pārraides tīkls. Radītie dati ir dažādu veidu, ieskaitot tradicionālos datus teksta formātā un arvien plašāk izmantotos audio un video datus.

⁴ Haupt M. 2017 What is a Smart Society? Toward the transcendent model society of 2030, Project 2030, <https://medium.com/project-2030/what-is-a-smart-society-92e4a256e852>

⁵ Marr B. 2018 How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read, The Forbes, <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/>

Lielo datu tehnoloģijām piemīt vēl viena īpašība – tās ir visaptverošas. Visi esam datu devēji, patērētāji un potenciālie labuma guvēji. Savās ikdienas aktivitātēs ikviens sabiedrības loceklis ģenerē datus. Vienlaikus mēs arī patērējam datus un informāciju. Lielo datu tehnoloģiju izmantošanas izaicinājums ir nodrošināt, ka datu devēji un patērētāji spēj savstarpēji nodot un saņemt viņiem lietderīgo informāciju. “Google” meklēšanas serviss katru dienu var uzlabot savu algoritmu, jo, meklējot datus, mēs netieši norādām, kurš meklēšanas rezultāts bija īstais. Nododot savus datus uzņēmumam, mēs pašlaik labumu gūstam, izmantojot bezmaksas servisu, bet aprēķini liecina, ka mūsu datu vērtība, ko mēs sniedzam “Facebook”, “Google” un citiem datu apstrādes uzņēmumiem, ir vismaz 200 eiro gadā.⁶

Diemžēl mūsu spējas pašlaik izmantot lielo datu tehnoloģijas un gūt labumu tieša veidā ir ļoti ierobežotas, salīdzinot ar “Facebook”, “Apple”, “Netflix” un “Google” jeb FANG.

Atklātie dati

Tomēr, lai sacenstos ar lielo datu milžiem, ir viena alternatīva – atklātie dati. Atklātie dati ir datus kopas, kuras dažādas organizācijas un iestādes nodod sabiedrības rīcībā, lai ikvienam būtu līdzvērtīgas iespējas izmantot datus un, viedvalsts kontekstā, pieņemt ar datiem pamatotus lēmumus. Viens no pazīstamākajiem atklāto datu projektiem ir Čikāgas municipalitātes piemērs, kura publicē internetā visus tās rīcībā esošos datus.⁷ Datu analīzes ceļā, piemēram, tika noskaidrots, ka ir izteikta korelācija starp ielu apgaismojumu un noziedzību. Ja kādā ielā ir ziņots par izdegušu lampiņu, tad būtiski pieaug varbūtība, ka šajā ielā notiks kāds noziegums.

Arī Latvijā atklāto datu ideja nav sveša, un daudzos gadījumos likumi nosaka, ka valsts iestāžu rīcībā esošajiem datiem ir jābūt plaši pieejamiem. Tomēr praksē atklāto datu izmantošanas iespējas ir ierobežotas, jo dati ir izkļaidēti dažādās vietās, tiem ir dažādi formāti, nepiemēroti piekļuves veidi un to interpretācijas un izmantošanas iespējas ir neskaidras. Piemēram, Latvijas Republikas Vides aizsardzības un reģionālās attīstības ministrija uztur atvērto datu portālu <https://data.gov.lv>, bet tajā publicēto datu kopu skaits un izmantotāju skaits ir salīdzinoši mazs. Ministrija lielo

⁶ Wibson 2018 How Much Is Your Data Worth? At Least \$240 per Year. Likely Much More. <https://medium.com/wibson/how-much-is-your-data-worth-at-least-240-per-year-likely-much-more-984e250c2ffa>

⁷ <https://data.cityofchicago.org/>

datu izmantošanas iespēju popularizēšanu ir izvirzījusi par vienu no savām prioritātēm. Datu portālos tiek nodrošināta datu pieejamība, tomēr tas nenodrošina lietotājiem un uzņēmumiem nepieciešamās spējas, lai lielos datus varētu izmantot izmantošanai.

Vēl viena kritiska problēma ir organizatoriska, jo dažādas organizācijas nespēj vienoties par dažādiem datu pārvaldības jautājumiem, un katra organizācija datus apstrādā atbilstoši savai varēšanai un saprašanai. Valstī ir bijuši vairāki mēģinājumi centralizēt vai vismaz koordinēt datu pārvaldību un apstrādi, bet panākumi ir ierobežoti. Problēma nav unikāla tikai Latvijai. Arī Eiropā, neraugoties uz Eiropas Savienības ieguldītajiem līdzekļiem gan pētniecībā, gan infrastruktūras izveidē, datu izmantošanas iespējas ir ierobežotas. Eiropa būtiski atpaliek no ASV, kur datu integrācijas ziņā iniciatīva pieder lielajiem uzņēmumiem, un Ķīnas, kur datu pārvaldības jautājumi ir stingrā valsts kontrolē aiz Ķīnas Lielā ugunsmūra.

Alternatīva centralizācijai un koordinācijai ir decentralizācija un pašorganizācija. Rīgas Tehniskās universitātes Informācijas tehnoloģijas institūta pētnieki ir izstrādājuši datu integrācijas un sistēmu adaptācijas platformu, kas ļauj visām iesaistītajām pusēm būt gan datu devējiem, gan datu patērētājiem. Viena no platformas iezīmēm ir, ka tā ļauj atklāt, kā datus izmanto dažādie patērētāji. Tas ļauj risināt klasisko datu pārvaldības problēmu, ka datu ir daudz, bet izpratnes par to lietošanu, t. i. zināšanu, pietrūkst. Platformas vīzija ir padarīt zināšanas, kā datus izmantot, par plaša patēriņa produktu visām iesaistītajām pusēm.

Viens no potenciālajiem platformas izmantošanas veidiem ir datu pārvaldība efektīvākā autoceļu uzturēšanā ziemā.

Ziemas ceļu uzturēšanas problēma

Autoceļu uzturēšana ir resursietilpīga. Regulārās uzturēšanas aktivitātes ir iespējas efektīvi plānot, bet ziemas ceļu uzturēšana, kas ietver ceļu attīrīšanu no sniega un ledus un apstrādi ar pretslīdes līdzekļiem, ir apgrūtināta, jo laikapstākļi ir neprognozējami. Neraugoties uz to, ziemas ceļu uzturēšanā kritiski svarīgi ir laikus reaģēt uz ceļu stāvokļa izmaiņām. Diemžēl informācija par ceļu stāvokli ir fragmentāra un nepilnīga. Ceļu novērošanas kameras un laikapstākļu mērīšanas meteostacijas ir izvietotas nevienmērīgi, un bieži ziņas par jau esošu apledojumu vispirms parādās sociālās tīklošanas mikroziņās.

No datu apstrādes viedokļa ziemas ceļu uzturēšanā ir iesaistīti vairāki datu devēji un datu patērētāji. Datu devēji ir “Latvijas Valsts ceļu” pārvaldītās ceļu novērošanas kameras, pašvaldību, ceļu uzturētāju un meteoroloģisko dienestu meteostacijas, mobilo datu pārraides tīklu operatori, ģeogrāfiskās informācijas sistēmas un viedtālruņi. Datu patērētāji ir ceļu uzturētāji, ceļu lietotāji, operatīvie dienesti un citi. Katra iesaistītā puse var būt gan datu devējs, gan patērētājs. Piemēram, ceļu uzturētāji saņem informāciju par nepieciešamajiem ceļu uzturēšanas pasākumiem un sniedz datus par ceļu pašreizējo stāvokli. Arī viedtālruņu īpašnieki gan sniedz datus par braukšanas apstākļiem, gan saņem brīdinājumus par bīstamajiem ceļu posmiem. Starp šiem datiem vairākas datu kopas, piemēram, meteostaciju mērījumi, ir atvērtie dati, kurus varētu izmantot visi interesenti.

Lai nodrošinātu savlaicīgu ceļu tīrīšanu un pretslīdes apkopi, ceļu uzturētājiem ir nepieciešamas prognozes par sagaidāmo ceļu stāvokli. Ceļu uzturēšanas uzņēmumi valstī un pašvaldībās savā starpā varētu apmainīties ar zināšanām par datiem, kuru izmantošana ļauj uzlabot prognozēšanu.

Lai iedzīvīnātu šādu decentralizētu datu devēju un patērētāju tīklu, ir nepieciešamas atbilstošas tehnoloģijas. Datu integrācijas platformā datu patērētāji specificē viņiem nepieciešamo informāciju, un datu devēji reģistrē viņu rīcībā esošos datus. Tā nodrošina liela apjoma datu straumju apstrādi tuvu reāllaikam un transformē datu straumes patērētājiem nepieciešamā informācijā. Katrs datu apmaiņas sistēmas dalībnieks var izmantot savu platformas instanci, kļūstot par decentralizēta datu apmaiņas tīkla dalībnieku. Datu apstrādes rezultāti tiek izmantoti, pieņemot lēmumu uzņēmumā vai organizācijā. Ceļu uzturēšanas gadījumā autoceļu uzturētājs pieprasa informāciju par iespējamo apledojumu. Meteodienesti, ceļu novērošanas kameras un sociālie tīkli piegādā aktuālos datus par iespējamo ceļu stāvokli, un datu apstrādes platforma, integrējot un apstrādājot saņemtās datu plūsmas, prognozē ceļu stāvokli, piegādājot informāciju ceļu uzturēšanas uzņēmumam. Ceļu uzturēšanas uzņēmums novēro, kuru datu plūsmu izmantošana ir devusi labāko rezultātu ceļu stāvokļa prognozēšanā, un ģenerē zināšanas par datu izmantošanas iespējām. Ja vairāki ceļu uzturēšanas uzņēmumi publicē šīs zināšanas vienotā repozitorijā, tad iesaistītās puses spēj datu integrācijas labo praksi izmantot savā darbībā. Datu integrācijas platforma ir modeļvirzīta, konfigurējama un augsti mērogojama, kas ļauj katram datu apmaiņas tīkla dalībniekam viegli piekļūt datu apstrādes tehnoloģijām.

Datu un zināšanu izmantošana sniedz lielus potenciālus ieguvumus, jo ceļu uzturētājiem viens nelietderīgs ceļu tīrīšanas izbraukums izmaksā aptuveni 25 000 eiro. Viens ceļu negadījumā bojā gājis cilvēks – tie ir 0,5 miljoni eiro zaudējumi valstij.

Nobeigums

Viedvalstī katram tās pilsonim ir pieejama informācija un spējas, kas nepieciešamas, lai varētu pamatoti pieņemt viņam svarīgus lēmumus.

Atklātie dati ir iespēja visiem: lietot datu tehnoloģijas ir viens no aktuālākajiem informācijas tehnoloģijas pašreizējās attīstības virzieniem. Tās var dot būtisku ieguldījumu viedvalsts attīstībā. Uzņēmumi jau ir pārvarējuši pirmo entuziasma vilni un konstatējuši, ka datu daudzums pats par sevi vēl nenodrošina apsolītos ieguvumus. Sabiedrība kopumā vēl ir tikai ceļā uz plašāku lieto datu tehnoloģiju izmantošanu, un atklātie dati, tos pareizi pārvaldot un apstrādājot, paver plašas iespējas.

Decentralizācijas posts un spožums: centralizēta datu apstrāde bieži vien ir efektīvāka par decentralizētu datu apstrādi, bet centralizācijas fundamentāls traucēklis ir grūtības vienoties par kopīgu darbību. Decentralizēta pieeja ir piemērotākā Latvijas gadījumā, bet tas nozīmē, ka visu dalībnieku rīcībā ir jābūt pietiekami vienkāršam mehānismam, kas viegli ļauj pievienoties datu apstrādes un zināšanu pārvaldības tīklam.

Dati pieder mums katram un ir nepieciešami mums visiem, atliek tikai tos sekmīgi izmantot.