

# Koncept i primjena jezera podataka u zdravstvu

---

Šimić, Ivan

Master's thesis / Diplomski rad

2023

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Economics and Business / Sveučilište u Zagrebu, Ekonomski fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:148:191896>

Rights / Prava: [Attribution-NonCommercial-ShareAlike 3.0 Unported/Imenovanje-Nekomercijalno-Dijeli pod istim uvjetima 3.0](#)

Download date / Datum preuzimanja: **2024-07-14**



Repository / Repozitorij:

[REPEFZG - Digital Repository - Faculty of Economics & Business Zagreb](#)



**Sveučilište u Zagrebu**

**Ekonomski fakultet**

**Integrirani preddiplomski i diplomski sveučilišni studij**

**Poslovna ekonomija - smjer Menadžerska informatika**

**KONCEPT I PRIMJENA JEZERA PODATAKA U  
ZDRAVSTVU**

**CONCEPT AND APPLICATION OF DATA LAKE IN  
HEALTHCARE**

Diplomski rad

**Student: Ivan Šimić**

**JMBAG: 0067540458**

**Mentor: Prof. dr. sc. Katarina Ćurko**

**Zagreb, lipanj 2023.**

IVAN ŠIMIĆ

Ime i prezime studenta/ice

## IZJAVA O AKADEMSKOJ ČESTITOSTI

Izjavljujem i svojim potpisom potvrđujem da je DIPLOMSKI RAD  
(vrsta rada)

isključivo rezultat mog vlastitog rada koji se temelji na mojim istraživanjima i oslanja se na objavljenu literaturu, a što pokazuju korištene bilješke i bibliografija. Izjavljujem da nijedan dio rada nije napisan na nedozvoljen način, odnosno da je prepisan iz necitiranog rada, te da nijedan dio rada ne krši bilo čija autorska prava. Izjavljujem, također, da nijedan dio rada nije iskorišten za bilo koji drugi rad u bilo kojoj drugoj visokoškolskoj, znanstvenoj ili obrazovnoj ustanovi.

U Zagrebu, 18.06.2023.

Student:

*Ivan Šimić*

(potpis)

## SAŽETAK

Količina podataka, kao i njihova dostupnost, na najvišim je razinama otkad je svijeta. S pojavom Velikih podataka pojavio se problem u skladištenju i iskorištavanju podataka na optimalan način. Tradicionalna skladišta podataka sve češće ne zadovoljavaju velike potrebe u procesiranju i skladištenju podataka koje se javljaju u sve više industrija, dok koncept jezera podataka postaje sve popularniji i prisutniji u brojnim industrijama. Jedna od tradicionalnih djelatnosti u kojoj dionici sve više shvaćaju potrebu za digitalizacijom svog poslovanja, iskorištavanjem velike količine podataka a time i korištenja jezera podataka je zdravstvo. U ovom radu će se pokušati prikazati kako jezera podataka i digitalne tehnologije odgovaraju na probleme poput skladištenja i upravljanja podacima te probleme u interoperabilnosti, sigurnosti i analizi podataka u zdravstvu.

Ključne riječi: zdravstvo, digitalna transformacija, jezero podataka, veliki podaci, strojno učenje

## **SUMMARY**

The amount of data, as well as its availability, is at the highest levels ever. With the appearance of Big Data, a problem in storing and using data in an optimal way arose. Traditional data warehouses increasingly do not meet the great needs in data processing and storage that arise in more and more industries, while the concept of data lakes is becoming more popular and present in numerous industries. One of the traditional industries in which stakeholders are increasingly realizing the need to digitize their business, exploit a large amount of data and thus use data lakes, is healthcare. This paper will try to show how data lakes and digital technologies respond to problems such as data management and governance and problems in interoperability, security and data analysis in healthcare.

Key words: healthcare, digital transformation, data lake, Big data, machine learning

## SADRŽAJ

1. UVOD .....	1
1.1. Cilj rada .....	1
1.2. Izvori podataka i metode prikupljanja.....	1
1.3. Sadržaj i struktura rada.....	1
2. POSLOVNO OKRUŽENJE U ZDRAVSTVU.....	3
2.1. Pojmovno određenje i značaj zdravstva .....	3
2.2. Osnovne karakteristike zdravstva u Republici Hrvatskoj i svijetu .....	4
2.3. Digitalna transformacija i poslovni zahtjevi.....	9
2.4. Utjecaj pandemije Covid – 19 na zdravstvo.....	13
3. JEZERO PODATAKA .....	15
3.1. Povijest i koncept jezera podataka.....	15
3.2. Vrste podataka .....	18
3.3. Skladište podataka i jezero podataka .....	19
3.4. Arhitektura jezera podataka.....	21
3.5. Vladanje podacima (engl. Data governance).....	23
3.6. Sustavi za pohranu podataka u jezeru podataka .....	25
3.7. Vodeća tehnološka rješenja .....	26
3.8. Veliki podaci i umjetna inteligencija.....	27
4. PRIMJENA JEZERA PODATAKA U ZDRAVSTVU.....	31
4.1. Veliki podaci u zdravstvu.....	31
4.2. Jezera podataka u zdravstvu .....	33
4.3. Slučajevi korištenja jezera podataka u zdravstvu .....	35
4.4. Umjetna inteligencija i strojno učenje u zdravstvu.....	37
4.5. Zaštita podataka .....	39
5. STUDIJA SLUČAJA AMAZON HEALTHLAKE PLATFORME .....	42
5.1. Opis Amazon HealthLake platforme .....	42
5.2. Prikaz uloga alata i popratnih tehnologija u Amazon Healthlake platformi .....	44
5.3. Primjeri primjene i doprinosa Amazon HealthLake-a u zdravstvu .....	51
5.4. Budućnost i izazovi jezera podataka.....	53
6. ZAKLJUČAK.....	55
7. LITERATURA .....	56
POPIS SLIKA.....	62
POPIS GRAFIKONA .....	63
POPIS TABLICA .....	64



# **1. UVOD**

## **1.1. Cilj rada**

Istraživanjem se želi prikazati poslovno okruženje u zdravstvenoj djelatnosti koje je u posljednje vrijeme sve dinamičnije i zahtijeva velike prilagodbe. Cilj je objasniti ulogu informacijsko – komunikacijskih tehnologija i načine putem kojih utječu na promjenu u poslovanju zdravstva. Pandemija Covid -19 je poslužila kao katalizator promjena u dijeljenju podataka među zdravstvenim organizacijama što je olakšalo poslovanje mnogim dionicima u sustavu zdravstva. Tehnologije poput umjetne inteligencije i strojnog učenja već uvelike mijenjaju zdravstvo kakvo smo poznavali i daju mogućnost donošenja boljih odluka na osnovi analiza podataka. Tradicionalna skladišta podataka više nisu u mogućnosti pohranjivati ogromne količine podataka i to dovodi do novih i prikladnijih rješenja za pohranu podataka, posebice u zdravstvenoj djelatnosti. Kroz slučajeve korištenja jezera podataka i studiju slučaja Amazon HealthLake će se prikazati na koji način jezero podataka i digitalne tehnologije u sklopu Amazon Web servisa mogu utjecati na zdravstvo.

## **1.2. Izvori podataka i metode prikupljanja**

Metode koje se koriste u radu su induktivna i deduktivna metoda, metode analize i sinteze, metoda generalizacije i specijalizacije te sustavno praćenje. Korištene su brojne knjige domaćih i stranih autora, znanstveni i stručni članci te internet izvori.

U prvom dijelu ovog rada, prikazat će se teoretski dio koji se odnosi na zdravstvo i aktualne probleme s kojima se ono suočava, kao i koncept jezera podataka i fenomen Velikih podataka. U drugom dijelu rada će se prikazati korist jezera podataka za zdravstvo kroz slučajeve korištenja a na samom kraju će se prikazati i studija slučaja Healthlake platforme tj. alata i tehnologija koje mogu napraviti velike pomake u zdravstvu.

## **1.3. Sadržaj i struktura rada**

Diplomski rad “Primjena i koncept jezera podataka u zdravstvu“ se sastoji od 6 tematskih cjelina. U prvom poglavlju predloženi su ciljevi rada, izvori podataka i metode prikupljanja



podataka te sadržaj i struktura rada. U drugom poglavlju je dan pregled poslovnog okruženja u zdravstvu s naglaskom na digitalnu transformaciju i utjecaj pandemije Covid-19. Treće poglavlje se bavi konceptom jezera podataka kao i njegovim tehnološkim sastavnicama. Četvrto poglavlje govori o primjeni jezera podataka i drugih tehnologija proizašlih iz Velikih podataka u zdravstvu. U petom poglavlju je napravljena studija slučaja Amazon HealthLake platforme a na samom kraju rad se zaključuje sintezom svih poglavlja i popisom korištene literature, popisom slika, tablica i grafikona te životopisom autora.

## 2. POSLOVNO OKRUŽENJE U ZDRAVSTVU

### 2.1. Pojmovno određenje i značaj zdravstva

Zdravstvo je djelatnost kojoj je cilj unaprjeđenje zdravlja, sprječavanje i pravodobno otkrivanje bolesti te liječenje i rehabilitacija bolesnika, s pomoću specifičnih sredstava, opreme i metoda rada; čine ga zdravstvene ustanove i djelatnici koji pružaju zdravstvenu zaštitu građanima u određenom sustavu organizacije zdravstvene službe, a regulira se posebnim zakonskim propisima. Na stupanj razvijenosti i organizaciju zdravstva utječu socijalne politike, gospodarske prilike, demografska kretanja i pokazatelji stanja stanovništva, državna organizacija te znanstvene spoznaje i dostignuća medicinske prakse.<sup>1</sup>

Prema Zakonu o zdravstvenoj zaštiti, svaka osoba ima pravo na zdravstvenu zaštitu i mogućnost ostvarenja najviše moguće razine zdravlja. Važnost zdravlja je neprocjenjiva i neupitna a država ima zadatak i obvezu stvoriti uvjete za provođenje zdravstvene zaštite kao i poboljšanje zdravlja stanovništva kroz gospodarske i socijalne politike.<sup>2</sup>

U posljednjih par desetljeća zdravstvo se razvijalo intenzivno i kontinuirano a učinak tog razvoja je vidljiv kroz produljenje životnog vijeka i kroz poboljšanje kvalitete života. Zdravstvena djelatnost ima veliki utjecaj na tokove u gospodarstvu. Što je zdravstvo razvijenije, svijest o važnosti zdravlja među populacijom će biti veća a liječenje uspješnije pa će i efikasnost radne snage biti veća. Razina kvalitete zdravstvene skrbi je jedan od bitnijih činitelja kako građani doživljavaju kvalitetu svog života a u mnogim zemljama pored ekonomskih pitanja, pitanje zdravstva je vodeće pitanje u nacionalnim politikama.

Bez obzira na veliki značaj zdravstva, mnoge zemlje u svijetu ne pružaju osnovnu besplatnu zdravstvenu zaštitu. Posljednjih godina se sve češće može čuti kako mnogobrojni čelnici svjetskih ekonomskih i zdravstvenih organizacija potiču države na veća ulaganja i osiguravanje besplatne zdravstvene zaštite kako bi se osigurala temeljna ljudska prava ali isto tako i osnažile ekonomije povećanjem stope produktivnosti rada.

---

<sup>1</sup> Zdravstvo, Hrvatska enciklopedija, mrežno izdanje. Leksikografski zavod Miroslav Krleža, 2021., Dostupno na: <http://www.enciklopedija.hr/Natuknica.aspx?ID=67025/>, Pristupano 21.08.2022.

<sup>2</sup> Zakon o zdravstvenoj zaštiti, Narodne novine NN 100/18, 125/19, 147/20, Dostupno na: <https://www.zakon.hr/z/190/Zakon-o-zdravstvenoj-za%C5%A1titi>, Pristupano 22.08.2022.

## 2.2. Osnovne karakteristike zdravstva u Republici Hrvatskoj i svijetu

Zdravstvo je jedna od važnijih djelatnosti u nacionalnim gospodarstvima i njen značaj u društvu je izniman. Shodno tome, zdravstvo predstavlja jednu od bitnijih komponenti u javnim politikama gospodarstava diljem svijeta. Pitanje zdravlja je osjetljivo pitanje koje zadire duboko u privatnost pojedinaca i upravo je to razlog zašto se veliki fokus stavlja na kvalitetnu zdravstvenu zaštitu. Ipak, zdravstvena djelatnost ne može doseći svoj domet ukoliko ne postoji izgrađen zdravstveni sustav koji je osnova i potpora za sve daljnje procese. Svaka zemlja ima zdravstveni sustav koji pokušava što uspješnije prilagoditi potrebama svojeg stanovništva s ciljem pružanja profesionalne usluge u što kraćem vremenskom roku.

U Republici Hrvatskoj standard zdravstvene zaštite je većinom zadovoljavajuć, s naglaskom na veću kvalitetu i dostupnost zdravstvenih usluga u većim gradovima. Raspon zdravstvenih usluga u područjima koja su udaljena od urbanih sredina i na otocima je manji ali još uvijek prihvatljiv.<sup>3</sup>

Ministarstvo zdravstva Republike Hrvatske ima zadatak upravljanja sustavom zdravstvene zaštite a to uključuje upravljanje zakonodavstvom iz područja zdravstvene zaštite, izradu proračuna, nadziranje zdravstvenog statusa i zdravstvenih potreba, edukaciju zdravstvenih djelatnika i nadgledanje procesa reforme zdravstvenog sustava.<sup>4</sup>

Sustavi pružanja i financiranja usluga i zdravstva su različiti od države do države. U Republici Hrvatskoj pružanje usluga u sektoru zdravstva je većinom javno iako postoje i privatni pružatelji usluga.

Zdravstvena zaštita u Republici Hrvatskoj je organizirana na tri razine a to su primarna, sekundarna i tercijarna te na razini zdravstvenih zavoda. Primarna zdravstvena zaštita podrazumijeva ustanove poput domova zdravlja, prijema za hitnu pomoć, ustanove za zdravstvenu njegu i palijativnu skrb te ljekarničke ustanove. Specijalističko – konzilijarna i bolnička djelatnost provodi se na sekundarnoj razini a tercijarna razina je namijenjena

---

<sup>3</sup> Opis zdravstvenog sustava, Dostupno na: <https://hzzo.hr/pravni-akti/opis-zdravstvenog-sustava>, Pristupano 16.08.2022.

<sup>4</sup> Džakula, A., Sagan, A., Pavić, N., Lončarek, K. (2014) Health Systems in Transition - Croatia Health System Review 2014. The European Observatory on Health Systems and Policies, Copenhagen. Međunarodna studija, Dostupno na: <https://apps.who.int/iris/handle/10665/128604>, Pristupano 16.08.2022.

znanstveno – istraživačkim radovima i nastavi. Ustanove s tercijarne razine su u vlasništvu države, dok su ustanove s prve dvije razine u vlasništvu županija.

Zdravstvo se financira iz različitih izvora a najveći dio se zasniva na uplatama osiguranika što je karakteristika Bismarckovog modela zdravstvene zaštite koji je nastao u Prusiji 1883. godine a danas je prisutan u većem dijelu središnje i zapadne Europe.<sup>5</sup>

Doprinosi za zdravstveno osiguranje se naplaćuju iz rada, tj. iz plaća a upravljanje sredstvima i sustavom je pod kontrolom države. Stopa doprinosa za zdravstveno osiguranje iznosi 16,5 % i utvrđuje je hrvatski Sabor.

Veliki problem u hrvatskom zdravstvu predstavljaju duge liste čekanja za sekundarnu i tercijarnu skrb koje su se u razdoblju pandemije Covid – 19 dodatno povećale. Sektor zdravstva se već godinama sektor bori s visokim dugom a preko 70 posto rashoda za zdravstvo u Republici Hrvatskoj financira javni sektor što je veliki problem za državu, uzmemo li u obzir činjenicu sve starijeg stanovništva i dug u zdravstvu koji je preko 6 milijardi kuna.

Broj zaposlenih u zdravstvu Republike Hrvatske 2020. godine je iznosio 72.929 od čega je 48.489 (66,4%) bilo zdravstvenih djelatnika. Najbrojniju grupu čine medicinske sestre i medicinski tehničari kojih je 33701 (46.2 %) od ukupnog broja zaposlenih a slijede ih doktori medicine kojih je bilo 15.417 (21.1%). Ukupni broj djelatnika u zdravstvu je u porastu ali i dalje nije dovoljan broj medicinskih sestara i tehničara kao ni doktora opće prakse u određenim područjima van urbanih sredina.

U javnom zdravstvu Republike Hrvatske veći broj djelatnika su žene a od ukupnog broja liječnika čak 64 posto su žene dok je u sestrinstvu 87 posto medicinskih sestara. Međutim, na vodećim položajima u zdravstvu, poput ravnatelja bolnica, u velikoj većini su muškarci.<sup>6</sup>

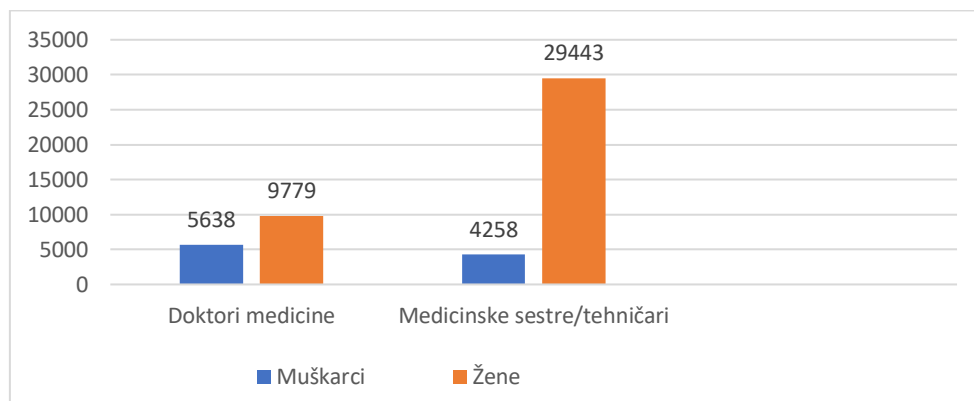
U Republici Hrvatskoj 2019. godine je od ukupnog broja liječnika, njih 23,7% su bili liječnici opće medicine. Liječničke usluge su ugovorene od strane Hrvatskog zavoda za zdravstveno osiguranje, dok su plaće određene po učinku i pokazateljima kvalitete kao i sustavu koeficijenata plaća određenih liječnika. Većina liječnika koji rade u primarnoj zdravstvenoj zaštiti su samozaposleni

---

<sup>5</sup> Kovač, N. (2013.), Financiranje zdravstva – situacija u Hrvatskoj, str. 553, Dostupno na: <https://hrcak.srce.hr/file/172162>, Pristupano 21.08.2022.

<sup>6</sup> Hrvatski zdravstveno-statistički ljetopis za 2020., Dostupno na: <https://www.hzjz.hr/hrvatski-zdravstveno-statisticki-ljetopis/hrvatski-zdravstveno-statisticki-ljetopis-za-2020-tablicni-podaci/>, Pristupano 19.08.2022.

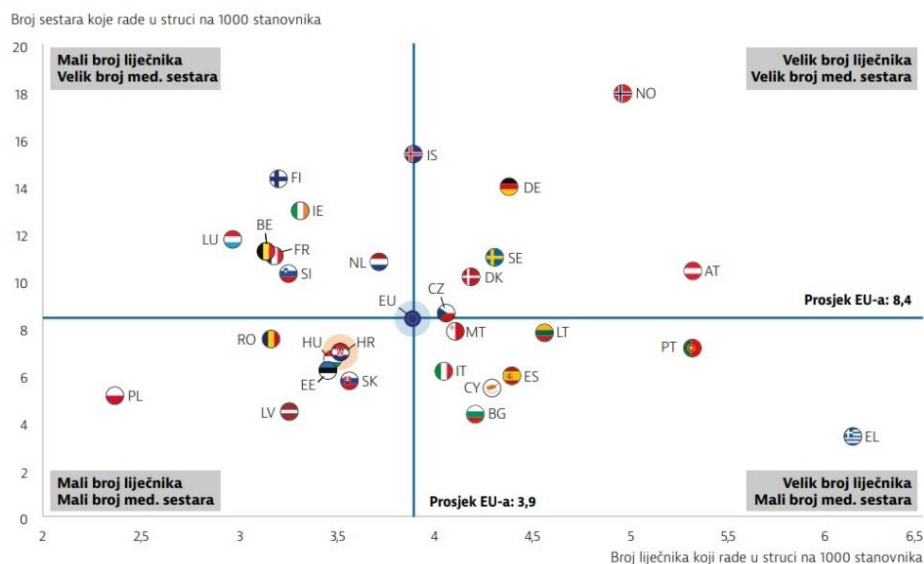
Grafikon 1 Struktura zaposlenih doktora medicine i medicinskih sestara/tehničara po spolu 2020.god. (u tisućama)



Izvor: Obrada autora prema podacima Hrvatskog zavoda za javno zdravstvo (HZJZ)

Unatoč stabilnom i konstantnom rastu broja liječnika i medicinskih sestara od 2013.godine Hrvatskoj i dalje nedostaje zdravstvenog kadra. Podaci iz 2019. godine pokazuju kako je Hrvatska na 1000 stanovnika imala 6,8 medicinskih sestara i 3,5 liječnika u odnosu na prosjek Europske unije od 8,4 medicinske sestre i 3,9 liječnika.<sup>7</sup>

Slika 1 Manjak zdravstvenih radnika u odnosu na prosjek EU



Izvor: Europska komisija

<sup>7</sup> State of Health in the EU - Hrvatska, Dostupno na: [https://health.ec.europa.eu/system/files/2022-01/2021\\_chp\\_hr\\_croatian.pdf](https://health.ec.europa.eu/system/files/2022-01/2021_chp_hr_croatian.pdf), Pristupano 19.08.2022.

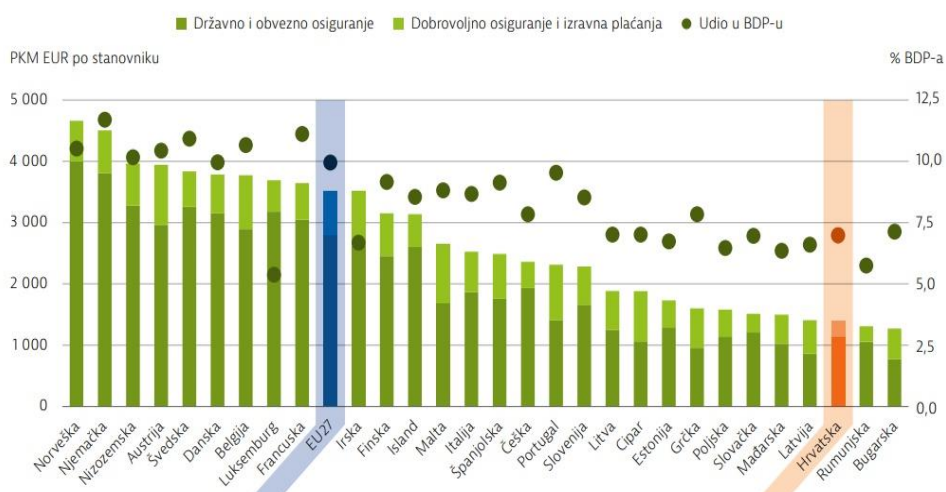
Jedan od pokazatelja stanja zdravstva je i kako ono utječe na očekivani životni vijek. Kvalitetne zdravstvene usluge nisu jedini faktor u produžavanju očekivanog životnog vijeka ali imaju veliku ulogu u tom procesu. U Hrvatskoj je 2010.godine očekivani životni vijek pri rođenju bio 76,7 godina a 2019.godine 78,6 godina. Usporedbe radi, na razini Europske unije prosjek očekivanog životnog vijeka je 2,7 godina veći.

Prije pandemije, Hrvatska je imala nisku stopu nezadovoljenih potreba za zdravstvenom skrbi (1,4%) a populacija koja je imala poteškoće s neizvršenim liječničkim pregledima, vremenom čekanja ili udaljenostima do zdravstvenih ustanova bila je ona s malim dohotkom (bez obzira na to što je oslobođena plaćanja participacije uslijed socijalnih uvjeta). U usporedbi s prosjekom nezadovoljenih zdravstvenih potreba u Europskoj uniji, Hrvatska je imala nižu stopu ali s naglašenim razlikama između dohodovnih grupa.

Iako je posljednjih godina u razvijenim zemljama svijeta znatno pojačana potrošnja u zdravstvenom sektoru, Hrvatska je u okviru Europske unije jedna od članica koja po glavi stanovnika troši daleko manje od većine članica. Točnije, Hrvatska je 2019. godine bila među tri države članice EU koja su najmanje trošile na zdravstvo tj. 7 posto BDP-a odnosno 1392 eura po glavi stanovnika, a prosjek potrošnje na zdravstvo u EU je 9.9 posto BDP-a.

S druge strane, jedna je od država kojoj je javni udio u ukupnim troškovima 2019. godine bio veći od 81.9 % što je stavlja u skupinu zemalja čiji je javni udio u troškovima izrazito visok.<sup>8</sup>

Slika 2 Potrošnja na zdravstvo u članicama EU



Izvor: OECD, 2019.

<sup>8</sup> Ibid., str.9

Nekoliko je reformi zdravstvenog sustava provedeno u zadnja dva desetljeća i neke od njih su bile uspješne a neke se još uvijek provode. Iako je u zdravstvu vidljiv napredak, ostaju izazovi kako odgovoriti na problem sve većeg iseljavanja zdravstvenih radnika u zemlje zapadne Europe ali i kompleksan problem financiranja zdravstva koji godinama unazad nije uspio riješiti nitko od vladajućih.

Hrvatska dijeli slične probleme u zdravstvu s ostatkom svijeta koji se također suočava s izazovima starenja stanovništva, pružanja visoko kvalitetne usluge i manjkom radne snage. Svi ovi izazovi su utjecali na drastično povećanu potrošnju na zdravstvo u svijetu u zadnjih nekoliko godina.

U posljednjih dva desetljeća, potrošnja se više nego udvostručila da bi 2019.godine dosegla skoro 9% svjetskog BDP-a. Dio povećane potrošnje na zdravstvene usluge odnosi se na kućanstva ali otprilike 60 posto izdataka za zdravstvo se odnosi na potrošnju iz državnih proračuna.<sup>9</sup> Oko 80% svjetske potrošnje na zdravstvo se odnosi na zemlje s visokim dohotkom.

U zemljama s visokim dohotkom troškovi koji su generirani od strane države dosežu i 70% dok zemlje s nižim dohotkom karakterizira imaju veći postotak potrošnje na zdravstvo od strane privatnih osoba i kućanstava. Većina zemalja koje su imale nisku zdravstvenu potrošnju bile su iz južne Azije i subsaharske regije dok je većina zemalja koja je imala visoku potrošnju bila u Europi, Sjevernoj Americi i istočnoj Aziji.<sup>10</sup> Europski i globalni zdravstveni podaci su iznimno fragmentirani, podaci se prikupljaju i kontroliraju mnoštvom lokalnih, regionalnih istraživanja, što otežava potpuni pregled pitanja vezana uz zdravstvenu skrb.<sup>11</sup>

Sjedinjene Američke Države i dalje imaju najveći trošak po glavi stanovnika kada je u pitanju zdravstvo (10948 dolara) a iza njih dolaze zemlje zapadne i sjeverne Europe koje troše znatno manje od SAD-a.<sup>12</sup>

---

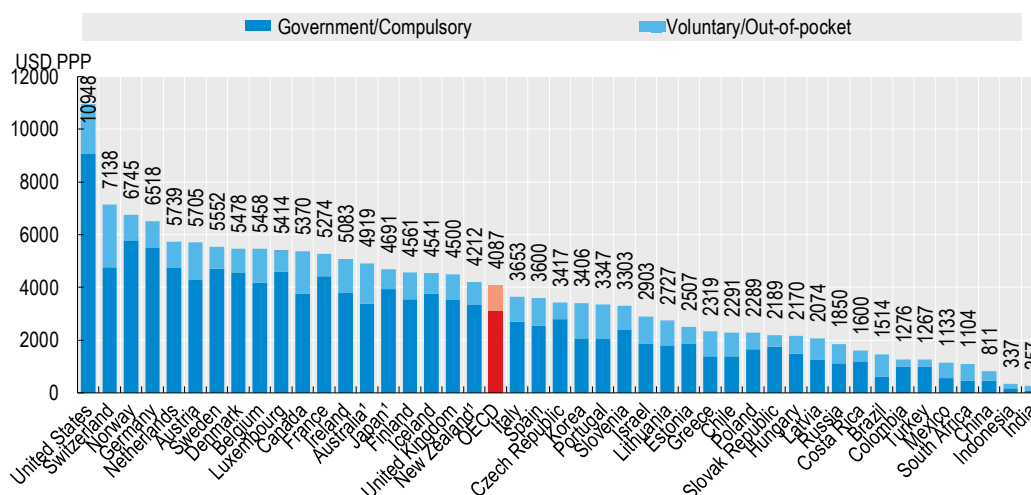
<sup>9</sup> World Health Organization (2019), Global spending on health: A world in transition, Dostupno na: <https://www.who.int/publications/i/item/WHO-HIS-HGF-HFWorkingPaper-19.4>, Pristupano 23.08.2022.

<sup>10</sup> World Health Organization (2021) Global expenditure on health: Public spending on the rise?, Dostupno na: <https://apps.who.int/iris/bitstream/handle/10665/350560/9789240041219-eng.pdf>, Pristupano 15.09.2022.

<sup>11</sup> Unified Data Lake for Healthcare Research and Analytics, Dostupno na: <https://www.bluematrix.com/post/unified-data-lake-for-healthcare-research-and-analytics> Pristupano 28.06.2022.

<sup>12</sup> OECD iLibrary, Health expenditure per capita, Dostupno na: <https://www.oecd-ilibrary.org/sites/154e8143-en/index.html?itemId=/content/component/154e8143-en>, Pristupano 23.08.2022.

Grafikon 2 Izdaci za zdravstvo po osobi u zemljama članicama OECD-a u 2019.god



Izvor: OECD Health Statistics

U mnogim zemljama svijeta raspravlja se o pitanju dostupnosti zdravstvenih podataka u više sustava unutar zdravstva. Potiče se standardizacija i izrada okvira kojim bi se omogućilo dijeljenje zdravstvenih podataka elektroničkim putem na način da liječnici iz različitih odjela imaju pristup pacijentovoj povijesti liječenja i medicinskoj dokumentaciji. U Sjedinjenim Američkim Državama interoperabilnost zdravstvenih podataka između zdravstvenih organizacija je već u primjeni i ima višestruke koristi.

Liječnici i zdravstvene organizacije u posljednje vrijeme sve češće naglašavaju kako je vidljiva promjena u pristupu pacijenata prilikom interakcije s zdravstvenim radnicima. Pacijenti se sve češće odnose prema zdravstvenim uslugama kao potrošači i pokazuju angažman, žele imati kontrolu nad liječenjem i njihova informiranost i očekivanja su na visokom nivou.

Zdravstvo je na izazove poput ovih odlučilo odgovoriti primjenom digitalnih tehnologija na optimalan način za pacijente, zdravstvene radnike i sustav u cijelosti.

### 2.3. Digitalna transformacija i poslovni zahtjevi

Razvoj informacijsko – komunikacijskih tehnologija je već od prošlog stoljeća donio mnogobrojne promjene u svakodnevnim životima svih nas. Danas pak, živimo u digitalnom svijetu gdje se promjene događaju nezabilježnom brzinom i tako postavljaju nove standarde u poslovanju na koje treba pronaći primjeren odgovor.



Digitalna transformacija poslovanja se odnosi na intenzivnu primjenu digitalne tehnologije i digitalnih resursa u svrhu stvaranja novih izvora prihoda, novih poslovnih modela i, općenito, novih načina poslovanja. Radi se o velikoj promjeni (transformaciji) načina poslovanja, organizacijske kulture i vrijednosti uz primjenu digitalnih tehnologija koje povezuju poslovne procese omogućujući konkurensku prednost.<sup>13</sup> Za organizacije i poduzeća koja žele primijeniti digitalne tehnologije u svoje poslovne modele vrlo je važno da stalno integriraju i primjenjuju digitalne tehnologije istovremeno.<sup>14</sup>

Zdravstveni sustavi se diljem svijeta suočavaju s problemima u pružanju kvalitetne i dostupne zdravstvene skrbi zbog promijenjene demografske slike i sve starijeg stanovništva. Današnje društvo je visoko informirano i na zdravlje se gleda puno obzirnije pa tako pacijenti očekuju i podrazumijevaju učinkovitu, pravednu i dostupnu zdravstvenu skrb.

U razvijenim zemljama je sve veći broj pacijenata starije životne dobi koji očekuju personaliziranu skrb a liste čekanja su sve duže. Izazovi poput ovih se često prelamaju na zdravstvenim radnicima koji rade pojačanim intenzitetom a od kojih se očekuje visoka kvaliteta u obavljanju usluga. Jedan od velikih problema za zdravstvenu djelatnost su i visoki troškovi u pružanju zdravstvene skrbi. Tradicionalni procesi više ne ispunjavaju sve zahtjeve pacijenata i zdravstveni resori i organizacije rješenja traže u novim tehnologijama i pristupima poput telemedicine, strojnog učenja, robotske automatizacije procesa, umjetne inteligencije, interneta stvari itd. Koristeći navedene tehnologije, transformacija se očekuje na svim razinama, od administrativne razine do poslovnih strategija i pristupa u liječenju.

Odluke su se u zdravstvu uvijek donosile na temelju podataka a sada liječnici mogu imati bolju dijagnostiku pomoću digitalnih tehnologija i odrediti najbolju terapiju pacijentu. Praćenje zdravstvenog stanja pacijenta u stvarnom vremenu bez da je fizički u bolnici može smanjiti redove čekanja a operativna inteligencija omogućuje praćenje troškova i njihovo smanjenje.<sup>15</sup>

---

<sup>13</sup>Spremić, M. (2017). Digitalna transformacija poslovanja. Zagreb:Ekonomski fakultet Zagreb. str.16

<sup>14</sup>Pejić Bach, M., Spremić, M. & Suša Vugec, D. (2018) Integrating Digital Transformation Strategies into Firms: Values, Routes and Best Practice Examples. U: Novo Melo, P. & Machado, C. (ur.) Management and Technological Challenges in the Digital Age. Boca Raton, Florida, Taylor & Francis Group: CRC press, str. 6

<sup>15</sup>Gopal, G., Suter-Crazzolara, C., Toldo, L. and Eberhardt, W. (2019) Digital transformation in healthcare – architectures of present and future information technologies. Clinical Chemistry and Laboratory Medicine (CCLM), Vol. 57 (Issue 3), pp. 328-335. <https://doi.org/10.1515/cclm-2018-0658>

Napretkom tehnologija je došlo i do pitanja interoperabilnosti podataka u zdravstvu. Interoperabilnost podataka u zdravstvu se odnosi na mogućnost dijeljenja podataka o pacijentima elektroničkim putem u više sustava, što donedavno nije bio slučaj. Zdravstveni djelatnici su se često susretali s problemima nedostupnosti podataka o pacijentu u sustavu pa su se oslanjali na dokumentaciju koju fizički donose pacijenti nakon obavljenih pregleda.

Slika 3 Interoperabilnost u zdravstvu



Izvor: Obrada autora prema Cabot Solutions

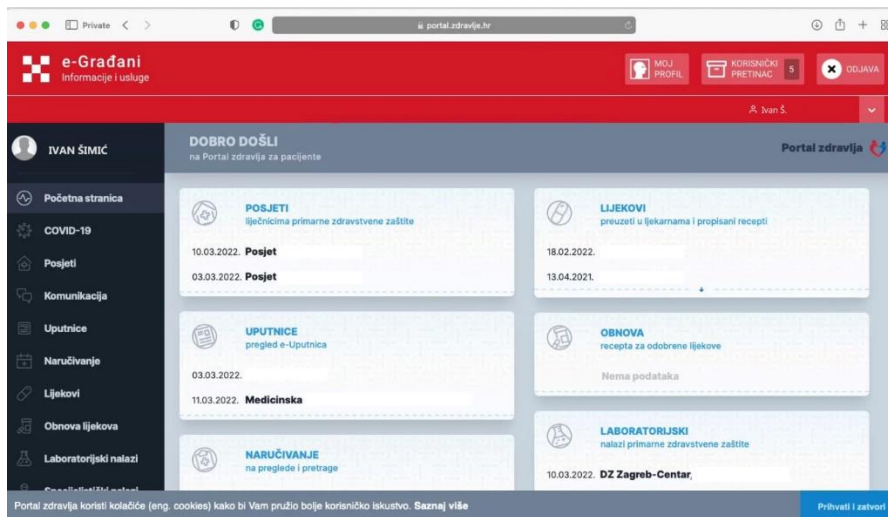
Pacijenti žele imati pristup svom online zdravstvenom kartonu i svim zapisima koji se tiču njihovog zdravstvenog stanja. Upravo je u Europskoj uniji pitanje interoperabilnosti zdravstvenih podataka u punom jeku i pokušava se riješiti kroz Europski okvir interoperabilnosti e-zdravlja.<sup>16</sup> U Hrvatskoj već postoji Portal zdravlja koji je u kreiran u sklopu portala e-Građani i pruža zdravstvene podatke poput propisanih recepata, posjeta liječnicima, laboratorijskih nalaza itd.<sup>17</sup>

<sup>16</sup> Cocir.org, eHealth Interoperability, Dostupno na: <https://www.cocir.org/regulations/digital-health/ehealth-interoperability.html>, Pristupano 23.08.2022.

<sup>17</sup> Ucionica.net, e-Građani Portal zdravlja: kako do nalaza, uputnica, lijekova, Dostupno na: <https://www.ucionica.net/aplikacije/e-gradani-portal-zdravlja-kako-do-nalaza-uputnica-lijekova-9971/>, Pristupano 23.08.2022.

Elektronički zdravstveni zapisi predstavljaju skup podataka i informacija o zdravstvenom stanju korisnika zdravstvene zaštite koji se pohranjuju i prenose u elektronskom obliku na zaštićen način i uz pristup osiguran isključivo ovlaštenim osobama.<sup>18</sup> Ovaj tip tehnologije također omogućava zdravstvenim djelatnicima da prate pacijenta na više načina; od praćenja zdravstvenog stanja preko osobnih informacija, plaćanja i proknjižavanja uplata pa sve do kontrole prijevara u zdravstvenom osiguranju.

Slika 4 Portal Zdravlje



Izvor: Portal Zdravlje

Zbog velikih troškova u zdravstvu koji se stvaraju kroz procese poput nabave opreme i naručivanja lijekova potrebno je upravljati lancem opskrbe uz pomoć raznih softvera koji mogu pratiti cijene dobavljača, vrijeme i kvalitetu dostave.

Analizom podataka koje zdravstvena organizacija dobije pomoću softvera, mogu se izabrati najprikladniji dobavljači i tako uštediti znatna sredstva koja se mogu prenamijeniti u područja gdje je najpotrebnija pomoć sustavu. Pomoću analitičkih alata za knjigovodstvo može se vidjeti dublji uvid u financijsku situaciju.

Takvi softveri mogu dati odgovore na pitanja gdje najviše odlazi novac, koji pacijenti su financijski najzahtjevniji u pogledu medicinskih troškova i gdje se novac gubi bez pokrića.

<sup>18</sup> Kern, J., Elektronički zdravstveni zapis kao predmet interesa korisnika zdravstvene zaštite, Dostupno na: <https://hrcak.srce.hr/file/309835>, Pristupano 24.08.2022.

Transformacija zdravstva ne implicira samo usvajanje i primjenu novih alata i tehnologija već kreiranje zdravstvenog okvira prilagođenog digitalnom dobu pomoću Velikih podataka i novih poslovnih modela koji poboljšavaju rezultate liječenja, smanjuju troškove i poboljšavaju iskustva pacijenata i osoblja.<sup>19</sup> Digitalna transformacija u zdravstvu je složen proces koji zahtijeva plan i strategiju a ne ad hoc pristup. Iako je većina organizacija nekim dijelom započela svoj put digitalne transformacije, taj proces je još uvijek u svojim počecima ali zahvaljujući tehnologijama koje se neprekidno nadopunjuju i razvijaju, za očekivati je da ćemo u skoroj budućnosti imati pristupačnije i efikasnije zdravstvo za sve.

## 2.4. Utjecaj pandemije Covid – 19 na zdravstvo

Pandemija COVID – 19 je početkom 2020. godine snažno pogodila cijeli svijet i pokazala brojne slabosti u zdravstvenim sustavima diljem svijeta. Mjere poput ograničavanja kretanja u počecima pandemije su utjecale i na snažnu redukciju zdravstvenih usluga i otkazivanje zdravstvenih pregleda i operativnih zahvata koji nisu bili hitnog karaktera. Bolnički sustavi su podnijeli veliki pritisak zbog velikog broja zaraženih a zbog nedostatka radne snage neki sustavi su se i potpuno urušili.

Zdravstveni djelatnici su bili na prvoj crti obrane od virusa i dali su nemjerljiv doprinos u borbi protiv bolesti unatoč njihovom nedovoljnom broju. Zbog nedostatka osoblja, većina europskih zemalja angažirala je studente medicine koji su bili pri kraju studija a neke zemlje su zaposlile umirovljene liječnike.<sup>20</sup> Mnogi od njih su osjetili pandemiju i na svom zdravstvenom stanju pa su tako pored zaraze, zdravstveni sindikati prijavljivali depresivna stanja, visoku dozu anksioznosti kao i povišenu stopu samoubojstava kod bolničkog osoblja.<sup>21</sup>

Kako je vrijeme prolazilo, sustavi zdravstvene zaštite su pokušavali naći rješenja za nesmetano odvijanje primarne zdravstvene zaštite, za skrb virusom zaraženih pacijenata ali i pacijenata

---

<sup>19</sup> Philips.com, Dostupno na: <https://www.philips.com/c-dam/corporate/newscenter/global/standard/resources/healthcare/2021/digital-transformation/philips-digital-transformation-position-paper.pdf> , Pristupano 23.08.2022.

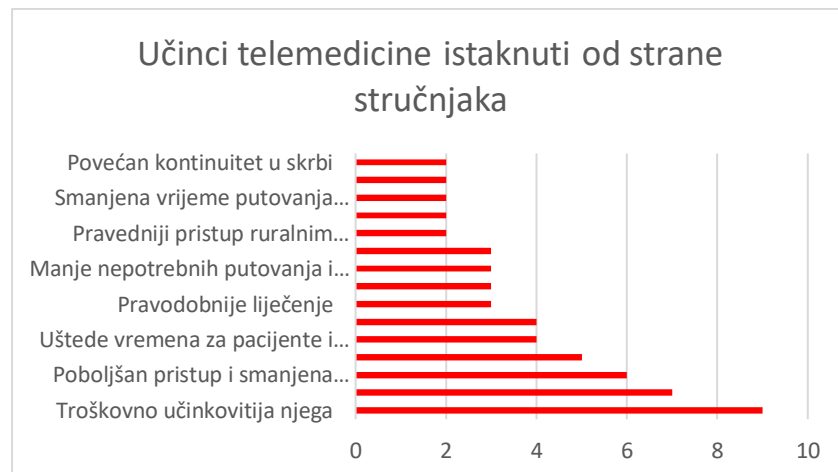
<sup>20</sup> Europska komisija, Health at a Glance: Europe 2020, State of health in the EU cycle, Dostupno na: [https://health.ec.europa.eu/system/files/2020-12/2020\\_healthatglance\\_rep\\_en\\_0.pdf](https://health.ec.europa.eu/system/files/2020-12/2020_healthatglance_rep_en_0.pdf), Pristupano 25.08.2022.

<sup>21</sup> Hall, Heather MBA, MA, MPAS, PA-C. The effect of the COVID-19 pandemic on healthcare workers' mental health. JAAPA: July 2020 - Volume 33 - Issue 7 - p 45-48 doi: 10.1097/01.JAA.0000669772.78848.8c

koji su imali zakazane operacije ili neke druge bolesti. Jedini racionalan i moguć odgovor na probleme u zdravstvu bila je primjena dostupnih tehnologija. Iako su neki koraci u digitalizaciji zdravstva već bili realizirani, pandemija je svakako taj proces ubrzala. Zamjenu posjeta liječnicima i klasičnu komunikaciju između pacijenta i liječnika, omogućila je telemedicina pa su tako putem digitalnih kanala pacijenti mogli dobiti savjete od svojih liječnika, naručiti se na preglede i slati svoju medicinsku dokumentaciju na pregled. Stručnjaci iz 13 zemalja članica Organizacije za ekonomsku suradnju i razvoj (OECD) proveli su istraživanje o utjecaju telemedicinese na zdravstvo i pacijente. U svojim zaključcima su naveli kako je telemedicina u posljednje vrijeme napravila zaokret i da su učinci već sada veliki te da se uz pravilno upravljanje tehnologijom mogu očekivati velike uštede, povećana efikasnost sustava kao i pravednija skrb.<sup>22</sup>

Intenzivna primjena telemedicinese tijekom pandemije COVID - 19 kao i predviđanja njenog korištenja u budućnosti, služe kao jasan dokaz kako je zdravstvo u fazi velikih promjena u poslovanju. Smatra se kako je visokokvalitetna njega ona koja je zasnovana na dobrom odnosu između pacijenata i liječnika. Telemedicina bi svakako trebala biti opcija koja će se koristiti na najbolji mogući način ali osnovni način pružanja skrbi u ordinaciji bi se trebao nastaviti koliko god je moguće kada su u pitanju kompleksni zdravstveni problemi.

Grafikon 3 Prednosti u pružanju skrbi telemedicinom



Izvor: Obrada autora prema OECD

<sup>22</sup> Oliveira Hashiguchi, T. (2020), "Bringing health care to the patient: An overview of the use of telemedicine in OECD countries", *OECD Health Working Papers*, No. 116, OECD Publishing, Paris, <https://doi.org/10.1787/8e56ede7-en>.

### 3. JEZERO PODATAKA

#### 3.1. Povijest i koncept jezera podataka

Koncept jezera podataka se prvi put pojavio u 2010. godini kada je predstavljen od Jamesa Dixona kao spremište u koje mogu ulaziti sirovi, neobrađeni podaci u velikim količinama. Nezabilježene količine i brzine generiranja nestrukturiranih podataka zahtijevale su revoluciju u pohrani i obradi podataka. Dotadašnje, tradicionalne baze podataka i softveri za poslovnu inteligenciju tog vremena se nisu mogli nositi s novim zahtjevima tržišta.

Jezero podataka je prošlo dug put razvoja od svog nastanka 2010. do danas. Iako je koncept relativno novijeg datuma može se činiti da je u pitanju potpuno novi model za skladištenje i obradu podataka ali to nije slučaj. Koncept jezera podataka je sustav koji je danas dobro uspostavljen obrazac dizajna i arhitekture podataka za pripreme duboke pripreme u skladištenju podataka, izvještavanju, znanosti o podacima i naprednoj analitici.<sup>23</sup> Dizajniranje ovog sustava je motivirano potrebom za pohranjivanjem velike količine uz troškovnu učinkovitost, a primjena tehnologija koje iskorištavaju podatke iz jezera se dogodila tek par godina poslije. Ovaj sustav omogućava skupljanje podataka za obradu kako bi zadovoljio zahtjeve rada sa velikim skupom podataka različite strukture. Svi potrebni podaci se trajno pohranjuju u jezero podataka u neizmjenjenom obliku.

Tijekom godina, tehnološki napredak omogućio je poboljšanja podatkovne platforme na kojoj se temelji sama tehnologija jezera podataka. Napredak digitalnog svijeta je nametnuo promjene u načinu na koji su jezera podataka dizajnirana, projektirana i postavljena. To znači da mnoga poduzeća moderniziraju svoja jezera podataka kako bi se prilagodile modernim poslovnim zahtjevima. Većina poboljšanja, odnosno promjena događa se na arhitektonskoj razini, što znači da se stari sustavi upravljanja bazama podataka i druge platforme napuštaju u korist modernijih i poboljšanih platformi koje olakšavaju pristup željenim podacima. Replatformiranje dovodi do promjena u arhitekturi jezera podataka kada se podaci

---

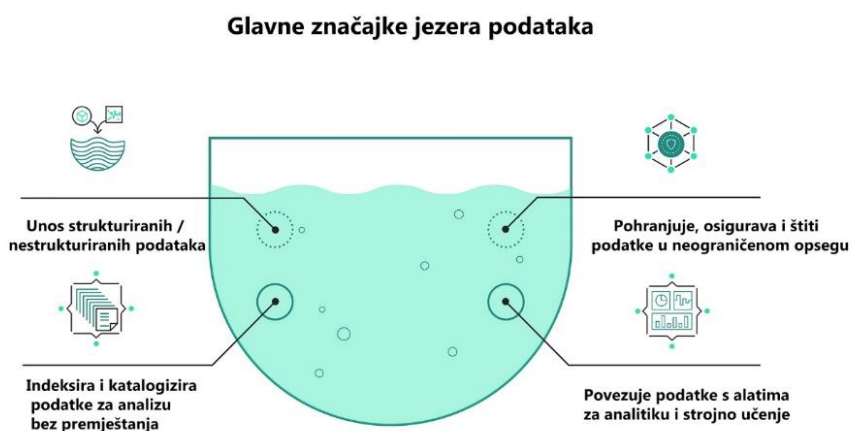
<sup>23</sup> Maini E., Venkateswarlu, B., Gupta, A. (2018). Data Lake-An Optimum Solution for Storage and Analytics of Big Data in Cardiovascular Disease Prediction System *International Journal of Computational Engineering & Management*, 21, 861-869.

redistribuiraju između nove mješavine platformi ili se ponovno moderniziraju i poboljšavaju tijekom migracije podataka.<sup>24</sup>

Jezera podataka u svojim nastajanjima su imala problem neiskorištavanja potencijala iz dostupnih podataka. S vremenom su korisnici jezera podataka ispravili te nedostatke i stvorili veće strukture, bolju kvalitetu vođenja i upravljanja podacima. Glavni cilj poboljšanja jezera podataka bio je napraviti poboljšanja koja neće narušiti duh izvornog jezera podataka. Zatim se poboljšano jezero podataka počinje koristiti kao spremište za ogromne količine neobrađenih podataka koji su prikladni za široko istraživanje, otkrivanje i mnoge analitičke pristupe. Moderno jezero podataka mora služiti širem krugu korisnika i njihovim potrebama.

Kad god se podaci unose zajedno iz različitih izvora i s raznim shemama, metapodaci su zaduženi za praćenje tih podataka. Oni osiguravaju da se podaci kasnije mogu pronaći i koristiti i da im se može vjerovati.<sup>25</sup>

Slika 5 Glavne značajke jezera podataka



Izvor: Obrada autora prema LakeFS

Jezera podataka se odnose na masivno skalabilnu pohranu koja sadrži veliku količinu neobrađenih podataka u izvornom formatu dok ne budu potrebni za obradu. Obično su

<sup>24</sup> Purra S. B., & Pasupuleti P. (2015.) *Data Lake development with Big Data*. Birmingham: Packt Publishing.

<sup>25</sup> Giebler, C., Gröger, C., Hoos, E., Schwarz, H. and Mitschang, B., 2019, August. Leveraging the data lake: Current state and challenges. In *International Conference on Big Data Analytics and Knowledge Discovery* (pp. 179-188). Springer, Cham.

izgrađena za rukovanje velikim količinama nestrukturiranih podataka koji brzo dolaze (za razliku od skladišta podataka koje karakteriziraju visoko strukturirani podaci) iz kojih se mogu izvesti dodatne informacije. Zbog toga jezera koriste dinamiku analitičke aplikacije<sup>26</sup>. Jezera podataka vrlo često uključuju semantičku bazu podataka kao i konceptualni model koji koristi iste standarde i tehnologije korištene za stvaranje internetskih hiperveza i dodaje sloj konteksta preko podataka koji definira značenje podataka i njihove međusobne veze s drugim podacima. Jezero podataka je mjesto za izvođenje preliminarne obrade podataka, dok se fleksibilno strukturiranje orijentirano na zadatak provodi samo tamo gdje je potrebno. Jezero podataka, kao i svaki drugi sustav, ima puno svojih prednosti, ali i nedostataka koji utječu na njegov rad a u nastavku su navedene glavne prednosti i rizici.

Tablica 1 Prednosti i nedostaci korištenja jezera podataka

Prednosti	Nedostaci
Nudi ekonomičnu skalabilnost i fleksibilnost	Nestrukturirani podaci mogu dovesti do neupotrebljivih podataka
Nudi vrijednosti iz neograničenih vrsta podataka	Ne postoji način da dobijete uvid od drugih koji su radili s podacima
Smanjuje dugoročne troškove	Unos podataka bez ikakvog nadzora
Omogućuje ekonomično pohranjivanje datoteka	Velika opasnost od nastajanja močvare podataka
Brza prilagodljivost promjenama	
Centralizacija različitih izvora sadržaja	
Fleksibilan pristup podacima	

Izvor: Obrada autora prema Stein (2014)

<sup>26</sup> Stein, B., Morrison, A. (2014). The enterprise data lake: Better integration and deeper analytics. Price water house Cooper., Dostupno na: <http://www.pwc.com/us/en/technology-forecast/2014/cloudcomputing/assets/pdf/pwc-technology-forecast-data-lakes.pdf> , Pristupano 26.08.2022.



### 3.2. Vrste podataka

Jezero podataka je centralizirano spremište koje omogućuje pohranjivanje podataka onakvima kakvi zapravo jesu, odnosno pohranjivanje podataka u izvornom obliku. Postoje različite vrste podataka pohranjenih u jezeru podataka i dijele se na strukturirane (relacijske baze podataka), polustrukturirane (web stranice, e-pošta, komprimirane datoteke) i nestrukturirane (audio i video datoteke, tekstualni dokumenti).

Slika 6 Strukturirani i nestrukturirani podaci



Izvor: Obrada autora prema Igneous

Jezero podataka omogućuje pohranjivanje podataka bez potrebe da se prvo strukturiraju podaci i da se pokrenu bilo koje vrste analitike za pohranjivanje podataka. Jezero se može promatrati kao repozitorij za različite vrste datoteka i objekata, gdje se podaci mogu pohranjivati u raznim oblicima. Velika je mogućnost da u bliskoj budućnosti jezero podataka bude glavno mjesto pohrane svih podataka iz poduzeća ili organizacije.

### 3.3. Skladište podataka i jezero podataka

Ovisno o zahtjevima, tipična organizacija podataka zahtijevat će i pohranu podataka i jezero podataka.

Pojam „skladište podataka“ prvi je upotrijebio Bill Inmon 1990. godine. Prema Inmonu skladište podataka je usmjereno području, integrirano, vremenski obilježeno te nepromjenjivo/neizbrisivo. Skladište podataka je baza podataka optimizirana za analizu relacijskih podataka koji dolaze iz transakcijskih sustava i linija poslovnih aplikacija. Skladišta podataka pružaju opće i konsolidirane podatke u višedimenzionalnom pogledu.<sup>27</sup> Strukture podataka i shema su unaprijed definirane zbog optimizacije pretraživanja SQL upita. Rezultat ovih upita se često koristi za operativno izvješćivanje i analizu. Podaci se obogaćuju, čiste i transformiraju kako bi djelovali kao "jedinствени izvor istine" kojem korisnici mogu pouzdano vjerovati.

Jezero podataka predstavlja svojstveno proširenje u konceptu skladištenja podataka i oslanja se na klasično skladište podataka. Drugačiji je jer osim strukturiranih podataka pohranjuje i nestrukturirane podatke čiji su izvori mobilne aplikacije, IoT senzori ili društvene mreže. Struktura podataka ili shema u jezeru podataka je nedefinirana, što znači da je moguće pohraniti sve podatke bez pažljivog dizajna ili potrebe da korisnici znaju svrhu za koju će im dati podaci trebati. Također, u jezerima podataka postoji demokratizacija podataka odnosno davanje pristupa podacima svim zaposlenima ili članovima organizacije, za razliku od skladišta podataka.

Nestrukturirani se podaci pretražuju na drugačiji način, pomoću strojnog učenja, algoritama pretraživanja teksta i analitike velikih podataka. Strategije jezera podataka mogu kombinirati SQL i NoSQL pristupe bazama podataka i mogućnosti mrežne analitičke obrade (OLAP) i mrežne obrade transakcija (OLTP). Kod skladišta podataka su prisutne datoteke za pohranu podataka dok je kod jezera podataka arhitektura ravna i svaki podatak ima metapodatak i identifikator uz sebe.<sup>28</sup>

---

<sup>27</sup> Ćurko, K. i Španić Kazen, M. (2016.), Skladištenje podataka, put do znanja i poslovne inteligencije, Ekonomski fakultet u Zagrebu

<sup>28</sup> Miloslavskaya N, and Tolstoy A., (2016.), Big Data, Fast Data and Data Lake Concepts, preuzeto s [https://www.researchgate.net/publication/309183107\\_Big\\_Data\\_Fast\\_Data\\_and\\_Data\\_Lake\\_Concepts](https://www.researchgate.net/publication/309183107_Big_Data_Fast_Data_and_Data_Lake_Concepts) (Datum pristupa: 24.08.2022.)

Tablica 2 Usporedba skladišta podataka i jezera podataka

	SKLADIŠTE PODATAKA	JEZERO PODATAKA
Podaci	Strukturirani, obrađeni	Neobrađeni, sirovi
Obrada podataka	Relacijske strukture podataka	Schema – on - read
Troškovi	Visoki	Relativno niski
Agilnost	Visoko strukturiran (promjena oduzima puno vremena)	Vrlo fleksibilan (lako se mijenja konfiguracija)
Sigurnost	Visoko sigurnosni protokoli	Zadovoljavajuća sigurnost, neka rješenja su u razvoju
Korisnici	IT / Poslovni analitičari	Podatkovni znanstvenici
Granularnost podataka	Podaci na niskoj razini detalja (granularnosti)	Podaci na sažetoj ili zbirnoj razini detalja (granularnosti)
Potencijalni problemi	Upravljanje troškovima pohrane i upita	Nastanak močvara podataka

Izvor: Obrada autora prema Adviso, 2020

Močvara podataka (*engl. Data swamp*) je proizašla iz jezera podataka. Nekontrolirano pohranjivanje podataka, odnosno pohranjivanje podataka bez ograničenja, bilo kakvih očekivanja i smjernica za pohranu, jezera podataka pretvara u močvaru podataka. Iz močvara podataka se može izvući jako malo kvalitetnih podataka tj. naknadnih koristi kroz analize. Močvara podataka je sveobuhvatna pohrana podataka i nastaje kada korisnici (organizacija koja pohranjuje podatke) ne znaju kako kategorizirati potrebne podatke. To nam govori da močvara podataka pohranjuje nepotrebne i zastarjele podatke, jer korisnici ne vode računa o

tome što unose u jezero podataka i to rade nekontrolirano, bez postavljenih smjernica relevantnosti. Močvarama podataka ne upravljaju redovni administratori i analitičari, kao što je slučaj s jezerima podataka. Prvi razlog zašto se podaci pronađeni u močvarama podataka ne mogu koristiti u analitici velikih podataka je taj što se podaci u močvarama pohranjuju bez ikakve kontrole i kategorizacije koja bi vodila korisnike, a drugi razlog je nedostatak metapodataka što korisnicima otežava pretraživanje ili organiziranje podataka.

Močvare podataka su opasnost za narušavanje usklađenosti podataka jer skrivaju detaljnije podatke, koji su korisnicima vrlo važni za daljnji rad.<sup>29</sup> Vrlo je važno spriječiti da jezero podataka ne postane močvara podataka koja u svojim dubinama skriva mnoštvo podataka koji se ne mogu pronaći kada su potrebni za daljnju transformaciju.

Jezera podataka trebaju dobru organizaciju i vođenje kako bi se omogućilo pružanje korisnih i relevantnih podataka. Jezera podataka moraju imati metapodatke koji se rutinski unose u sustav pohrane, a podaci moraju biti regulirani, precizni i dobro organizirani.

### **3.4. Arhitektura jezera podataka**

Arhitektura jezera predstavlja kombinaciju alata i tehnologija koje pomažu u kreiranju cjeline u kojoj se mogu obrađivati podaci. Radi se o alatima za prihvatanje i unos podataka, zatim alata za čišćenje podataka i izvlačenje značajki iz istih pa sve do analitičkih alata i alata za pretraživanje.

Podaci u jezero podataka ulaze iz više izvora a neki od podataka mogu biti iz sustava evidencije, streaming servisa, aplikacija koje daju serijske interne podatke i eksternih izvora. Jezero podataka se “punilo” tako što podaci, koji mogu biti različitih vrsta, ulaze u jezero putem zajedničkog sloja Unos podataka.

Kako se radi o podacima različitih vrsta, poželjno je standardizirati podatke i centralizirati ih u zajedničko spremište za pohranu. Iako ovo nije nužno, ima svoje prednosti jer analitičari ističu kako je bolje raditi s kopiranim podacima i njihovim izvorima a ne s originalnim izvorima podataka jer u daljnjem tijeku procesa imaju veću slobodu prilikom analiza.

---

<sup>29</sup> Brackenbury, W. & Liu, R., & Mondal, M., Elmore, A., Ur, B., Chard, K., & Franklin, M. (2018). Draining the Data Swamp: A Similarity-based Approach. *Hilda* '18, 10, 1-7.

Slika 7 Prikaz arhitekture jezera podataka



Izvor: Obrada autora prema IBM

Ponekad je potrebno očistiti podatke i pripremiti ih za korištenje a ukoliko se žele kreirati nove značajke radi se izvlačenje značajki iz podataka kako bi se kasnije mogle kreirati i analizirati nove značajke. Nakon što su podaci očišćeni i kreirane značajke za daljnje analize, radi se napredna analitika i vježba modela strojnog učenja. Prilikom svake od navedenih radnji, kreiraju se novi derivirani setovi podataka koji su povezani s prvotnim setom podataka. U slučaju da je bila potreba za ispravljanjem određenih podataka, ispravljeni podaci bi bili povezani sa onima na početku procesa. Ova veza je jako bitna i predstavlja bitnu ulogu u vladanju podacima.

Vladanje podacima je zaduženo za prikupljanje metapodataka i za razumijevanje odnosa između podataka kroz različite formate, primjerice kroz tablice.<sup>30</sup> Potrebno je osigurati provođenje organizacijskih politika na način da se podaci koriste na način koji će organizaciju obilježavati i donijeti koristi u budućnosti.

Nakon što su u jezeru podataka napravljene sve potrebne radnje, u sloju Upotreba podataka se događaju ključne stvari zbog kojih jezero podataka i postoji. Postoje razne mogućnosti koje se mogu primijeniti uz pomoć analiziranih podataka a neke od njih su nadzorne ploče koje pomažu u donošenju racionalnijih odluka. Nadzorne ploče mogu odgovoriti na pitanja budućih investicija i novih projekata.

<sup>30</sup> IBM, Data architecture:IBM's POV – IBM Cloud Architecture Center, Dostupno na: <https://www.ibm.com/cloud/architecture/architectures/dataArchitecture/>, Pristupano 28.08.2022.

Pomoću strojnog učenja se mogu dizajnirati pametne aplikacije koje pružaju vrijedne preporuke korisnicima tih aplikacija na osnovu podataka o njihovim prethodnim kupnjama ili ponašanju. Još jedna od korisnih stvari koja se može primijeniti u zadnjem sloju arhitekture jezera podataka je automatizacija procesa. Inteligentni modeli mogu automatizirati repetitivne poslovne procese i kreirati bolje iskustvo pomoću obogaćenih podataka u jezeru podataka.

Cijeli proces se ponavlja više puta jer nove pametne aplikacije kreiraju nove podatke koji ponovno ulaze u jezero podataka i započinju isti process. Arhitektura jezera podataka se može objasniti i kroz četiri procesa. Prvi proces je bio skupljanje podataka u sloju Unos podataka, zatim slijedi proces organizacije kroz pripremu i izvlačenje značajki iz podataka iz spremišta podataka zajedno s treniranjem modela strojnog učenja i vladanjem podaci. Naposljetku je proces unosa saznanja i uvida iz jezera podataka u aplikacije odnosno sloj Upotreba podataka.

### **3.5. Vladanje podacima (*engl. Data governance*)**

Vladanje podacima na efikasan način je ključno ukoliko korisnik ne želi da od jezera podataka dobije močvaru podataka. Prije nego se odluči koji sustav pohrane želi imati u svojoj organizaciji, potrebno je imati strategiju vladanja podacima.

Iako su mnogi podatkovni analitičari govorili kako je za jezera podataka potrebna minimalna kontrola i usmjeravanje fokusa na modele strojnog učenja i umjetne inteligencije, sve više stručnjaka se ne slaže s tom tezom. Naime, mnoge organizacije su krenule u pokretanje jezera podataka i na kraju završili s velikim troškom i močvarom podataka. Vladanje podacima se vrši na temelju internih podatkovnih standarda i politika koje kontroliraju korištenje podataka. Vrlo je važno da samo vladanje podacima bude učinkovito, jer se tada osigurava konzistentnost i pouzdanost podataka, kao i sigurnost da se podaci ne zlouporabe. Dobro osmišljen program vladanja podacima vrlo često uključuje upravljački tim koji djeluje kao upravljačko tijelo. Kao temeljna komponenta cjelokupne strategije upravljanja podacima je vladanje podacima, ali ono ne može funkcionirati bez korisnika koji se trebaju fokusirati na očekivane poslovne koristi koje očekuju od samog sustava. Bez učinkovitog vladanja podacima neće biti moguće formirati konzistentne podatke koji su nužni za dobro funkcioniranje sustava, odnosno jezera podataka. Također, loše vladanje podacima može spriječiti inicijative koje su potrebne za usklađivanje s propisima koji su nametnuti. Takav način funkcioniranja mogao bi poduzeću koje koristi taj

sustav stvoriti određene probleme u samom korištenju podataka. Vladanje podacima uključuje razvoj zajedničkih definicija podataka i standardnih formata podataka koji se primjenjuju u svim poslovnim sustavima. Schneider i Vom Brocke razlikuju tri kategorije mehanizama vladanja podacima.<sup>31</sup>

- Strukturalni mehanizam u odnosu na strukture vladanja
- Proceduralni mehanizam povezan s politikama za vladanja podacima
- Mehanizam odnosa koji se odnosi na komunikaciju dioničara

Glavni cilj vladanja podacima je razbiti silose podataka u organizaciji. Ovi silosi najčešće nastaju kada pojedine poslovne jedinice primjenjuju zasebne sustave obrade podataka bez centralizirane koordinacije, odnosno podatkovne arhitekture tvrtke. Vladanje podacima ima za cilj uskladiti podatke u tim sustavima kroz proces suradnje. Drugi cilj vladanja podacima je osigurati pravilnu upotrebu podataka, kao i izbjeći unošenje grešaka u sustav. To znači da vladanje podacima automatski blokira potencijalnu zlouporabu podataka. To se postiže stvaranjem jedinstvenih pravila o korištenju podataka zajedno s postupcima za kontinuirano praćenje, korištenje i provedbu pravila. Osim toga, vladanje podacima može pomoći pri uspostavljanju režima između prakse i privatnosti. Uz točniju analitiku i veću usklađenost podataka, vladanje podacima također uključuje poboljšanje kvalitete podataka, niže troškove upravljanja podacima i povećan pristup podacima za znanstvenike, analitičare i poslovne korisnike. Vladanje podacima pomaže poboljšati donošenje poslovnih odluka pružajući korisnicima bolje i točnije podatke.

Neke od najčešćih izazova u vladanju jezerom podataka su identifikacija i održavanje ispravnih izvora podataka. Ukoliko metapodaci nisu dostupni cijeli sadržaj jezera podataka može biti doveden u pitanje. Mnoge organizacije učitavaju podatke u jezero podataka bez provjere valjanosti podataka i to čini podatke manje korisnima u analitičkom dijelu. Često je veliki problem loša koordinacija u kvaliteti podataka i vladanju podacima što može dovesti do loših poslovnih odluka.

Okvir za vladanje podacima sastoji se od politika korištenja, pravila, procesa, organizacijskih struktura i tehnologija koje su postavljene kao dio programa upravljanja. Okvir za vladanje podacima treba dokumentirati i interno dijeliti, tako da bude dostupan svim korisnicima koji

---

<sup>31</sup> Derakhshannia, M. & Gervet, C. & Hajj-Hassan, H. & Laurent, A, Martin, A. (2020). Data Lake Governance: Towards a Systemic and Natural Ecosystem Analogy. *Future Internet*. 12. 126. doi: 10.3390/fi12080126.

su uključeni u proces korištenja tog sustava vladanja. Vrlo važna komponenta vladanja podacima je komponenta koja se odnosi na inicijativu digitalne transformacije i može pomoći drugim korporativnim procesima, poput upravljanja rizicima, poslovnih procesa i upravljanja podacima. Gledano s tehnološke strane, sustav vladanja podacima može se koristiti za automatizaciju aspekata upravljanja programom vladanja podacima. Alati za upravljanje sustavom upravljanja podacima nisu obavezna komponenta okvira upravljanja, ali se mogu koristiti zajedno s alatima za kvalitetu podataka, upravljanje metapodacima i matičnim podacima. Kako se korištenje podataka širi i razvija, tako rastu i nove tehnologije vladanja podacima koje dobivaju sve širu primjenu.

### **3.6. Sustavi za pohranu podataka u jezeru podataka**

Za očuvanje sigurnosti i trajnosti podataka od posebne je važnosti sustav pohrane podataka u jezeru podataka. Ono što je dobra karakteristika jezera podataka je to da se mogu izgraditi hibridni pristupi pohranjivanju koji omogućavaju korištenje prednosti određenih vrsta pohranjivanja bez da postoji obavezan odnos prema njima.

Postoji niz različitih vrsta pohrane koje se mogu integrirati u arhitekturu jezera podataka ili širu analitičku infrastrukturu. To uključuje slojeve pohrane koji su uvelike neovisni o računalnim kapacitetima kao i pristupe koji integriraju to dvoje i mogu pohraniti podatke u zaštićenom formatu.

Distribuirani sustav datoteka (DFS) je najjednostavniji oblik pohrane izvan pohrane na lokalnom disku. Distribuirani datotečni sustavi obično pružaju toleranciju na pogreške, redundanciju i mogućnost paralelnog čitanja i pisanja podataka iz klasteriranih čvorova. U ovom slučaju, podaci se pohranjuju u tekstualne datoteke i pohranjuju se ili u tabličnom, razgraničenom formatu ili u strukturiranijem formatu kao što je JSON. Prisutan je u Amazon HealthLake i Amazon S3 platformi.

Datoteke s ravnim tekstom mogu biti učinkovite za određene slučajeve korištenja. Konkretno, komprimirani formati orijentirani na stupce s particioniranjem imaju vrlo dobre performanse za niz analitičkih aplikacija. Dohvat i prikupljanje podataka je ubrzano jer se dohvat vrši za samo potrebne stupce, a podaci u istom stupcu su kontinuirani.



NoSQL baze podataka i platforme temeljene na indeksu pretraživanja mogu biti korisne za brzo generiranje metrike za velike skupove podataka. Karakteristike izvedbe i značajke ovih platformi uvelike se razlikuju, stoga je najbolje u potpunosti razumjeti specifične potrebe i prepustiti stručnjaku

### **3.7. Vodeća tehnološka rješenja**

Hadoop je važan element arhitekture jezera podataka koji se koristi za izgradnju jezera podataka. Riječ je o jezeru podataka koje je izgrađeno na platformi koja se sastoji od Hadoop klastera. Hadoop u jezeru podataka čuva podatke u izvornom obliku i prati promjene podataka u životnom ciklusu podataka. Ovaj je pristup posebno koristan za unutarnje revizije a korisnici Hadoopa su često napredni podatkovni znanstvenici s iskustvom u analitici, bazama podataka i rudarenju podataka.<sup>32</sup>

Hadoop je postao posebno popularan u arhitekturi jezera podataka jer je otvorenog koda. Korištenje Hadoop jezera podataka značajno smanjuje troškove izgradnje velikih skladišta podataka. Podaci pohranjeni na Hadoop klasterima su iracionalni i mogu uključivati datoteke, slike, web objave. Može podnijeti velike količine strukturiranih, polustrukturiranih i nestrukturiranih podataka s lakoćom i jeftiniji je za implementaciju budući da može započeti korištenjem običnog hardvera.<sup>33</sup>

Hadoop arhitektura nije izgrađena za obradu transakcija, već je usmjerena na podršku analitičkim aplikacijama. Ovaj sustav omogućuje istovremenu obradu podataka, jer se podaci prilikom unosa razbijaju na segmente i distribuiraju kroz različite čvorove u Hadoop klasteru. Hadoop može sadržavati različite nizove nestrukturiranih, strukturiranih i polustrukturiranih podataka. Ovo jezero podataka pruža potrebne sigurnosne alate i omogućuje sigurniji i sigurniji prijenos podataka.

Izvorna jezera podataka bila su isključivo na Hadoop tehnologiji, što se danas smatra nezadovoljavajućim. Mnogi korisnici koji koriste podatkovna jezera na Hadoopu vidjeli su koliko on ima nedostataka, odnosno da samo korištenje podataka zahtijeva veću i bolju racionalnu funkcionalnost nego što je sam Hadoop nudio.

---

<sup>32</sup> Zicari, R.V, (2014.), Challenges and opportunities u : Akerkar, R., Big Data Computing, Sogndal: Taylor and Francis

<sup>33</sup> Tomcy J., Panhay M. (2018.) Data lake for Enterprises, Birmingham, Packt Publishing Ltd.

Osim Hadoop jezera podataka, postoje i drugi primjeri, poput Azure Data Lake Store ili Amazon S3 rješenja u oblaku.

Azure Data Lake je izgrađen na vrhu Azure Blob pohrane i predstavlja jezero podataka u oblaku. Izgrađeno je prema HDFS standardu i nije ograničeno količinom podataka. Ovo jezero podataka uključuje sve mogućnosti potrebne kako bi programerima, znanstvenicima i analitičarima olakšalo pohranjivanje podataka bilo koje veličine, oblika i brzine te obavljanje svih vrsta obrade i analitike na svim platformama. Sastoji se od 3 glavne značajke: Azure Data Lake Storage, Azure Data Lake Analytics i Azure HDInsight i svaka je zadužena za svoj dio procesa iako zajednički surađuju u razmjeni podataka. Microsoft Azure je zajedno s Amazon S3 jezerom podataka jedna od vodećih tehnologija u spremanju podataka u oblaku.

### **3.8. Veliki podaci i umjetna inteligencija**

Pojam „Veliki podaci“ predstavlja skupove podataka čija je veličina izvan mogućnosti tradicionalnih softverskih alata u sklopu baza podataka za snimanje, pohranu, upravljanje i analizu.<sup>34</sup> Tehnologija Velikih podataka je u posljednjih nekoliko godina snažno unaprijedila i promijenila suvremeno poslovanje. Bitno je naglasiti kako su Veliki podaci kompleksna tehnologija i kako su najvažnije odrednice pored veličine, raznovrsnost i brzina kojom se podaci generiraju, distribuiraju i prikupljaju.

Organizacije i poduzeća se prije pojave Velikih podataka često nisu mogle nositi s izazovima koji su s brzim napretkom tehnologija postajali sve veći. Pojavom i razvojem Velikih podataka mnogi su svoja rješenja našli u primjeni novih tehnologija proizašlih iz Velikih podataka. Koristi od tehnologije Velikih podataka su brojne i svakim novim priljevom podataka se dodatno povećavaju. Brojni stručnjaci navode kako su najvažnije koristi cjeloviti odgovori na mnoga pitanja u svakodnevnom poslovanju i donošenje jasnijih odluka, uz pomoć znanja koje je proizašlo iz analize podataka. Krajnje rezultate od primjene Velikih podataka se često može vidjeti u kratkom vremenskom razdoblju kroz uštede, sigurnije poslovanje i povećanje vrijednosti poduzeća ili organizacije.

---

<sup>34</sup> McKinsey Global Institute (2011), Big Data: The Next Frontier for Innovation, Competition, and Productivity, Dostupno na: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation>, Pristupano 15.09.2022.

Veliki podaci daju mogućnost transformacije i napretka u poslovanju ali postoje i rizici koji dolaze uz njihovu primjenu. Kako bi se bilo kakva korist mogla dobiti iz podataka, podaci se prvo moraju prikupiti. No, posjedovanje velike količine podataka sasvim sigurno nije jamstvo za značajan uspjeh. Naprotiv, može postati i veliki teret te zakomplicirati poslovanje ukoliko podaci nisu kvalitetni i konzistentni, stoga je jako bitno pristupiti svim procesima prikupljanja podataka planirano i temeljito.

Potrebno je razvijati i dizajnirati nove modele za analizu podataka a broj stručnjaka za Velike podatke koji pored informatičkih znanja imaju i znanja iz financija i menadžmenta je nedovoljan, što predstavlja dugoročan problem. Pored ovog problema, postoji i problem brzih promjena u podacima koji se ogleda u činjenici da poduzeća ne mogu tako brzo prikupiti najnovije podatke što se naposljetku očituje kroz zastarjele informacije.<sup>35</sup>

Osim ranije spomenute pohrane podataka, postoje još tri glavne kategorije u tehnologiji Velikih podataka a to su rudarenje podataka, analitika podataka i vizualizacija podataka. Ono u čemu mnogi vide smisao Velikih podataka je krajnji proces u analitici podataka tj. izvlačenje zaključaka iz podataka i prepoznavanje obrazaca koji pomažu u predviđanju i donošenju odluka pomoću umjetne inteligencije.

Pojam „umjetna inteligencija“ se prvi put spominje 1956. godine a danas je izrazito aktualan zbog sve veće količine podataka i mogućnosti primjene ove tehnologije. Označava granu računalne znanosti i uključuje razvoj računanih programa za obavljanje zadataka koji bi inače zahtijevali ljudsku inteligenciju. Algoritmi se mogu baviti učenjem, percepcijom, rješavanjem problema, razumijevanjem jezika i logičkim zaključivanjem.<sup>36</sup>

Umjetna inteligencija automatizira ponavljajuće učenje i otkrića pomoću podataka. Iako se povezuje uz Velike podatke čija je primjena postala trend u proteklih nekoliko godina, umjetna inteligencija je već uvelike prisutna u našem bliskom okruženju. Različita je od robotske automatizacije koja je vođena hardverom jer ne automatizira ručne poslove već elektroničke zadatke ali je još uvijek potrebna osoba koja će postaviti upit. Za umjetnu inteligenciju je karakteristično da već postojećim predmetima daje dodatnu inteligenciju i često se primjenjuje

---

<sup>35</sup> Aburawi, Y., & Albaour, A., Big Data: Review Paper. International Journal Of Advance Research And Innovative Ideas In Education, 7, 2021, 2021.

<sup>36</sup> Saleh, Z. (2019). Artificial Intelligence Definition, Ethics and Standards, Dostupno na: [https://www.researchgate.net/publication/332548325\\_Artificial\\_Intelligence\\_Definition\\_Ethics\\_and\\_Standards](https://www.researchgate.net/publication/332548325_Artificial_Intelligence_Definition_Ethics_and_Standards), Pristupano 15.09.2022.

u sklopu nekog proizvoda ili usluge. Najpoznatiji primjeri umjetne inteligencije se nalaze u velikom broju kućanstava a to su virtualni asistenti Siri, Alexa, Google Assistant i Cortana koji tumače ljudski govor i odgovaraju putem sintetiziranih glasova.

Postoji nekoliko tehnologija umjetne inteligencije a to su:

- Strojno učenje
- Duboko učenje
- Obrada prirodnog jezika
- Računalni vid

Strojno učenje je vrsta umjetne inteligencije koja omogućuje softverskim aplikacijama da postanu preciznije u predviđanju ishoda bez da su za to eksplicitno programirane. Algoritmi strojnog učenja koriste povijesne podatke kao ulaz za predviđanje novih izlaznih vrijednosti. Ova tehnologija ima izniman značaj i veliku primjenu u modernim spremištima podataka, poput jezera podataka.

Postoji više vrsta umjetne inteligencije. Trenutno se upotrebljava umjetna uska inteligencija i to je jedina od tri vrste umjetne inteligencije koja je primjenjena u praksi. Programirana je tako da obavlja zadatak u stvarnom vremenu ali nema mogućnost rada van sustava za koji je dizajnirana. To je razlog zašto je mnogi nazivaju i slabom, nema samosvijest daleko je ispod ljudske inteligencije. Neki od primjera umjetne uske inteligencije su samovozeći automobili, preporuke filmova na streaming servisima, slikovna dijagnostika u medicini itd.<sup>37</sup>

Umjetna opća inteligencija se odnosi na strojeve koji mogu donositi odluke samostalno i na razini ljudske inteligencije. I dalje će postojati komunikacija sa strojevima no oni će biti u mogućnosti donositi zaključke i prijedloge, biti kreativni u svojim rješenjima i integrirati prethodna znanja u buduće odluke.

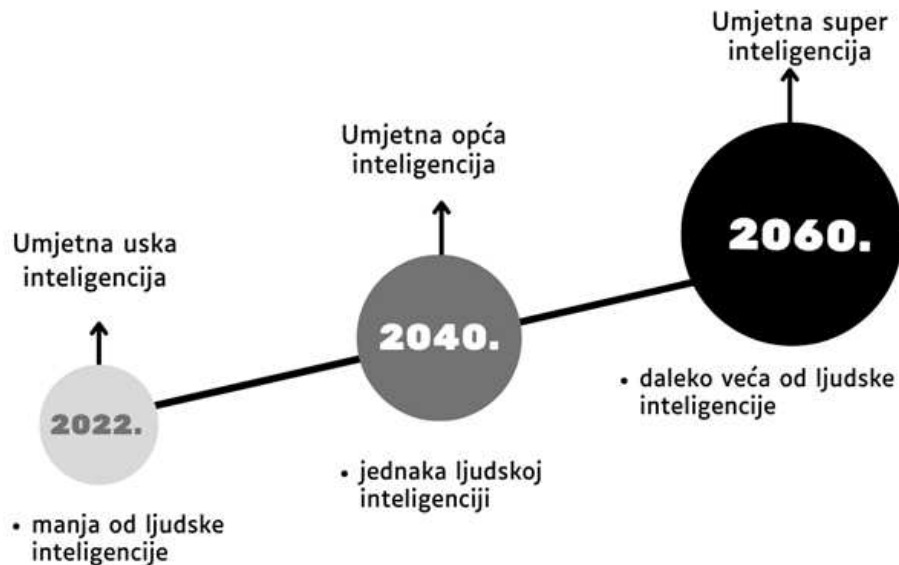
Treća vrsta umjetne inteligencije se zove umjetna super inteligencija. Ona je kao i prethodna još uvijek samo teorijski koncept, ali u svojoj postavi nadmašuje ljudsku inteligenciju. Iako se procjenjuje da će zaživjeti u praksi za manje od 40 godina, još uvijek se ne znaju sve

---

<sup>37</sup> Medium.com (2018.), Distinguishing between Narrow AI, General AI and Super AI , Dostupno na: <https://medium.com/mapping-out-2050/distinguishing-between-narrow-ai-general-ai-and-super-ai-a4bc44172e22>, Pristupano 15.09.2022.

karakteristike koje će posjedovati. Mnogi strahuju od umjetne super inteligencije jer smatraju da će naštetiti ljudskoj rasi i da će ljudi postati taoci svojih izuma.

Slika 8 Vrste umjetne inteligencije



Izvor: Obrada autora prema Artificial Intelligence Definition, Ethics and Standards

Umjetna inteligencija ima cijeli niz svojih prednosti ali i nedostataka.<sup>38</sup> Jedna od najvećih prednosti je svakako smanjenje ljudske pogreške i povećanje točnosti i preciznosti. Smatra se da uz umjetnu inteligenciju stopa pogrešaka može biti smanjena na nulu. Uz to, mnogi procesi koji su u svojoj prirodi opasni mogu biti zamijenjeni robotima iz umjetne inteligencije (npr. deminiranje, putovanje svemirom i oceanima). Za razliku od ljudi, softveri i roboti umjetne inteligencije se ne mogu umarati i mogu postići visoke stope produktivnosti. Isto tako, roboti nemaju emocije i mogu posao obavljati objektivno na osnovu informacija koje skupe.

No, postoje i loše strane umjetne inteligencije. Jedna od njih je svakako pitanje nezaposlenosti. Mnogi vjeruju da će se uvođenjem robota i softvera u sve procese naštetiti zaposlenima samo kako bi se ostvarila maksimalna efikasnost. Umjetna inteligencija bi mogla utjecati na ljude u budućim generacijama na način da manje razmišljaju i propituju svijet oko sebe jer je sve već dostupno i optimizirano.

<sup>38</sup> Simplilearn.com (2022.) Advantages and Disadvantages of Artificial Intelligence, Dostupno na: [https://www.simplilearn.com/advantages-and-disadvantages-of-artificial-intelligence-article#advantages\\_&\\_disadvantages\\_of\\_artificial\\_intelligence](https://www.simplilearn.com/advantages-and-disadvantages-of-artificial-intelligence-article#advantages_&_disadvantages_of_artificial_intelligence), Pristupano 15.09.2022.

## 4. PRIMJENA JEZERA PODATAKA U ZDRAVSTVU

### 4.1. Veliki podaci u zdravstvu

Zdravstvo prolazi kroz velike promjene uzrokovane eksponencijalnim rastom podataka i njihovih izvora. Takvi trendovi nameću brojna pitanja poput potencijalnih koristi od dostupnih podataka, načina skladištenja i pristupa podacima. Evolucija zdravstvene skrbi prije svega je evolucija načina razmišljanja: zdravlje treba promatrati kao društvenu i gospodarsku investiciju, pokretač rasta koji stvara blagostanje među onima koji pružaju tehnološku opremu (tvrtke), onima koji koriste u hitnim slučajevima i rutinskoj njezi (bolnice i medicinska profesija), i onima koji imaju koristi od toga (opća javnost).<sup>39</sup> Veliki podaci su iznimno bitni, ali su slabo organizirani, heterogeni i često neuredni.<sup>40</sup> Iako se zbog velike količine ne može inzistirati na potpunoj točnosti i uređenosti svih podataka, Veliki podaci se smatraju jednim od najboljih alata za smanjenje povezanih i funkcionalnih troškova pružatelja zdravstvenih usluga diljem svijeta.<sup>41</sup>

Dugo vremena u zdravstvu je postojalo odupiranje velikim tehnološkim promjenama ali pojavom Velikih podataka je postalo jasno da su primjene novih tehnologija neizbježne. Veliki podaci u zdravstvu imaju značajnu ulogu i korist za zdravstvo i pacijente jer se uz njihovo ispravno korištenje očekuje kvalitetnije i efikasnije zdravstvo za sve. Pohranjivanje i dohvat podataka zahtijevaju svi pružatelji zdravstvenih usluga, a to je moguće samo zahvaljujući efikasnom upotrebom računala i informacijske tehnologije.<sup>42</sup> Izvori podataka u zdravstvu su obično u u kliničkim bilješkama, senzornim i genomskim podacima i sve češće u nosivim uređajima koji mjere biološke pokazatelje korisnika. Prilikom donošenja odluka u liječenju, zdravstveni djelatnici su se često oslanjali na podatke koji su spremljeni u bazama podataka

---

<sup>39</sup> Shan, Keerthi & John, E P. (2022). Adoption of Digital Health Care -A Reality in Future. International Journal of Innovative Research in Education. 8. str. 370-380.

<sup>40</sup> Varga, M. (2014). Upravljanje podacima, 2. izdanje, Zagreb: Element

<sup>41</sup> Hussain, Mohammad Javed. (2020). Big Data in Healthcare. International Journal of Recent Technology and Engineering. 8 (6), str. 2127-2131, <http://www.doi.org/10.35940/ijrte.F8100.038620>

<sup>42</sup> Bakri, M. (2020). Big data healthcare paper, International Journal of Recent Technology and Engineering (IJRTE) ISSN: str. 2277-3878, 8 (6) Preuzeto sa: [https://www.researchgate.net/publication/343971630\\_Big\\_data\\_healthcare\\_paper](https://www.researchgate.net/publication/343971630_Big_data_healthcare_paper) (Datum pristupa 26.08.2022.)

koje nisu bile namijenjene nestrukturiranim podacima. Nestrukturirani podaci su u zdravstvu ali i u svim drugim djelatnostima su standard i u njima se nalazi mnoštvo informacija iz kojih se mora maksimizirati korist kako bi se donijele najbolje odluke.

Laney je 2001. definirao tri osnovne karakteristike posebnih obilježja Velikih podataka i to: volumen, raznovrsnost i promjenljivost.<sup>43</sup> Kasniji razvoj i unapređenje big data uzdravstvenom sustavu doveo je do definiranja šest osnovnih karakteristika Velikih podataka koje su prikazane<sup>44</sup>.

*Tablica 3 Osnovne karakteristike velikih podataka u zdravstvu*

VOLUMEN	BRZINA	RAZNOVRSNOST	ISTINITOST	VRIJEDNOST	POVJERLJIVOST
Količina za manipulaciju	Dinamika prikupljanja i obrade	Oblik, strukturiranost	Pouzdanost izvora	Važnost obzirom na istraživanje	Tajnost, legislativa
Velik	Velika	Velika	Visoka	Visoka	Visoka

Izvor: Izrada autora prema Bilten Hrvatskog društva za medicinsku informatiku

Volumen predstavlja najvažniju i najočitiju značajku Velikih podataka. Autor ovog koncepta je upozorio da rastom volumena podataka proporcionalno pada kvaliteta podataka i da je to odnos koji se mora promatrati.

Brzina opisuje dinamiku stvaranja i obrade podataka. U području zdravstva brzina Big Data je velika, jer su podaci ti koji uvijek moraju biti dostupni.

Raznolikost podataka je bitna karakteristika Velikih podataka. Postoje tri različite vrste podataka koji se koriste u sustavima Velikih podataka, a to su: strukturirani, polustrukturirani i nestrukturirani podaci.

Istinitost podataka označava kvalitetu i podrijetlo podataka i izvora. Ovo je značajka koju uvijek treba propitivati jer može doći do velikih troškova i problema ukoliko se naknadno primjete nelogičnosti i neistinitosti u podacima.

<sup>43</sup> Laney, D., 2001. 3D data management: Controlling data volume, velocity and variety. META group research note, 6(70), p.1.

<sup>44</sup> Somek, M. (2021) Što skrivaju masovni podaci u zdravstvu? Bilten Hrvatskog društva za medicinsku informatiku, 27 (2), 23-30.

Vrijednost podataka usko je povezana s procesima analize i predstavlja dodatnu vrijednost koju se može dobiti iz poznatih podataka.

Povjerljivost podataka je iznimno važna značajka zbog prirode zdravstvenih podataka. Podaci pacijenata su moraju ostati tajni i podložni su pravnoj regulativi .

Količina podataka, koja se svakodnevno koristi u zdravstvu je vrlo velika, te je stoga potrebno podatke pohranjivati i stvarati sigurnosne kopije koje će spriječiti gubitak podataka. Za pravilno pohranjivanje zdravstvenih podataka potrebno je planirati samu pohranu zbog upravljanja podacima o povijesti bolesti pacijenata.

Vrlo je važno poticati modernizaciju i unaprijeđenje cjelokupnog zdravstvenog sustava jer samo uvođenje sustava Velikih podataka u zdravstvene sustave i organizacije unaprijedit će i sam proces liječenja. Razvijanje tehnologija koje su proizašle iz Velikih podataka i njihova usklađenost s postojećim protokolima i standardima je ključna za ukupni napredak zdravstva. Iako profit nije i ne bi trebao biti primarni razlog dobrog poslovanja, od ključne je važnosti za zdravstvene organizacije nabaviti dostupne alate, infrastrukturu i tehnike za učinkovito korištenje Velikih podataka kako bi poslovali s manjim troškovima. <sup>45</sup>

## **4.2. Jezera podataka u zdravstvu**

Zdravstvo se sve više oslanja na tehnologiju pa tako danas postoje organizacije koje su već uvele ili imaju u planu uvesti jezera podataka u poslovanje. Prelazak zdravstvenih organizacija s klasičnih oblika skladištenja na jezera podataka mogao bi riješiti probleme pogrešnih dijagnoza kao i predviđanja tijekova bolesti.

Jezero podataka može se učinkovito koristiti u različitim domenama zdravstvene skrbi kao što su pružanje skrbi za pacijente, klinička istraživanja, upravljanje zdravljem stanovništva i sigurnosna analitika. To osigurava prednosti kao što su bolji klinički ishodi i protokoli, smanjeni troškovi, brže i točnije kliničke istraživačke studije i poboljšana sigurnost podataka. Pored toga, jezera podataka mogu imati veliku ulogu i u administrativnim i računovodstvenim

---

<sup>45</sup> Raghupathi, V., Raghupathi, W. (2014). Big data analytics in healthcare: Promise and potential, Health Information Science and Systems, 2 (3) <https://doi.org/10.1186/2047-2501-2-3>



poslovima, procesima nabave i brige o popunjenosti smještajnih kapaciteta ali i u istraživanju i razvoju novih načina liječenja.<sup>46</sup>

Kroz podatkovnu znanost i predviđanje analitike, zdravstvene organizacije mogu prepoznati potencijalne ishode i istraživati scenarije koje im se čine vjerojatnim. S takvim uvidima usmjerenim na budućnost, pružatelji usluga u zdravstvu mogu unaprijediti inicijative odgovorne zdravstvene skrbi, stvaranje novog područja znanosti o podacima za otkrivanje trendova i otkrića koja utječu na integriranu skrb o pacijentu. Jezera podataka također otvaraju mogućnosti za integraciju informacija iz nosivih uređaja baziranih na internetu stvari poput pametnih satova, uređaja za praćenje rada pacemakera, pametnih olovaka za praćenje glukoze u krvi itd.

Zdravstvene organizacije iz ovakvih tehnologija mogu imati višestruke koristi poput smanjenja bolničkih liječenja, personalizirane medicine i poboljšanih zdravstvenih ishoda. U kliničkom okruženju bitno je da liječnički tim može brzo i jednostavno pristupiti svim relevantnim informacijama o pacijentu. Međutim, interakcija s digitalnim zdravstvenim sustavima i jezerom podataka često zna predstavljati prepreke za liječnike i zbog toga je jako bitno osigurati pristup svim informacijama o pacijentu ali i jednostavno sučelje.<sup>47</sup>

Najveći dio medicinskih podataka je nestrukturiran i kao takvi često nisu bili predmet analiza već su prolazili ispod radara u skladištima podataka ostavljajući iza sebe neiskorištene mogućnosti manipuliranja podacima.<sup>48</sup> S jezerom podataka se to mijenja jer svi podaci mogu ući u jezero bez da se odmah rade analize, pretraživanja i modeli strojnog učenja. Usporedbe analiza podataka iz klasičnih skladišta podataka sa strukturiranim podacima i svih vrsta podataka iz jezera podataka pokazala je veću vjerodostojnost jezera podataka s obzirom na obuhvat većeg i raznovrsnijeg podatkovnog materijala.

---

<sup>46</sup> Sigma Data Systems (2020), What is data lake: Storage and data processing, Dostupno na: <https://www.sigmadatasy.com/data-lake-essentials-storage-data-processing/>, Pristupano 28.08.2022.

<sup>47</sup> AWS, Derive AI/ML-driven insights from healthcare data using Amazon HealthLake, Dostupno na: <https://aws.amazon.com/blogs/industries/derive-ai-ml-driven-insights-from-healthcare-data-using-amazon-healthlake> (27.08.2022.)

<sup>48</sup> Dinesh K. G., Arumugaraj K. , Santhosh K.D., Mareeswari V., (2018.), Prediction of Cardiovascular Disease Using Machine Learning Algorithms, International Conference on Current Trends towards Converging Technologies (ICCTCT), 2018, pp. 1-7, <https://doi.org/10.1109/ICCTCT.2018.8550857>

### 4.3. Slučajevi korištenja jezera podataka u zdravstvu

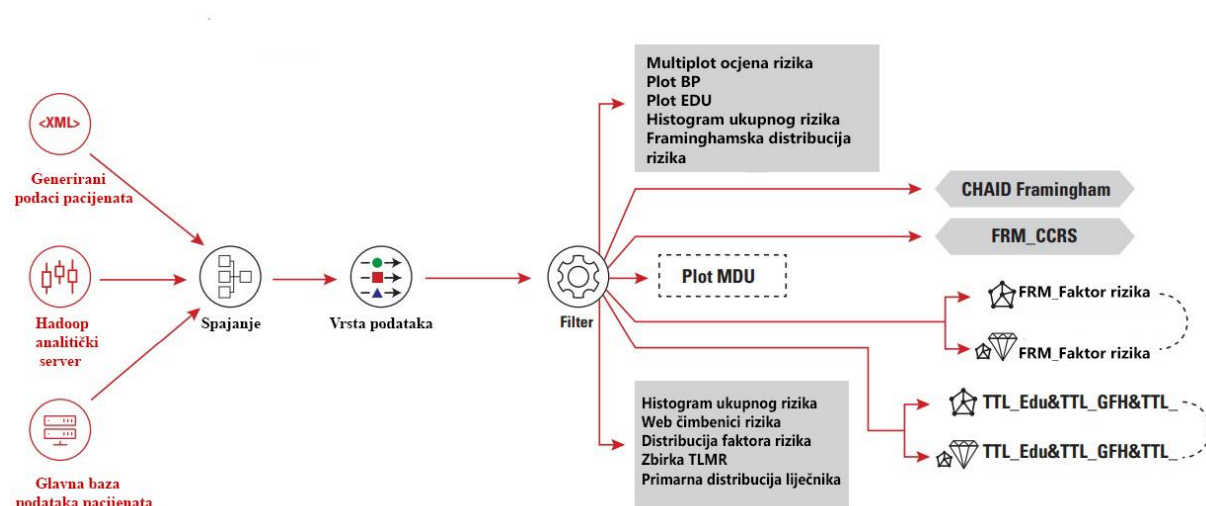
Najčešći upiti zdravstvenih organizacija prema poduzećima koja implementiraju digitalne tehnologije u velike sustave su vezani uz isplativost jezera podataka u odnosu na klasična skladišta podataka. Jezero podataka je primjereno zdravstvenim organizacijama najviše zbog znatne količine podataka koja dolazi iz više izvora u različitim oblicima.

Važno je istaknuti kako kvaliteta jezera podataka ne umanjuju veliki značaj skladišta podataka u svakodnevnim poslovnim odlukama i kako se najbolji rezultati dobiju uz zajedničku suradnju.

U ovom poglavlju će biti prikazani neki od slučajeva korištenja jezera podataka u zdravstvu. Prvi slučaj korištenja tiče se Framinghamske studije srca i prediktora osobnih zdravstvenih kartona. Framinghamska studija je višegeneracijska studija koja je osmišljena za prepoznavanje čimbenika koji doprinose razvoju kardiovaskularnih bolesti i iz nje je proizašao termin „faktor rizika“. <sup>49</sup> Pokazuje na koji način su srčane bolesti povezane s dvije grupe čimbenika. Prva se odnosi na krvni tlak, kolesterol, pušenje i dob a u drugoj grupi su čimbenici poput životnih navika, okolina i genetika.

Slika 9 Model rizika za kardiovaskularne bolesti pacijenta

Što



<sup>49</sup> Br Gallestey, J. Bacallao (2020). Framingham Heart Study, Dostupno na: <https://www.britannica.com/event/Framingham-Heart-Studayitannica.com/>, Pristupano 28.08.2022.

Izvor: Obrada autora prema Perficient

Na slici se može vidjeti kako se koristi jezero podataka zajedno sa skladištem podataka. Jezero podataka tvore podaci pacijenata odabranih za analizu, dok je u skladištu podataka glavna baza podataka svih pacijenata. Ovi podaci su spojeni za formiranje učinkovitog modela rizika kardiovaskularnih bolesti. Podaci o pacijentima iz jezera podataka su izvezeni iz osobnih zdravstvenih kartona u XML formatu i čine oslonac modela. Obrada spaja podatke odabranih pacijenata u XML formatu sa strukturiranim podacima iz skladišta podataka kako bi izračunala ocjene rizika i odredila relevantne faktore. Nakon toga se kreiraju izvješća s ocjenama stanja srca i uspoređuju s novim modelima rizika koristeći širi spektar varijabli.

Kada je analiza gotova i pregledana od strane liječnika, podaci se mogu objaviti na više načina, npr. u aplikaciji ili u elektroničkom zdravstvenom kartonu zajedno s preventivnim mjerama koje je potrebno poduzeti prije otpušanja pacijenata iz bolnice. Osnovni elementi koji su proučavani u studiji (kolesterol, pušenje, krvni tlak, spol) daju odgovore o stanju kardiovaskularnog sustava. Na osnovu primarne analize mogu se raditi i detaljnije analize pomoću strojnog učenja i umjetne inteligencije koje bi dale detaljniji uvid i otkrile glavne čimbenike rizika za zdravlje.

U drugom slučaju korištenja, radi se o učinkovitom korištenju smještajnih jedinica u bolničkim sustavima.<sup>50</sup> Primjetan je trud da se prosječno vrijeme boravka pacijenta u bolnici smanji a da se pritom ne smanji zadovoljstvo pacijenata tijekom boravka u bolnici. Identificirajući pacijente s rizikom za produženi boravak, bolnica može poduzeti proaktivne mjere za izradu plana liječenja kako bi se smanjila očekivana duljina boravka. Bolnicama je zbog velikih troškova u cilju predvidjeti prosječnu duljinu boravka svakog pacijenta u trenutku prijema i dati te informacije medicinskoj sestri ili osoblju koje prima bolnicu u svrhu raspodjele resursa. Modeli mogu biti izrađeni samo na osnovu podataka koje su svakom pacijentu dostupne u trenutku prijema. Podaci dobiveni od pacijenata na prijemu se mogu analitičkim metodama usporediti s arhivskim podacima pacijenata sa sličnim stanjima i na taj način predvidjeti prosječnu duljinu boravka u bolnicama. S tim podacima, ustanove imaju lakše upravljanje zauzetošću krevetima i smanjene troškove.

---

<sup>50</sup> Diving in: Navigating a data lake for predictive care, EMC, 2015. Dostupno na: <https://www.dell.com/community/s/vjauj58549/attachments/vjauj58549/solutions-ch/673/1/h14417-emc-navigating-data-lake-predictive-care.pdf> (28.06.2022.)

U trećem slučaju korištenja, radi se o pronalaženju izvora problema za organizaciju i pacijenta pomoću tehnologija u jezeru podataka. Za ovaj slučaj korištenja, pružatelji usluga su ispunili jezero podataka podacima iz višestrukih izvora poput laboratorija, elektroničkih zdravstvenih zapisa, nosivih uređaja i izvješća. Naime, u bolničkim ustanovama se često zna dogoditi razvijanje dodatnih stanja bolesti nakon što je pacijent već u procesu liječenja. Takva stanja dovode do velikih organizacijskih problema po zdravstvene djelatnike i upravu i naravno velikog rizika za pacijenta. Koristeći analitiku jezera podataka, zdravstvene organizacije mogu pronaći izvore problema i riješiti ih. Najčešće se radi o bakterijskim infekcijama tj. posljedičnim trovanjima krvi za vrijeme liječenja. Koristeći logove i podatke iz kliničkih bilješki i liječničkih posjeta, rezultati analize mogu ukazati na anomalije kod pacijenata za vrijeme radnog vremena određenih liječnika. Takve se liječnike mora upozoriti i provesti edukacija o higijenskim procesima za vrijeme liječenja kako bi se takvi, često smrtonosni, propusti iskorijenili.

#### **4.4. Umjetna inteligencija i strojno učenje u zdravstvu**

Tehnologije umjetne inteligencije koje postaju sve prisutne u suvremenom poslovanju i svakodnevnom životu također se kontinuirano primjenjuju u zdravstvu. Iako je umjetna inteligencija mlada znanost, naslijedila je mnoge pristupe i tehnike iz drugih znanosti koje joj danas daju prednost u širokoj upotrebi.<sup>51</sup> Korištenje umjetne inteligencije u zdravstvu ima potencijal pomoći pružateljima zdravstvenih usluga u mnogim načinima skrbi za pacijente i administrativnim procesima, pomažući im da poboljšaju postojeća rješenja i brže prevladaju izazove. Većina umjetne inteligencije i zdravstvenih tehnologija ima veliku važnost za područje zdravstva, ali načini primjene se mogu značajno razlikovati između bolnica i drugih zdravstvenih organizacija

Aplikacije strojnog učenja imaju značajan utjecaj na zdravstvo. Strojno učenje je podvrsta tehnologije umjetne inteligencije koja ima za cilj poboljšati brzinu i točnost rada liječnika. Zemlje u cijelom svijetu se trenutno suočavaju s preopterećenim zdravstvenim sustavom s nedostatkom kvalificiranih liječnika, pri čemu umjetna inteligencija daje veliku nadu. Tehnike temeljene na strojnom učenju pomažu u otkrivanju ranih pokazatelja epidemije ili pandemije. Ovaj algoritam ispituje satelitske podatke, vijesti i izvješća društvenih medija, pa čak i video

---

<sup>51</sup> Prister, V. (2019). Umjetna inteligencija, Media, culture and public relations, 10(1), str. 67-72. <https://doi.org/10.32914/mcpr.10.1.7>

izvore kako bi utvrdio hoće li bolest izmaći kontroli. Korištenje strojnog učenja može produljiti vrijeme pružatelja zdravstvenih usluga da se usredotoče na brigu o pacijentima umjesto na pretraživanje ili unos informacija.<sup>52</sup>

Razvoj algoritama i aplikacija umjetne inteligencije u kombinaciji s lakim pristupom podacima omogućuje primjenu umjetne inteligencije u zdravstvu. Ovi se algoritmi koriste u radiološkim metodama snimanja organa koje se koriste za otkrivanje karcinoma. Uz CT probir, odnosno kompjutoriziranu tomografiju, umjetna inteligencija koristi se u radu magnetskih skenera, PET skenera, kao i drugih medicinskih uređaja koji koriste slične algoritme rada.<sup>53</sup> Provedena su mnoga istraživanja koja su potvrdila važnost umjetne inteligencije u zdravstvu. Jedno od takvih istraživanja je istraživanje koje je potvrdilo da je umjetna inteligencija dijagnosticirala rak dojke s 90% točnosti u usporedbi sa stručnjacima radiolozima. Osim toga, bezbroj je slučajeva gdje umjetna inteligencija, odnosno CT pregledi, PET skeneri i sl. uspjeli dijagnosticirati rak u ranoj fazi razvoja, odnosno u fazi razvoja koja se ne može ustanoviti standardnim pregledom kod radiologa, bez umjetne inteligencije. Danas, u modernom svijetu, umjetna inteligencija radi revoluciju kada je u pitanju detekcija stanica raka i drugih bolesti. To znači da umjetna inteligencija postupno, ali sigurno preuzima jednu od najvažnijih uloga kada je u pitanju dijagnosticiranje teških bolesti. Sama inteligencija omogućuje da se pregled obavi pod puno većom i jasnijom strukturom obrade svih dobivenih podataka nego što to može učiniti bilo koji liječnik, bez obzira na razinu znanja i informiranosti.

Strojno učenje predstavlja granu umjetne inteligencije koja se temelji na ideji da sustavi mogu učiti iz podataka bez prisutnosti ljudi kao važnog faktora. Pogodan je za primjenu u situacijama kada se rješava problem koji se vrlo lako može formulirati i savladati, ali se ne može eksplicitno definirati. Da bi strojno učenje bilo moguće, potrebno je prethodno pripremiti i prilagoditi podatke, zatim pronaći najbolji način isporuke podataka, odnosno naći metodu koja će podatke dostaviti. Strojno učenje se koristi se u različitim kliničkim domenama s analizom stotina kliničkih parametara u djelotvornim i učinkovitim modelima za poboljšanje rezultata i kvalitete modela medicinske skrbi.<sup>54</sup>

---

<sup>52</sup> Coh T. & Brody, J. (2021). Applications of Machine Learning in Healthcare, DOI: 10.5772/intechopen.92297

<sup>53</sup> Franjić, D., & Miljko, M. (2020). Umjetna inteligencija u radiologiji: Etički problemi. *Zdravstveni glasnik*, 6, 61-68.

<sup>54</sup> Lopez, Fernando & Núñez Valdez, E., García Díaz, V., & Bursac, Z. (2020). A Case Study for a Big Data and Machine Learning Platform to Improve Medical Decision Support in Population Health Management. *Algorithms* 13 (4), 102 <https://doi.org/10.3390/a13040102>

Metode strojnog učenja mogu se kategorizirati u tri vrste

1. Nadzirano učenje je metoda u kojoj se daju podaci na temelju kojih se uči, ali se za dane podatke dobivaju i željeni rezultati. Ova metoda strojnog učenja mora naučiti odrediti odgovarajuće izlaze za dane podatke.

2. Nenadzirano učenje je metoda u kojoj se podaci dostavljeni za strojno učenje ne podudaraju s odgovarajućim rezultatima. Ova metoda strojnog učenja potrebna je za prepoznavanje uzoraka u danim podacima.

3. Učenje pojačanja je metoda koja potiče proces strojnog učenja koristeći princip nagrade i kazne. Učenje pojačanja u zdravstvu ima najčešću primjenu za pacijente s dugotrajnijim stanjima.

Neki od najčešćih primjera strojnog učenja u zdravstvu su sustavi za podršku u kliničkom odlučivanju, personalizirana medicina, strojno učenje u medicinskim slikama i prediktivni pristup liječenju.

#### **4.5. Zaštita podataka**

Iako usvajanje tehnologija Velikih podataka u zdravstvu nosi mnoge prednosti i očekivanja, ono također postavlja neke prepreke i izazove. Tajnost i privatnost zdravstvenih podataka je standard koji treba biti osiguran u svakom trenutku. Ipak, zabrinutost oko sigurnosti osjetljivih informacija i privatnosti u zdravstvu se povećava iz godine u godinu zbog nekoliko rastućih trendova u zdravstvu, kao što su mobilnost liječnika i bežično umrežavanje, razmjena zdravstvenih informacija, računalstvo u oblaku itd. Podaci koji se prikupljaju u zdravstvu se mogu vrlo lako prodati kao roba, koja se može zloupotrijebiti na mnogo načina.

Zbog velike lakoće prikupljanja podataka, jedno od najvažnijih pitanja u suvremenom svijetu je zaštita podataka. Korištenje umjetne inteligencije, čak i sofisticiranog softvera uz sve svoje koristi povećava mogućnost krađe podataka. Kao rezultat napretka tehnologije i brze proizvodnje podataka, ključno je da organizacije implementiraju rješenja za sigurnost zdravstvenih podataka, a istovremeno zadovoljiti zahtjeve za usklađenošću zdravstvene zaštite.

Postoji više tehnologija koje se koriste u zaštiti zdravstvenih podataka a jedna od njih je autentifikacija. Autentifikacija služi zaštiti identiteta korisnika i osigurava autentičnost korisnika pomoću kriptografskih protokola. Najčešća vrsta napada na zdravstvene sustave u

cilju krađe podataka je „čovjek u sredini“ (*eng.* man in the middle), koji funkcionira na način da haker zloupotrebljava slabosti sustava tako što postavlja zlonamjerni softver koji uspješno prolazi kroz sigurnosni sustav i presreće zdravstvene podatke pacijenata.<sup>55</sup>

Jedan od načina zaštite zdravstvenih podataka je i enkripcija podataka. Zdravstvene organizacije ili pružatelji usluga moraju osigurati da je shema šifriranja učinkovita, laka za korištenje i od strane pacijenata i zdravstvenih radnika, te da se lako može proširiti za uključivanje novih elektroničkih zdravstvenih zapisa. Nadalje, broj ključeva koje drži svaka strana trebao bi biti minimiziran a vanjski suradnici zaduženi za implementaciju novih tehnoloških rješenja bi trebali biti pod posebnim nadzorom.

Maskiranje podataka zamjenjuje osjetljive podatke s vrijednostima koje ne mogu biti identificirane i često se koristi u deidentificiraju podataka pacijenata. Značajna prednost ove tehnike je smanjenje troškova osiguravanja implementacije Velikih podataka. Kako se sigurni podaci migriraju iz sigurnog izvora u platformu, maskiranje smanjuje potrebu za primjenom dodatnih sigurnosnih kontrola na te podatke dok se nalaze na platformi.

Važna mjera u zaštiti podataka je kontrola pristupa podacima. Svaki pacijent ima pravo uvida u svu medicinsku dokumentaciju koja se tiče njegova zdravstvenog stanja. Pravo na pristup, odnosno uvid pacijenta u svoju medicinsku dokumentaciju odnosi se, kako na dokumentaciju koja se uručuje pacijentu nakon završenog liječničkog pregleda, tako i na njegovu cjelokupnu medicinsku dokumentaciju.<sup>56</sup>

Problem nastaje kada pacijentovim podacima pokuša pristupiti netko drugi i upravo zato su kontrole pristupa zdravstvenim podacima izuzetno stroge. Prilikom pristupa informacijskom sustavu, postavljen je sustav autentifikacije. Nakon provjere autentičnosti, korisnici mogu ući u informacijski sustav, ali će njihov pristup i dalje biti reguliran politikom kontrole pristupa koja se obično temelji na privilegiji i pravu svakog praktičara kojeg ovlasti pacijent ili treća strana od povjerenja. Tako je osiguran snažan i fleksibilan mehanizam za dodjelu dopuštenja korisnicima. Pruža sofisticirane kontrole autorizacije kako bi se osiguralo da korisnici mogu obavljati samo one aktivnosti za koje imaju dopuštenja, kao što je pristup podacima, predaja

---

<sup>55</sup> Spremić, M. (2017). Sigurnost i revizija informacijskih sustava u okruženju digitalne ekonomije. Zagreb:Ekonomski fakultet Zagreb. str.50

<sup>56</sup> Čizmić J., Pravo na pristup podacima u medicinskoj dokumentaciji (1991) Zbornik pravnog fakulteta Sveučilišta u Rijeci 30(1), str. 93

poslova, administracija klastera, itd. Predložen je niz rješenja za rješavanje pitanja sigurnosti i kontrole pristupa a neke od primjera možemo vidjeti i u hrvatskom zdravstvenom sustavu.

Slika 10 Autentifikacija HZZO pametnom karticom



Izvor: HZZO

U sustavu e-građani, postoji više sigurnosnih razina putem kojih korisnici mogu pristupiti svojim podacima ili podacima pacijenata. U zdravstvu je već duže vrijeme prisutna pametna kartica koja pruža sigurnost i pouzdanost zdravstvenih podataka zbog visoke razine zaštite.

U svim sustavima, pa tako i sustavima Velikih podataka, bitno je da postoji transparentnost u poslovanju i regulativama koje se odnose na zaštitu podataka. Bitno je da pacijenti uvijek imaju informacije kako se njihovi podaci koriste i tko sve ima pristup podacima.

Sva poduzeća i organizacije moraju se držati zahtjeva iz Opće uredbe o zaštiti podataka u kojoj je navedeno da u slučaju da u slučaju situacije da osobni podaci budu zabilježeni u sustavu a nisu dobiveni od pacijenta moraju javno pružiti informacije o administratoru i voditelju obrade podataka, o razlozima i svrsi zašto su ti podaci uzeti i čemu namijenjeni.<sup>57</sup>

<sup>57</sup> Gdpr Text, Članak 14. GDPR, Dostupno na: <https://gdpr-text.com/hr/read/article-14/>, Pristupano 15.09.2022.



## 5. STUDIJA SLUČAJA AMAZON HEALTHLAKE PLATFORME

### 5.1 Opis Amazon HealthLake platforme

Zdravstvene organizacije svaki dan proizvode i dobivaju golemu količinu podataka o svojim pacijentima i na osnovu tih podataka pokušavaju stvoriti potpunu sliku o zdravstvenom stanju pacijenata. Problem za zdravstvene organizacije nastaje zbog nestandardiziranih formata u kojima se ti podaci spremaju i prenose. Većina medicinskih podataka su nestrukturirane ili polustrukturirane informacije i potrebno je dugo vremena za prikupljanje podataka a potom i transformaciju i analizu. Ključno pitanje je kako napraviti smislenu cjelinu od svih podataka koji su dostavljeni od zdravstvene organizacije na što jeftiniji način i vremenski prihvatljiv okvir.

Slika 11 Različiti izvori podataka koji ulaze u jezero podataka



Izvor: Obrada autora prema AWS Events

Amazon HealthLake je platforma koja omogućuje pružateljima zdravstvenih usluga, zdravstveno osiguravajućim društvima i farmaceutskim poduzećima pohranu, transformaciju,

upite i analizu zdravstvenih podataka.<sup>58</sup> Platforma je izgrađena u sklopu poduzeća Amazon Web Services, koje je kao svoju podružnicu 2006. godine osnovala najveća svjetska online trgovina na malo, Amazon. Platforma se trenutno koristi samo u Sjedinjenim Američkim Državama i podlaže Zakonu o prenošenju i odgovornosti zdravstvenog osiguranja (engl. HIPAA) koje pruža zakonodavni okvir za zaštitu i privatnost medicinskih podataka.

Uzimajući u obzir znanja o problemima s kojima se zdravstvene organizacije susreću, Amazon je kreirao platformu koja je bila zamišljena kao spremište kojim se može upravljati i vršiti razne manipulacije podacima kroz moderne tehnologije. Koristeći HealthLake programsko sučelje, zdravstvene organizacije imaju mogućnost na jednostavnije načine kopirati zdravstvene podatke koji mogu biti medicinski izvještaji ili personalizirane bilješke pacijenata. To čine tako da podatke iz lokalnih sustava premještaju u jezero podataka u oblaku, koje smatraju najsigurnijim okruženjem.

Podaci koji su premješteni u jezero podataka su sirovi i neobrađeni i najčešće predstavljaju uspostavljene dijagnoze, propisane lijekove i bilješke o postupcima u liječenju u različitim oblicima (PDF, snimljeni razgovori, podaci vremenskih serija itd.) Pomoću višestrukih modela strojnog učenja platforma uspijeva prepoznati i izdvojiti bitne informacije iz medicinske dokumentacije na način da se kasnije mogu pretraživati bez poteškoća. Takve informacije i podatke Healthlake organizira i indeksira prije nego ih pohrani po FHIR standardu koji pobliže opisuje podatkovne formate i elemente.<sup>59</sup>

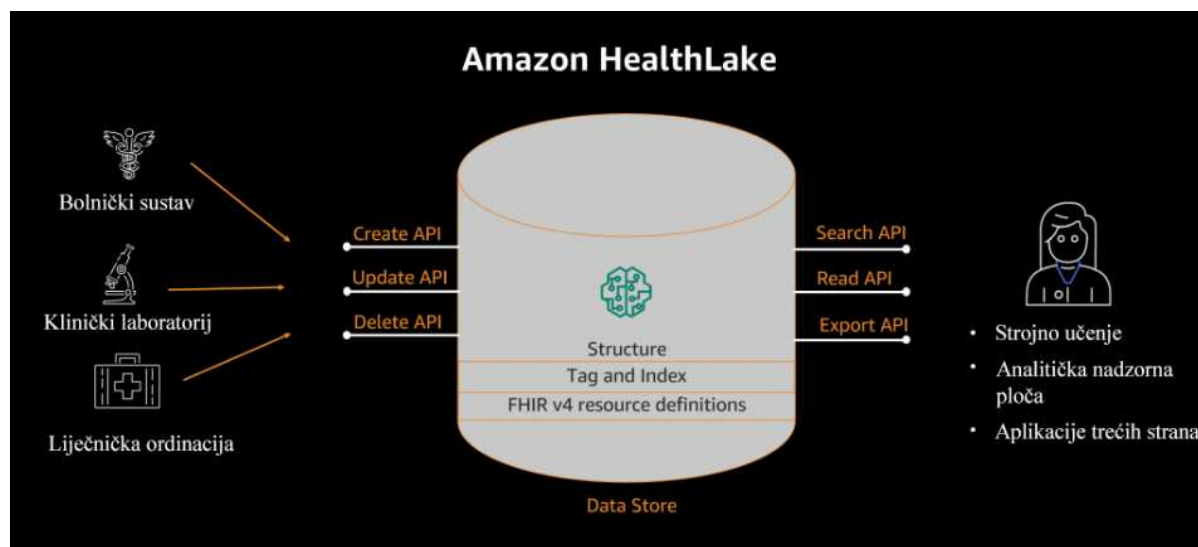
Indeksiranje je proces skeniranja i izrada popisa riječi za pretraživanje, kao i uvjet brzog i učinkovitog pretraživanja. FHIR format pripada HL7 standardu za izmjenjivanje informacija unutar zdravstvenog sustava. Nakon što su svi početni procesi uspješno odrađeni, podaci o pacijentu i svi povijesni podaci su na jednom mjestu i mogu se lako pronaći, nadopuniti i analizirati.

---

<sup>58</sup> AWS (2020), Introducing Amazon HealthLake to make sense of health data, Dostupno na: <https://aws.amazon.com/about-aws/whats-new/2020/12/introducing-amazon-healthlake-to-make-sense-of-health-data/>, Pristupano 26.08.2022.

<sup>59</sup> IBM (2021.), Healthcare: FHIR Transformation pattern, Dostupno na: <https://www.ibm.com/docs/en/integration-bus/10.0?topic=pack-healthcare-fhir-transformation-pattern>, Pristupano 26.08.2022.

Slika 12 Prikaz sučelja Amazon HealthLake platforme



Izvor: Obrada autora prema AWS

## 5.2 Prikaz uloga alata i popratnih tehnologija u Amazon Healthlake platformi

Platforma koristi posebne modele strojnog učenja poput obrade prirodnog jezika (engl. NLP) kako bi se nestrukturirani podaci mogli automatski transformirati. Obrada prirodnog jezika predstavlja područje u umjetnoj inteligenciji u kojoj računala analiziraju, razumiju i izvlače značenja podataka iz ljudskog jezika.<sup>60</sup> Zadatak obrade prirodnog jezika je da detektira entitete poput zdravstvenog stanja i lijekova te da uspostavi vezu između tih entiteta kako bi sustav na kraju mogao npr. ponuditi točnu dozu lijekova za određenog pacijenta ili izraditi račun.

Jednom kada se podaci uvezu, nalaze se u formatu u kojem se mogu postavljati upiti i izvlačiti smislenu cjelinu iz podataka koji su bili u neupotrebljivom formatu za pretraživanje i kreiranje upita. Kada se radi izvlačenje zaključaka iz tih podataka, platforma daje postotak pouzdanosti u zaključke pa korisnik može na osnovu toga odlučiti hoće li vjerovati modelu strojnog učenja.

Kada svi podaci budu spojeni, HealthLake platforma indeksira podatke kako bi se mogli raditi upiti i pretraživati odgovori. Pretrage se mogu raditi pomoću unosa posebnih zdravstvenih stanja ili kodova. Sa svim podacima na jednom mjestu, mogu se raditi vježbe modela strojnog učenja i kreirati upravljačke ploče za poslovnu inteligenciju. Platforma radi pomoću više alata

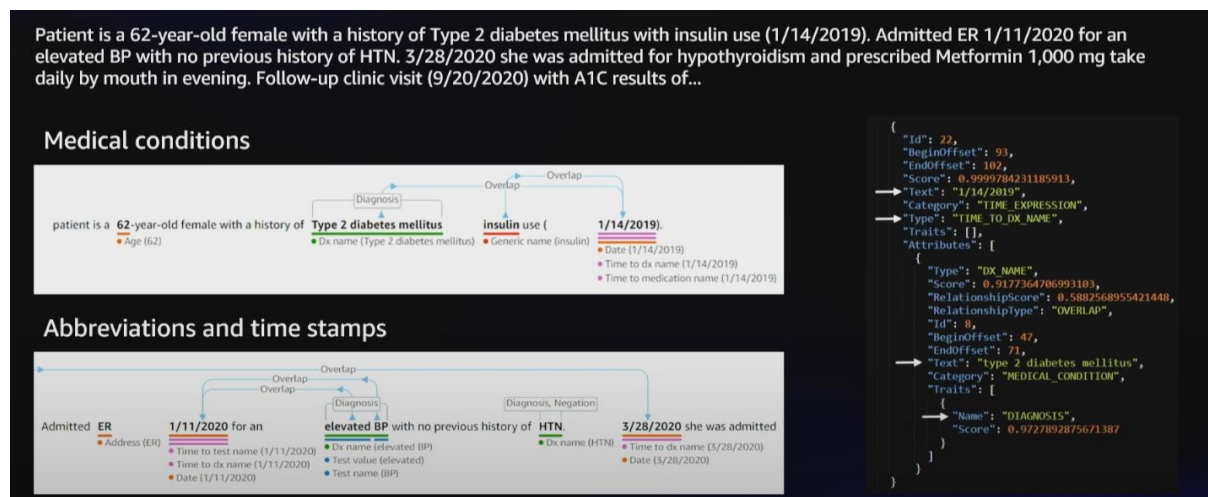
<sup>60</sup> Data Robot (2016), What is Natural Language Processing? Introduction to NLP, Dostupno na: <https://www.datarobot.com/blog/what-is-natural-language-processing-introduction-to-nlp/>, Pristupano 26.08.2022.

a neki od bitnijih su Amazon SageMaker koji kreira modele strojnog učenja dok QuickSight analizira trendove, objašnjava odnose između entiteta i radi predviđanja pomoću zdravstvenih podataka.<sup>61</sup> Zbog standardiziranja podataka u FHIR format, svi ovi podaci se mogu dijeliti među pacijentima i zdravstvenim organizacijama i djelatnicima.

Na sljedećoj slici je primjer kako se izvlači značenje iz nestrukturiranih podataka. Navedeni su sljedeći podaci o pacijentici:

Pacijent je 62 - godišnja ženska osoba sa šećernom bolesti i koristi inzulin od 14.01.2019. Primljena na hitni prijem 11.01.2019. zbog povišenog krvnog tlaka. Nema povijest hipertenzije. 28.03.2020. je primljena zbog hipotireoze i propisan joj je Metformin 1000mg koji se uzima svaki dan navečer.

Slika 13 Izvlačenje značenja iz nestrukturiranih zdravstvenih podataka iz kliničke bilješke



Izvor: AWS

Na slici je prikazan primjer što obrada prirodnog jezika zajedno s HealthLake platformom radi s podacima iz ovih bilješki. Prvo se prepoznaje dijagnoza koja je u ovom slučaju dijabetes tipa 2, zatim se ustanovljuje datum kada je dijagnoza postavljena i dodjeljuje se kod iz ICD- 10 međunarodne klasifikacije bolesti (*engl. International Classification of Diseases*). Obrada prirodnih jezika prepoznaje i negaciju odnosno nepostojanje prethodnih problema s hipertenzijom. Sve se pretvara u FHIR format i na samom kraju liste kodova s desne strane se može vidjeti rezultat pouzdanosti prikazan u relativnom obliku. Kod pretraživanja lijekova i

<sup>61</sup> AWS, Amazon QuickSight – The most popular cloud-native, serverless BI service, Dostupno na: <https://aws.amazon.com/quicksight/>, Prisupano 27.08.2022.

doze koja je dodijeljena pacijentu, koriste se kodovi svih dostupnih lijekova putem kojih se izvršavaju upiti.

Slika 14 Dodijeljivanje medicinskih kodova tekstu

The screenshot displays two panels: RxNorm on the left and ICD-10 on the right. The RxNorm panel shows two columns of medication suggestions for the text 'METFORMIN 1,000 mg tablet E 1 AND 1/2 TABLETS BY MOUTH IN THE MORNING...'. The top suggestion in each column is highlighted with a red box. The ICD-10 panel shows a list of diabetes-related codes for the text 'Type 2 diabetes mellitus with hyperglycemia, without long-term current use of insulin (HCC)...', with the top suggestion also highlighted in red.

RxNorm		ICD-10	
861004	Metformin hydrochloride 1000 MG Oral Tablet 0.9904 score	E11.65	Type 2 diabetes mellitus with hyperglycemia Score: 0.7905
1243020	Linagliptin 2.5 MG / Metformin hydrochloride 1000 MG Oral Tablet 0.8243 score	E11.9	Type 2 diabetes mellitus without complications Score: 0.7658
1807894	Osmotic 24 HR Metformin hydrochloride 1000 MG Extended Release Oral Tablet 0.7614 score	E11.22	Type 2 diabetes mellitus with diabetic chronic kidney disease Score: 0.7353
1807888	Modified 24 HR Metformin hydrochloride 1000 MG Extended Release Oral Tablet 0.7532 score	E11.42	Type 2 diabetes mellitus with diabetic polyneuropathy Score: 0.7144
861007	Metformin hydrochloride 500 MG Oral Tablet 0.7297 score	E11.621	Type 2 diabetes mellitus with foot ulcer Score: 0.6959

Izvor: AWS

Jedna od bitnijih odrednica HealthLake platforme je i interoperabilnost pa se tako podaci iz bolnica, laboratorija, farmaceutskih poduzeća pretvoreni u FHIR format mogu naći na jednom mjestu i imati detaljan pregled.

Slika 15 Interoperabilnost unutar zdravstvene djelatnosti

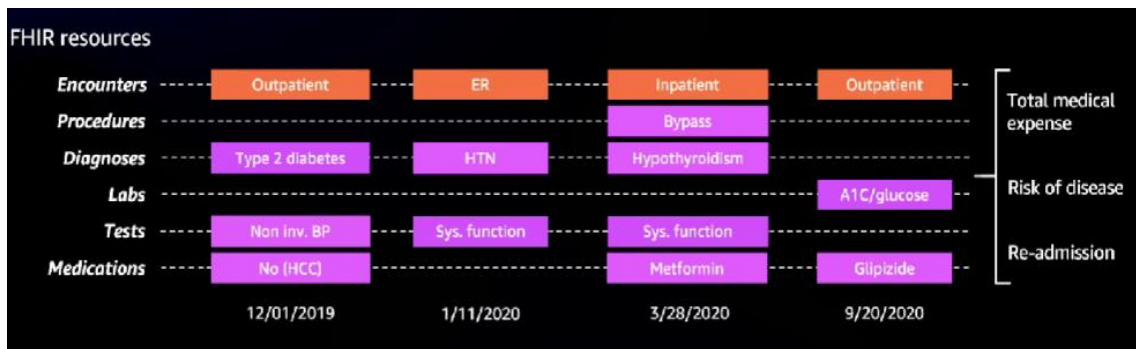


Izvor: Obrada autora prema AWS

Zahvaljujući suradnji i dijeljenju podataka između dionika, lako je napraviti kronološki pregled zdravstvenog stanja pacijenta. Uz dobar i potpun kronološki pregled podataka o pacijentima koji su pretvoreni u novi format mogu se donositi i kvalitetnije odluke u liječenju. Primjerice, ako se za primjer uzme pacijentica s šećernom bolesti može se vidjeti kako HealthLake ima mogućnost kreiranja holističkog i kronološkog pregleda stanja pacijenta pomoću podataka

pretvorenih u FHIR format. Nakon što se svi podaci kronološki poredaju uz pomoć obrade prirodnog jezika, mogu se raditi predviđanja poput vremenskog trajanja oporavka pacijenta u bolnici, visine troškova u liječenju, rizika novih bolesti ili ponovnog prijema u bolnicu. Zdravstvenim organizacijama i ustanovama je jako bitno imati što točnije modele jer su liječenja jako skupa i pomoću tehnologija i informacija moraju smanjivati troškove.

Slika 16 Kronološki pregled događaja tijekom liječenja

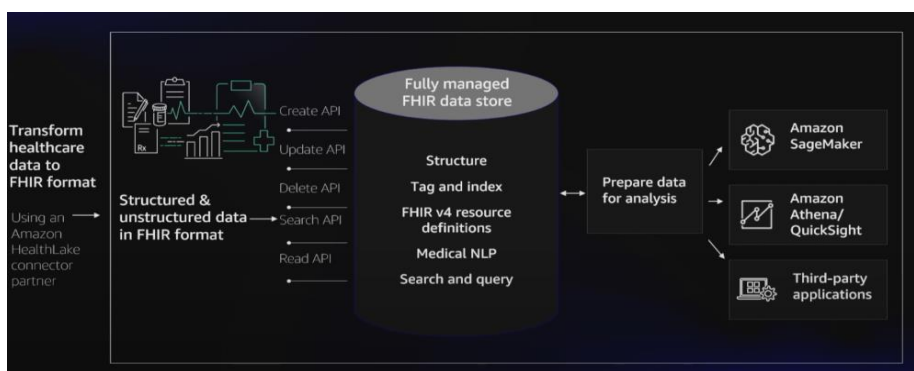


Izvor: AWS

Nakon što su potrebni podaci odabrani i za potrebe daljnje analize pretvoreni u standardizirani format i nakon što je HealthLake kronološki poredao sve događaje radi se analiza podataka i predviđanja. Kada su podaci pretvoreni u FHIR format, alatima QuickSight, SageMaker i aplikacijama treće strane više nije problem provoditi modele strojnog učenja i umjetne inteligencije nad tim podacima.

Amazon SageMaker pomaže u identificiranju trendova i razvijanju interaktivnih upravljačkih ploča, Amazon QuickSight kreira prediktivne modele poput predviđanja mogućnosti razvijanja određenog zdravstvenog stanja (npr. sepsa u bolničkim intenzivnim jedinicama).

Slika 17 Korištenje HealthLake platforme za provođenje analitike



Izvor: AWS



Na sljedećim slikama bit će prikazano kako korisnici HealthLakea pristupaju i koriste svoja spremišta podataka.

Kako bi se podaci uopće mogli uvezati, analizirati i naposljetku izvoziti, prvo je potrebno otići na web stranicu Amazon Health Lake i kreirati spremište podataka. Svi podaci koji će biti u Amazon HealthLakeu su podaci koji su bili nestrukturirani ili polustrukturirani ali zbog lakših izvođenja operacija nad njima pretvoreni su u format FHIR R4. Zbog sigurnosti podataka platforma nudi enkripciju svih podataka u spremištu tako da kreira sigurnosni ključ koji posjeduje i kojim upravlja ali daje mogućnost i korisniku da uveze svoj sigurnosni ključ tako da kroz sve vrijeme korištenja, korisnik ima kontrolu nad svojim podacima.

Slika 18 Kreiranje spremišta podataka

**Create datastore** [Info](#)  
Create a datastore to import, analyze, and export data.

**Datastore settings**

Name of datastore - *optional*  
MyFHIRDataStore123  
The name can have up to 256 characters. Valid characters: A-Z, a-z, 0-9, space, and % + . / : = @ \_ - (hyphen).

**Format**  
We currently support only FHIR version R4.  
FHIR R4

**Preload data - optional** [Info](#)  
If you would like to explore Health AI before importing your own data, you can use the preloaded Synthea dataset.  
 Preloaded sample. The only version supported is R4 from Synthea.

Cancel **Create datastore**

Izvor: AWS

Nakon što se kreira spremište podataka, nude se opcije za upravljanje njime. To su opcije brisanja spremišta podataka, pretraživanja spremišta, unosa novih podataka u spremište i izvoza podataka van spremišta podataka.

Slika 19 Naredbe u spremištu podataka

Amazon HealthLake > Data Stores > satbtestdatastore

**satbtestdatastore** [Info](#) Delete Run query Import Export

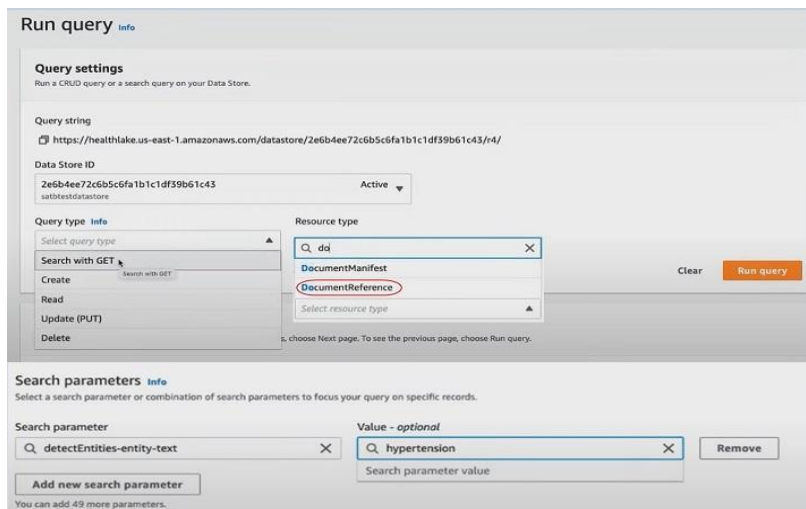
**Data Store details**

Data Store ID 2e6b4ee72c6b5c6fa1b1c1df39b61c43	Format FHIR R4	Encryption key -
Data Store name satbtestdatastore	Created at January 30, 2021, 20:57 (UTC-08:00)	Endpoint <a href="https://healthlake.us-east-1.amazonaws.com/datastore/2e6b4ee72c6b5c6fa1b1c1df39b61c43/r4/">https://healthlake.us-east-1.amazonaws.com/datastore/2e6b4ee72c6b5c6fa1b1c1df39b61c43/r4/</a>
Status <span>Active</span>	Preloaded with Synthea Yes	ARN <a href="arn:aws:healthlake:us-east-1:205474857088:datastore/fhir/2e6b4ee72c6b5c6fa1b1c1df39b61c43">arn:aws:healthlake:us-east-1:205474857088:datastore/fhir/2e6b4ee72c6b5c6fa1b1c1df39b61c43</a>

Izvor: AWS

Ukoliko korisnik želi pretraživati podatke unutar HealthLake platforme, to može učiniti tako što ode na odjeljak Run query. Podatke može pretraživati, kreirati, čitati, ažurirati ili brisati a prilikom pretraživanja podatke može filtrirati i po vrsti izvora kojih ima čak 71. Na slici 14 je odabran izvor DocumentReference jer se radi o području u koje dolaze nestrukturirani medicinski podaci a to u sljedećem koraku definira listu parametara za pretraživanje. Za ovu svrhu zadana je naredba da se pretraže svi entiteti koji sadrže vrijednost *hypertension* odabirom opcije Run query.

Slika 20 Pretraživanje upita



Izvor: AWS

Rezultat koji se dobije je zapisan u kodiranom nizu formata base64 u kojem su sadržani svi nestrukturirani podaci poput bilješki koje je medicinsko osoblje naznačilo prilikom pacijentovog posjeta ordinaciji ili bolničkoj ustanovi. Dakle, kada bi se ovaj kodirani tekst dekodirao korisnici bi u obliku izvornog teksta vidjeli podatke o pacijentu poput dobi, spola, povijesti bolesti itd. Sada je radnja pretraživanja podataka u HealthLakeu puno jednostavnije zbog korištenja generiranih kodova za bolesti ili lijekove.

Model strojnog učenja tj. obrada prirodnog jezika prepoznaje zdravstveno stanje pacijenata kroz različite formate u koje doktori upisuju stanja npr. tekstualne bilješke ili kodovi iz međunarodne klasifikacije bolesti i pruža rezultat pouzdanosti prediktivnih modela. Unos nestrukturiranih podataka u prediktivne analize će uvelike pomoći u kvaliteti rezultata proizašlih iz modela strojnog učenja jer su medicinski zapisi većinom u nestrukturiranom obliku.





### 5.3. Primjeri primjene i doprinosa Amazon HealthLake-a u zdravstvu

Amazon HealthLake je osmišljen kao pomoć zdravstvenim i farmaceutskim poduzećima u upravljanju podacima koje proizvode. No, platforma uz pomoć Amazonovih specijaliziranih aplikacija može imati i širu upotrebu, poput upravljanja troškovima i kvalitetom njege, zdravljem populacije, smanjiti nepotrebne procedure, upravljati kapacitetom bolničkih smještajnih jedinica itd.

U ovom poglavlju će biti ilustriran primjer izrade prediktivnih modela bolesti pomoću Amazon SageMaker s normaliziranim podacima HealthLakea. Podaci koji su korišteni u ovoj analizi su šifrirani podaci pacijenata koji su bili primljeni na odjel intenzivne njege. Podaci uključuju attribute poput korištenih lijekova, dobi, spoli, zdravstvenog stanja i kliničke bilješke. Početni modeli koje je razvila aplikacija SageMaker oslanjali su se na strukturirane podatke iz HealthLakea, da bi se kasnije dodali i modeli koji su prošireni dodatnim podacima, prvotno nestrukturiranim, iz kliničkih bilješki. Modeli su postavljeni na taj način zbog krajnjeg cilja a to je bila usporedba rezultata i izvedbe modela.

Prilikom ovakvih eksperimenata, modeli se mogu modelirati kao problemi ranije spomenutog nadziranog učenja ili nenadziranog učenja. Kod nadziranog učenja, modeli se „treniraju“ i kreiraju uz pomoć podataka koji su ušli u sustav kroz eksperimentiranje ili preuzimanje iz sekundarnih izvora a sve kako bi se mogao predvidjeti krajnji rezultat na temelju ulaznih podataka.

Ukoliko analitičari žele predvidjeti npr. srčani udar kod pacijenta, imaju poznate demografske i zdravstvene podatke. Te podatke uz pomoć algoritama mogu usporediti s onima iz arhive pacijenata koji su prije doživjeli srčani udar i dobiti predviđanja za svog pacijenta.

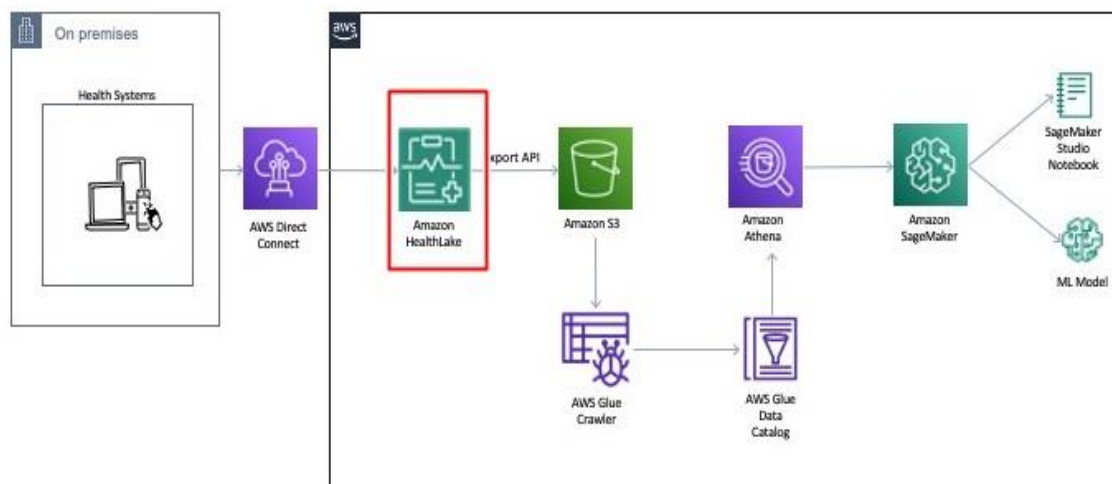
S druge strane, nenadzirano učenje ima za cilj pronaći skrivene uzorke i zajedničke strukture bez prethodnog poznavanja podataka.<sup>62</sup> Kada pronađe skrivene uzorke i zajedničke strukture, svi se podaci tada grupiraju prema sličnim karakteristikama i na kraju prezentiraju u komprimirani obliku.

---

<sup>62</sup> Bolf, N., i Bolf (ur.), N. (2021). 'Osvježimo znanje:

Strojno učenje', *Kemija u industriji*, 70(9-10), str. 591-593. Preuzeto s: <https://hrcak.srce.hr/263495>, Pristupano 28.08.2022.

Slika 23 Aplikacije potrebne za provođenje eksperimenta



Izvor: AWS

Prilikom kreiranja ovog modela, analitičari su preuzeli sve dostupne podatke, normalizirali ih i sve značajke su pretvorene u popise kako bi se kasnije mogao koristiti set za obuku u modelu.

U eksperimentu su određene pozitivna i negativna klasa. Pozitivna klasa je imala 500 pacijenata kojima je bilo dijagnosticirano zatajenje srca a za negativnu klasu je određeno nasumičnih 500 pacijenata koji nisu imali dijagnozu zatajenja srca. Sva klinička stanja iz pozitivne klase koja bi ukazivala na zatajenje srca su bila uklonjena kako bi bilo sigurno da model ne odgovara kliničkom stanju. Osnovni model sa strukturiranim podacima imao je točnost od 85% a dodavanjem značajki izdvojenih iz HealthLakea točnost modela je povećana za 4%. Modeli su pokazali kako je nakon dodavanja značajki iz HealthLakea smanjen i broj lažno pozitivnih s 20 na 13 i broj lažno negativnih s 27 na 20.

Podaci iz HealthLakea su pokazali velik značaj u obogaćivanju modela i analogno tome otkrivanju bolesti. Na sljedećoj slici se može vidjeti kako je pacijent koji je bio lažno negativan u modelu sa samo strukturiranim podacima bio puno siromašniji podacima iz kojih se moglo zaključiti da postoji opasnost od srčanog udara.

Za razliku od prvog modela, u modelu sa dodanim nestrukturiranim podacima vidi se više zdravstvenih komplikacija i indikacija na zatajenje srce.



analize ili situacije u kojoj je potrebno pregledati određene podatke dobilo se na fleksibilnijem pristupu u korištenju spremišta.

Ipak, postoji i određen broj podatkovnih analitičara koji nisu naklonjeni jezeru podataka i ne vide svijetlu budućnost tom konceptu. Većina ih navodi kako su unatoč velikim novčanim i vremenskim ulaganjima izazovi i problemi ostali isti. Najprihvaćeniji argumenti za napuštanje ovog koncepta među kritičarima su problemi poput sporog učitavanja podataka i naknadne izmjene istih, praćenja podataka u stvarnom vremenu, dugotrajnog pronalaženja podataka zbog njihove brojnosti i manjka stručnjaka za jezera podataka koji su i veliki financijski teret. Također, mnogi navode kako odlaganje podataka u jezero podataka i korištenje istih samo kada oni nekome zatrebaju za analizu nije odlika kvalitetnog pristupa podacima. Primjerice, ukoliko sustav šalje podatke koji govore da će se dogoditi neka katastrofa ukoliko se ne reagira pravovremeno, nije prikladno takve podatke samo odložiti u jezero i naknadno ih koristiti. <sup>63</sup>

U posljednje vrijeme sve više poslovnih analitičara zagovara koncept Lakehouse koji kombinira arhitekturu i procese skladišta i jezera podataka. Ovaj koncept je specijaliziran za upravljanjem podacima u zdravstvu i društvenim znanostima a nudi veće uštede i lakše upravljanje podacima.<sup>64</sup> Za razliku od novih metoda, neki analitičari smatraju kako već postoje rješenja koja svakodnevno koristimo ali da su se nepotrebno zamijenila jezerom podataka.

Jedno od takvih rješenja je metoda ODS (Pohrana operativnih podataka) koja prema mišljenju zagovaratelja ima najbrži i troškovno prihvatljiviji pristup za analizu i odgovaranje na podatkovne događaje u stvarnom vremenu.

Unatoč negativnim kritikama koncepta jezera podataka, broj organizacija i poduzeća koje implementiraju jezera podataka u svoje informatičke sustave iz dana u dan raste. Po svemu sudeći, jezera podataka će se koristiti još dugo jer su potrebe za upravljanjem svih vrsta podataka sve veće. Promjene koje se očekuju se tiču prilagodbe i nadogradnje alata i tehnologija u jezeru podataka, potrebama koje poduzeća ili organizacije iskažu u svojim zahtjevima.

---

<sup>63</sup> Drowning in Data: Why It's Time to End the Healthcare Data Lake, Dostupno na :<https://blog.hl7.org/drowning-in-data-why-its-time-to-end-the-healthcare-data-lake>, Pristupano 29.06.2022.

<sup>64</sup> Databricks Introduces Lakehouse for the Healthcare and Life Sciences Industries, Dostupno na: <https://www.datanami.com/this-just-in/databricks-introduces-lakehouse-for-the-healthcare-and-life-sciences-industries/> (30.06.2022.)

## 6. ZAKLJUČAK

Ovaj rad obrađuje primjenu informacijsko – komunikacijskih tehnologija u dinamičnom poslovnom okruženju zdravstva. Cilj rada je pokazati kako se uz pomoć modernih tehnologija mogu poboljšati i unaprijediti velike djelatnosti poput zdravstva.

Zdravstveni sustavi se diljem svijeta suočavaju s problemima u pružanju kvalitetne i dostupne zdravstvene skrbi zbog promijenjene demografske slike i sve starijeg stanovništva. Tradicionalni procesi više ne ispunjavaju sve zahtjeve pacijenata i zdravstveni resori i organizacije traže rješenja u novim tehnologijama i pristupima poput telemedicine, strojnog učenja, robotske automatizacije procesa, umjetne inteligencije, interneta stvari itd. Koristeći navedene tehnologije, transformacija se očekuje na svim razinama, od pristupa u liječenju do poslovne strategije i organizacijskih procesa.

Jezero podataka predstavlja svojstveno proširenje u konceptu skladištenja podataka i oslanja se na klasično skladište podataka. Drugačije je jer osim strukturiranih podataka pohranjuje i nestrukturirane podatke čiji su izvori mobilne aplikacije, IoT senzori ili društvene mreže. Struktura podataka ili shema u jezeru podataka je nedefinirana, što znači da je moguće pohraniti sve podatke bez pažljivog dizajna ili potrebe da znate svrhu za koju će vam podaci trebati.

U zdravstvenoj industriji različiti izvori Velikih podataka uključuju bolničke kartone, medicinske kartone pacijenata, rezultate liječničkih pregleda i uređaje koji su dio Interneta stvari. Složenost i porast podataka u zdravstvu znači da će se umjetna inteligencija i strojno učenje sve više primjenjivati u tom području. Nekoliko vrsta modela umjetne inteligencije se već koristi od strane platitelja i pružatelja skrbi, te zdravstvenih organizacija.

Amazon HealthLake je platforma koja omogućuje pružateljima zdravstvenih usluga, zdravstveno osiguravajućim društvima i farmaceutskim poduzećima pohranu, transformaciju, upite i analizu zdravstvenih podataka. Posjeduje alate koji mogu analizirati trendove i raditi prediktivne modele za zdravstvena stanja pacijenata i mnogih drugih organizacijskih procesa.

Podaci se u doba Velikih podataka smatraju bogatstvom koje u sebi sadrži brojne koristi. Uz nove koncepte pohrane podataka i korištenja digitalnih tehnologija, zdravstvo očekuje veliki napredak u svom poslovanju i načinu donošenja odluka.

## 7. LITERATURA

1. Aburawi, Y., & Albaour, A., Big Data: Review Paper. International Journal Of Advance Research And Innovative Ideas In Education, 7, 2021, 2021.
2. AWS (2020), Introducing Amazon HealthLake to make sense of health data, Dostupno na: <https://aws.amazon.com/about-aws/whats-new/2020/12/introducing-amazon-healthlake-to-make-sense-of-health-data/>, Pristupano 26.08.2022.
3. AWS, Derive AI/ML-driven insights from healthcare data using Amazon HealthLake, Dostupno na: <https://aws.amazon.com/blogs/industries/derive-ai-ml-driven-insights-from-healthcare-data-using-amazon-healthlake> (27.08.2022.)
4. AWS QuickSight – The most popular cloud-native, serverless BI service, Dostupno na: <https://aws.amazon.com/quicksight/>, Pristupano 27.08.2022 (27.08.2022.)
5. Bakri, M. (2020). Big data healthcare paper, International Journal of Recent Technology and Engineering (IJRTE) ISSN: str. 2277-3878, 8 (6) Preuzeto sa: [https://www.researchgate.net/publication/343971630\\_Big\\_data\\_healthcare\\_paper](https://www.researchgate.net/publication/343971630_Big_data_healthcare_paper) (Datum pristupa - 29.06.2022.)
6. Bludov, S. Head in the clouds: what the future of cloud computing means for media, Dostupno na: <https://www.techradar.com/news/head-in-the-clouds-what-the-future-of-cloudcomputing-means-for-media>, Pristupano 24.08.2022.
7. Bolf, N., i Bolf (ur.), N. (2021). 'Osvježimo znanje: Strojno učenje', Kemija u industriji, 70(9-10), str. 591-593. Preuzeto s: <https://hrcak.srce.hr/263495>, Pristupano 28.08.2022.
8. Br Gallestey, J. Bacallao (2020). Framingham Heart Study, Dostupno na: <https://www.britannica.com/event/Framingham-Heart-Studayitannica.com/>, Pristupano 28.08.2022.
9. Brackenbury, W. & Liu, R., & Mondal, M. & Elmore, A., Ur, B., Chard, K., Franklin, M. (2018). Draining the Data Swamp: A Similarity-based Approach. str. 1-7. <https://doi.org/10.1145/3209900.3209911>
10. Cocir.org, eHealth Interoperability, Dostupno na: <https://www.cocir.org/regulations/digital-health/ehealth-interoperability.html>, Pristupano 23.08.2022.
11. Coh T. & Brody, J. (2021). Applications of Machine Learning in Healthcare, DOI: 10.5772/intechopen.92297



12. Čizmić J., Pravo na pristup podacima u medicinskoj dokumentaciji (1991) Zbornik pravnog fakulteta Sveučilišta u Rijeci 30(1), str. 93
13. Ćurko, K. i Španić Kazen, M. (2016.), Skladištenje podataka, put do znanja i poslovne inteligencije, Ekonomski fakultet u Zagrebu
14. Data Robot (2016), What is Natural Language Processing? Introduction to NLP, Dostupno na: <https://www.datarobot.com/blog/what-is-natural-language-processing-introduction-to-nlp/>, Pristupano 26.08.2022.
15. Databricks Introduces Lakehouse for the Healthcare and Life Sciences Industries, Dostupno na: <https://www.datanami.com/this-just-in/databricks-introduces-lakehouse-for-the-healthcare-and-life-sciences-industries/> (30.06.2022.)
16. Derakhshannia, M. & Gervet, C. & Hajj-Hassan, H. & Laurent, A, Martin, A. (2020). Data Lake Governance: Towards a Systemic and Natural Ecosystem Analogy. Future Internet. 12. 126. doi: 10.3390/fi12080126.
17. Dinesh K. G., Arumugaraj K. , Santhosh K.D., Mareeswari V., (2018.), Prediction of Cardiovascular Disease Using Machine Learning Algorithms, International Conference on Current Trends towards Converging Technologies (ICCTCT), 2018, pp. 1-7, <https://doi.org/10.1109/ICCTCT.2018.8550857>
18. Diving in: Navigating a data lake for predictive care, EMC, 2015. Dostupno na: <https://www.dell.com/community/s/vjauj58549/attachments/vjauj58549/solutions-ch/673/1/h14417-emc-navigating-data-lake-predictive-care.pdf> (28.06.2022.)
19. Drowning in Data: Why It's Time to End the Healthcare Data Lake, Dostupno na: <https://blog.hl7.org/drowning-in-data-why-its-time-to-end-the-healthcare-data-lake>, Pristupano 29.06.2022.
20. Džakula, A., Sagan, A., Pavić, N., Lončarek, K. (2014) Health Systems in Transition - Croatia Health System Review 2014. The European Observatory on Health Systems and Policies, Copenhagen. Međunarodna studija, Dostupno na: <https://apps.who.int/iris/handle/10665/128604>, Pristupano 16.08.2022.
21. Europska komisija, Health at a Glance: Europe 2020, State of health in the EU cycle, Dostupno na: [https://health.ec.europa.eu/system/files/2020-12/2020\\_healthatglance\\_rep\\_en\\_0.pdf](https://health.ec.europa.eu/system/files/2020-12/2020_healthatglance_rep_en_0.pdf), Pristupano 25.08.2022.
22. Fang, H. (2015). Managing data lakes in big data era: What's a data lake and why has it become popular in data management ecosystem. IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems, 56, 820-824.



23. Franjić, D., & Miljko, M. (2020). Umjetna inteligencija u radiologiji: Etički problemi. *Zdravstveni glasnik*, 6, 61-68.
24. Gopal, G., Suter-Crazzolaro, C., Toldo, L. and Eberhardt, W. (2019) Digital transformation in healthcare – architectures of present and future information technologies. *Clinical Chemistry and Laboratory Medicine (CCLM)*, Vol. 57 (Issue 3), pp. 328-335. <https://doi.org/10.1515/cclm-2018-0658>
25. Hai R., Geisler, S. & Quix, C. (2016). Constance: An intelligent Data Lake system. *Conference proceedings*, 82, 2097-2100
26. Hall, Heather MBA, MA, MPAS, PA-C. The effect of the COVID-19 pandemic on healthcare workers' mental health. *JAAPA: July 2020 - Volume 33 - Issue 7 - p 45-48* doi: 10.1097/01.JAA.0000669772.78848.8c
27. Hrvatski zdravstveno-statistički ljetopis za 2020., Dostupno na: <https://www.hzjz.hr/hrvatski-zdravstveno-statisticki-ljetopis/hrvatski-zdravstveno-statisticki-ljetopis-za-2020-tablicni-podaci/>, Pristupano 19.08.2022.
28. Hussain, Mohammad Javed. (2020). Big Data in Healthcare. *International Journal of Recent Technology and Engineering*. 8 (6), str. 2127-2131, <http://www.doi.org/10.35940/ijrte.F8100.038620>
29. IBM, Data architecture:IBM's POV – IBM Cloud Architecture Center, Dostupno na: <https://www.ibm.com/cloud/architecture/architectures/dataArchitecture/>, Pristupano 28.08.2022.
30. Kern, J., Elektronički zdravstveni zapis kao predmet interesa korisnika zdravstvene zaštite, Dostupno na: <https://hrcak.srce.hr/file/309835>, Pristupano 24.08.2022.
31. Kovač, N. (2013.) *Financiranje zdravstva – situacija u Hrvatskoj*, str. 553, Dostupno na: <https://hrcak.srce.hr/file/172162>, Pristupano 21.08.2022.
32. Laney, D., 2001. 3D data management: Controlling data volume, velocity and variety. *META group research note*, 6(70), p.1.
33. Lopez, Fernando & Núñez Valdez, E. , García Díaz, V., & Bursac, Z. (2020). A Case Study for a Big Data and Machine Learning Platform to Improve Medical Decision Support in Population Health Management. *Algorithms* 13 (4), 102 <https://doi.org/10.3390/a13040102>
34. Maini E., Venkateswarlu, B., Gupta, A. (2018). Data Lake-An Optimum Solution for Storage and Analytics of Big Data in Cardiovascular Disease Prediction System *International Journal of Computational Engineering & Management*, 21, 861-869.

35. McKinsey Global Institute (2011), Big Data: The Next Frontier for Innovation, Competition, and Productivity, Dostupno na: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation>, Pristupano 15.09.2022.
36. Medium.com (2018.), Distinguishing between Narrow AI, General AI and Super AI, Dostupno na: <https://medium.com/mapping-out-2050/distinguishing-between-narrow-ai-general-ai-and-super-ai-a4bc44172e22>, Pristupano 15.09.2022.
37. Miloslavskaya N, and Tolstoy A., (2016.), Big Data, Fast Data and Data Lake Concepts, preuzeto [https://www.researchgate.net/publication/309183107\\_Big\\_Data\\_Fast\\_Data\\_and\\_Data\\_Lake\\_Concepts](https://www.researchgate.net/publication/309183107_Big_Data_Fast_Data_and_Data_Lake_Concepts) (Datum pristupa: 24.08.2022.)
38. Novi list, 2020 Digitalno doba je već tu: Telemedicina bi mogla zamijeniti čak 75 posto posjeta doktoru, Dostupno na: [https://www.novolist.hr/ostalo/sci-tech/tehnologija/digitalno-doba-je-vec-tu-telemedicina-bi-mogla-zamijeniti-cak-75-posto-posjeta-doktoru/?meta\\_refresh=true](https://www.novolist.hr/ostalo/sci-tech/tehnologija/digitalno-doba-je-vec-tu-telemedicina-bi-mogla-zamijeniti-cak-75-posto-posjeta-doktoru/?meta_refresh=true) Pristupano 24.08.2022.
39. OECD iLibrary, Health expenditure per capita, Dostupno na: <https://www.oecd-ilibrary.org/sites/154e8143-en/index.html?itemId=/content/component/154e8143-en>, Pristupano 23.08.2022.
40. Oliveira Hashiguchi, T. (2020), "Bringing health care to the patient: An overview of the use of telemedicine in OECD countries", OECD Health Working Papers, No. 116, OECD Publishing,
41. Opis zdravstvenog sustava, Dostupno na: <https://hzzo.hr/pravni-akti/opis-zdravstvenog-sustava>, Pristupano 16.08.2022.
42. Pejić Bach, M., Spremić, M. & Suša Vugec, D. (2018) Integrating Digital Transformation Strategies into Firms: Values, Routes and Best Practice Examples. U: Novo Melo, P. & Machado, C. (ur.) Management and Technological Challenges in the Digital Age. Boca Raton, Florida, Taylor & Francis Group: CRC press, str. 6
43. Philips.com, Dostupno na: <https://www.philips.com/c-dam/corporate/newscenter/global/standard/resources/healthcare/2021/digital-transformation/philips-digital-transformation-position-paper.pdf>, Pristupano 23.08.2022.
44. Prister, V. (2019). Umjetna inteligencija, Media, culture and public relations, 10(1), str. 67-72. <https://doi.org/10.32914/mcpr.10.1.7>

45. Purra S. B., & Pasupuleti P. (2015.) Data Lake development with Big Data. Birmingham: Packt Publishing
46. Raghupathi, V., Raghupathi, W. (2014). Big data analytics in healthcare: Promise and potential, Health Information Science and Systems, 2 (3) <https://doi.org/10.1186/2047-2501-2-3>
47. Rouse, M. (2018.) Data storage, Dostupno na: <https://searchstorage.techtarget.com/definition/storage>, Pristupano 24.08.2022.
48. Saleh, Z. (2019). Artificial Intelligence Definition, Ethics and Standards, Dostupno na: [https://www.researchgate.net/publication/332548325\\_Artificial\\_Intelligence\\_Definition\\_Ethics\\_and\\_Standards](https://www.researchgate.net/publication/332548325_Artificial_Intelligence_Definition_Ethics_and_Standards), Pristupano 15.09.2022.
49. Shan, Keerthi & John, E P. (2022). Adoption of Digital Health Care -A Reality in Future. International Journal of Innovative Research in Education. 8. str. 370-380.
50. Simplilearn.com (2022.) Advantages and Disadvantages of Artificial Intelligence, Dostupno na: [https://www.simplilearn.com/advantages-and-disadvantages-of-artificial-intelligence-article#advantages\\_&\\_disadvantages\\_of\\_artificial\\_intelligence](https://www.simplilearn.com/advantages-and-disadvantages-of-artificial-intelligence-article#advantages_&_disadvantages_of_artificial_intelligence), Pristupano 15.09.2022.
51. Somek, M. (2021) Što skrivaju masovni podaci u zdravstvu? Bilten Hrvatskog društva za medicinsku informatiku, 27 (2), 23-30.
52. Spremić M. (2020.) Sigurnost i revizija informacijskih sustava u okruženju digitalne ekonomije, Sveučilište u Zagrebu, Ekonomski fakultet
53. Spremić, M. (2017). Digitalna transformacija poslovanja, Sveučilište u Zagrebu, Ekonomski fakultet
54. State of Health in the EU - Hrvatska, Dostupno na: [https://health.ec.europa.eu/system/files/2022-01/2021\\_chp\\_hr\\_croatian.pdf](https://health.ec.europa.eu/system/files/2022-01/2021_chp_hr_croatian.pdf), Pristupano 19.08.2022
55. Stein, B., Morrison, A. (2014). The enterprise data lake: Better integration and deeper analytics. Price water house Cooper. URL: <http://www.pwc.com/us/en/technology-forecast/2014/cloudcomputing/assets/pdf/pwc-technology-forecast-data-lakes.pdf>
56. Tomcy J., Panhay - Tomcy J., Panhay M. (2018). Data lake for Enterprises. Birmingham: Packt Publishing
57. Ucionica.net, e-Gradani Portal zdravlja: kako do nalaza, uputnica, lijekova, Dostupno na: <https://www.ucionica.net/aplikacije/e-gradani-portal-zdravlja-kako-do-nalaza-uputnica-lijekova-9971/>, Pristupano 23.08.2022.

58. Unified Data Lake for Healthcare Research and Analytics, Dostupno na: <https://www.bluematrix.com/post/unified-data-lake-for-healthcare-research-and-analytics> Pristupano
59. Varga, M. (2014). Upravljanje podacima, 2. izdanje, Zagreb: Element
60. World Health Organization (2019), Global spending on health: A world in transition, Dostupno na: <https://www.who.int/publications/i/item/WHO-HIS-HGF-HFWorkingPaper-19.4>, Pristupano 23.08.2022
61. Zakon o zdravstvenoj zaštiti, Narodne novine NN 100/18, 125/19, 147/20, Dostupno na: <https://www.zakon.hr/z/190/Zakon-o-zdravstvenoj-za%C5%A1titi>, Pristupano 22.08.2022.
62. Zburivsky, D. & Partner, L. (2020). Designing Cloud Data Platforms. Manning Publications, 19, 564-570.
63. zdravstvo. Hrvatska enciklopedija, mrežno izdanje. Leksikografski zavod Miroslav Krleža, 2021. Pristupljeno 9. 8. 2022. <<http://www.enciklopedija.hr/Natuknica.aspx?ID=67025>>.
64. Zicari, R.V, (2014.), Challenges and opportunities u : Akerkar, R., Big Data Computing, Sogndal: Taylor and Francis

## POPIS SLIKA

Slika 1 Manjak zdravstvenih radnika u odnosu na prosjek EU .....	6
Slika 2 Potrošnja na zdravstvo u članicama EU .....	7
Slika 3 Interoperabilnost u zdravstvu.....	11
Slika 4 Portal Zdravlje.....	12
Slika 5 Glavne značajke jezera podataka .....	16
Slika 6 Strukturirani i nestrukturirani podaci .....	18
Slika 7 Prikaz arhitekture jezera podataka .....	22
Slika 8 Vrste umjetne inteligencije .....	30
Slika 9 Model rizika za kardiovaskularne bolesti pacijenta .....	35
Slika 10 Autentifikacija HZZO pametnom karticom.....	41
Slika 11 Različiti izvori podataka koji ulaze u jezero podataka.....	42
Slika 12 Prikaz sučelja Amazon HealthLake platforme.....	44
Slika 13 Izvlačenje značenja iz nestrukturiranih zdravstvenih podataka iz kliničke bilješke .	45
Slika 14 Dodjeljivanje medicinskih kodova tekstu.....	46
Slika 15 Interoperabilnost unutar zdravstvene djelatnosti .....	46
Slika 16 Kronološki pregled događaja tijekom liječenja .....	47
Slika 17 Korištenje HealthLake platforme za provođenje analitike .....	47
Slika 18 Kreiranje spremišta podataka.....	48
Slika 19 Naredbe u spremištu podataka .....	48
Slika 20 Pretraživanje upita.....	49
Slika 21 Kodirani niz nestrukturiranih medicinskih podataka .....	50
Slika 22 Identificiranje zdravstvenog stanja.....	50
Slika 23 Aplikacije potrebne za provođenje eksperimenta .....	52
Slika 24 Usporedba rezultata modela 1 i modela 2.....	53

## **POPIS GRAFIKONA**

Grafikon 1 Struktura zaposlenih doktora medicine i medicinskih sestara/tehničara po spolu 2020.god. (u tisućama).....	6
Grafikon 2 Izdaci za zdravstvo po osobi u zemljama članicama OECD-a u 2019.god.....	9
Grafikon 3 Prednosti u pružanju skrbi telemedicinom .....	14

## POPIS TABLICA

Tablica 1 Prednosti i nedostaci korištenja jezera podataka .....	17
Tablica 2 Usporedba skladišta podataka i jezera podataka .....	19
Tablica 3 Osnovne karakteristike velikih podataka u zdravstvu .....	32

# ŽIVOTOPIS



Ivan Šimić

## WORK EXPERIENCE

---

[ 03/03/2017 – 06/03/2020 ] **Back office administrators**

*Hrvatski Telekom d.d*

**City:** Zagreb

**Country:** Croatia

**Main activities and responsibilities:**

Administracija

Kreiranje ugovora

Podrška prodajnim mjestima

Rješavanje tehničkih problema

## EDUCATION AND TRAINING

---

[ 01/09/2011 – 18/06/2015 ] **Opća gimnazija Lovre Monti**

*Srednja škola Lovre Monti*

**Address:** Ikičina 30, 22300, Knin, Croatia

**Celonis Fundamentals Training**

*Celonis* [www.celonis.com](http://www.celonis.com)

**Address:** Trg John F. Kennedy 1, 10000, Zagreb, Croatia

## LANGUAGE SKILLS

---

**Mother tongue(s):** English

**Other language(s):**

**engleski**

LISTENING C2 READING B1 WRITING B2

SPOKEN PRODUCTION B2 SPOKEN INTERACTION C1

**German**

LISTENING A2 READING B1 WRITING A2

SPOKEN PRODUCTION A2 SPOKEN INTERACTION A2

## DIGITAL SKILLS

---

Microsoft Office | Social Media | Zoom | Google Docs | Skype | Microsoft Powerpoint  
| Outlook | Instagram | Google Drive | Power Point | Organizational and planning  
skills | Microsoft Excel | Good listener and communicator | Motivated | Written and  
Verbal skills | Decision-making | Team-work oriented | LinkedIn | Facebook | Twitter  
| Presenting | Critical thinking | Reliability



