Marko Petek, Kristina Gruden in Špela Baebler

Standardizacija in upravljanje z odprtimi podatki v biotehnoloških raziskavah

Abstract

Standardization and Open Data Management in Biotechnology Research

Biotechnological research generates large amounts of diverse data that is challenging to manage in a findable, accessible, interoperable and reusable (FAIR) manner. Various ways of standardization such as ontologies, terminologies, data standards and minimum reporting standards are being established in the research community. Establishing effective data management in international projects also presents an additional challenge. In the article, the authors present an example of good practice in establishing the management of international biotechnological projects. The approach to data management was determined through a data management plan. Data is managed using a combination of the pISA-tree local data management system, a synthetic-biological data management system and online data repositories. Such good practices will lead to the wider introduction of open data management following FAIR guidelines, which will lead to faster and cheaper biotechnological discoveries in the long run.

Keywords: biotechnology, data management, FAIR principles, data repositories, standards

Marko Petek, Kristina Gruden and Špela Baebler are researchers at the National Institute of Biology's Department of Biotechnology and Systems Biology. Marko Petek works in the field of plant biotechnology and molecular biology, where he is mostly involved in bioinformatic data analysis of high-throughput methods (marko.petek@nib.si). Kristina Gruden leads an interdisciplinary group in the field of plant biology that both generates and analyzes data (kristina.gruden@nib.si). Špela Baebler is a project leader in the field of biotechnology, where she connects laboratory procedures with data analysis (spela.baebler@nib.si).

Povzetek

V biotehnoloških raziskavah generiramo velike količine raznovrstnih podatkov, zato je njihovo obvladovanje, tj. da so najdljivi, dostopni, interoperabilni in ponovno uporabni (ang. *Findable, Accessible, Interoperable, Reusable* – FAIR), velik izziv. V skupnosti

se tako uveljavljajo različni načini standardizacije, kot na primer ontologije, terminologije, podatkovni standardi in minimalni standardi poročanja. Dodaten izziv je vzpostavitev učinkovitega upravljanja s podatki v mednarodnih projektih. V prispevku predstavljamo primer dobre prakse vzpostavitve upravljanja biotehnoloških mednarodnih projektov. Način upravljanja podatkov smo določili z načrtom upravljanja s podatki. Podatke upravljamo s kombinacijo sistema za lokalno upravljanje s podatki pISA-tree, sistema za upravljanje s sintezno-biološkimi podatki in spletnih podatkovnih repozitorijev. Prav take dobre prakse bodo vodile do širše uveljavitve upravljanja odprtih podatkov v skladu s smernicami FAIR, kar bo dolgoročno vodilo do hitrejših in cenejših biotehnoloških odkritij.

Ključne besede: biotehnologija, upravljanje s podatki, načela FAIR, podatkovni repozitoriji, standardi

Marko Petek, Kristina Gruden in Špela Baebler so raziskovalci na Nacionalnem inštitutu za biologijo, na Oddelku za biotehnologijo in sistemsko biologijo. Marko Petek deluje na področju rastlinske biotehnologije in molekularne biologije, kjer se ukvarja predvsem z bioinformatsko obdelavo masovnih bioloških podatkov (marko.petek@nib.si). Kristina Gruden vodi interdisciplinarno skupino na področju biologije rastlin, ki podatke ustvarja in hkrati analizira (kristina.gruden@nib.si). Špela Bebler vodi projekte s področja biotehnologije, kjer laboratorijske postopke povezuje z analizo podatkov (spela.baebler@nib.si).

Zakaj potrebujemo odprte podatke v biotehnologiji?

Dandanes je za kakovostno raziskovanje na področju biotehnologije nujno povezovanje v obliki mednarodnih in interdisciplinarnih delovnih skupin, v katerih poteka stalna komunikacija in odprta izmenjava izkušenj, mnenj in raziskovalnih podatkov. Revolucija visokozmogljivih metod v bioznanostih je v preteklih dveh desetletjih omogočila hitro pot do novih odkritij, a je hkrati ustvarila ogromne količine podatkov, znotraj enega samega raziskovalnega projekta lahko tudi nekaj terabajtov. Zaradi velikih količin podatkov smo raziskovalci, tako bioinformatiki kot biologi, dokaj hitro spoznali, da je treba podatke in z njimi povezane metapodatke zbrati in urediti že med raziskovanjem, še bolje pa je, da o urejanju podatkov razmislimo že pred izvedbo eksperimentov. Če tega ne storimo, se kaj kmalu zgodi, da se izgubimo v poplavi datotek, kar podaljša čas za analizo podatkov, interpretacijo in objavo rezultatov, v najslabšem primeru pa končnih rezultatov sploh ne pridobimo ali pa so zaradi slabo opisanih podatkov napačno interpretirani. Poskrbeti moramo tudi za varno shranjevanje in deponiranje surovih podatkov v prosto dostopnih bazah, kar omogoča transparentnost in njihovo ponovno uporabo. Tak način deponiranja zahtevajo vse ugledne znanstvene revije in javne raziskovalne agencije.

Zato je znanstvena skupnost načela upravljanja podatkov opredelila pod akronimom FAIR (ang. *Findable, Accessible, Interoperable, Reusable*; najdljivi, dostopni, interoperabilni, ponovno uporabni; Wilkinson idr., 2016). Upoštevanje teh načel zagotavlja dolgoročno dostopnost do raziskovalnih podatkov v digitalni obliki, njihovo izmenjavo, medsebojno povezovanje ter ponovno uporabo. Podatkom, ki so skladni s temi načeli, pravimo »FAIR podatki«. Sistemi upravljanja FAIR podatkov naj bi bili zasnovani na način, ki stremi k računalniški avtomatizaciji procesov s čim manj potrebe po uporabnikovem posredovanju za ponovno uporabo ali integracijo podatkov. Pri tem pa se je treba zavedati, da podatki nimajo vrednosti, če niso dobro opisani, tj. opremljeni z metapodatki.

K izboljšanju kulture deljenja podatkovnih nizov in večji razširjenosti FAIR podatkov pripomorejo tudi splošne in področne podatkovne znanstvene revije. Te omogočajo objavo prostobesedilnih opisov podatkovnih nizov s povezavami do repozitorijev, v katerih so podatki dostopni, ter primeri uporabnosti podatkov. Prednost objave v takšnih revijah je priznanje vsem raziskovalcem za izvedbo študije, predvsem pa to, da lahko podatke in rezultate delijo z znanstveno skupnostjo brez potrebe po poglobljeni biološki interpretaciji rezultatov, kar omogoča, da so lahko takšni podatki sproščeni veliko prej. Prednost je tudi, da omogočajo objavo kvalitetnih negativnih rezultatov, kar vodi nove študije, da se usmerijo v še neraziskana vprašanja. Najuglednejše revije, ki objavljajo opise bioloških nizov podatkov, so Scientific Data (Nature Research), GigaScience (Oxford University Press, BGI), G3: Genes|Genomes|Genetics (Genetics Society of America) in BMC Research Notes (BioMedCentral).

Zavedati se moramo, da v raziskovalnih projektih, tudi tistih, financiranih iz javnih virov, sodelujejo industrijski partnerji, katerih namen je uporaba ali komercializacija nekaterih rezultatov. Zaradi interesov po varovanju intelektualne lastnine v obliki patentov takojšnja javna objava nekaterih raziskovalnih rezultatov ni možna. To pa ni v konfliktu s konceptom odprte znanosti, saj je eno od načel podatkov FAIR, da so kolikor mogoče odprti, zaprti pa le, kolikor je potrebno (ang. *as open as possible, as closed as necessary*). To načelo je vgrajeno tudi v navodila EU programa Obzorje 2020 (ang. *Horizon 2020*) za odprti dostop in upravljanje podatkov (Evropska komisija, 2020), ki narekuje, da mora načrt za upravljanje podatkov opredeliti, kdaj bodo pridobljeni podatki na voljo za ponovno uporabo, pod kakšnimi pogoji in v primeru zahteve za zakasnitev objave (embargo) zaradi na primer pridobitve patentne zaščite, kako dolgo bo ta trajala in zakaj. Podatki so lahko v skladu s tem načelom dostopni z omejitvami zaradi pravic iz naslova intelektualne lastnine, kar mora biti jasno opredeljeno z uporabo ustreznih

licenc odprtega dostopa (Carroll, 2015). Prav tako so obdobja embarga objave za odprte podatke načeloma kratka in morajo biti dobro utemeljena.

Potreba po standardizaciji

Interdisciplinarnost raziskav pomeni, da v projektnih skupinah sodelujejo različni profili – v bioznanostih prednjačijo biologi, zdravniki, farmacevti, biokemiki ipd., ki izvajajo predvsem poskuse v laboratorijih ali na terenu, ter bioinformatiki, statistiki, matematiki, računalničarji ipd., ki se ukvarjajo predvsem z analizo podatkov. Neizbežno je, da včasih pride do nesporazumov ali nejasnosti, saj pogosto ne poznamo žargona drugih strok. Določeni izrazi in okrajšave namreč lahko imajo različen pomen v različnih strokah ali tudi znotraj iste stroke med različnimi laboratoriji. Ta problem rešimo z uvedbo t. i. področnih slovarjev, ki enoznačno določajo pomen strokovnih têrminov. Termini so lahko zbrani v ontologije, ki za določeno področje znanosti predpisujejo njihovo prednostno rabo in razmerja med njimi. Primer ene najbolj znanih in uporabnih ontologij v biotehnologiji je genska ontologija (Gene Ontology, geneontology.org), ki opisuje funkcijo genov z uporabo terminov za molekulske funkcije, lokacijo v celici in procese, pri katerih sodelujejo genski produkti (The Gene Ontology Consortium, 2019).

V zadnjih dveh desetletjih je izziv tudi zelo hitro uvajanje vedno novih visokozmogljivih metod za določanje genomskih zaporedij, 3D strukture proteinov, izražanje genov (mikromreže, visokozmogljivo sekvenciranje RNA) in fenotipizacijo, na primer spremljanje rasti in vidnih simptomov bolezni ali tretmajev (obdelava slik, videov). Te metode generirajo veliko količino podatkov, vendar ob uvedbi novih metod podatkovni formati in standardi poročanja zanje še niso oblikovani, zato raziskovalci večinoma le proizvajajo končne rezultate in o njih poročajo »nekontrolirano«. Šele s časom se metode »stabilizirajo« ter se uvedejo standardni postopki in standardi poročanja.

Potrebo po standardizaciji lahko opišemo z odmevnim primerom iz medicine (Bustin idr., 2008). Leta 2002 so Uhlman in sodelavci v reviji *Journal of Clinical Pathology* objavili članek, v katerem so pri otrocih z razvojnimi težavami prisotnost nukleinskih kislin virusa rdečk pokazali s takrat še novo metodo, tj. kvantitativno verižno reakcijo s polimerazo (qPCR).[1] To je nakazovalo, da bi delci oslabljenega virusa iz cepiva NMR ostali v organizmu, kar bi lahko posredno potrdilo tezo začetnika nasprotovanja cepljenju,

1 qPCR velja za zlati standard za kvantitativno določanje vsebnosti nukleinskih kislin v vzorcu in se rutinsko uporablja v medicini, na primer za določanje vsebnosti virusnih delcev.

prof. Wakefielda, da cepivo MMR povzroča avtizem. Avtorji v članku niso poročali o podrobnostih izvedbe raziskave, kasneje pa se je izkazalo, da so bili rezultati posledica kontaminacije vzorcev dotičnih otrok z virusom rdečk, ki so ga raziskovali v sosednjem laboratoriju. Raziskava je imela za posledico uveljavitev proticepilskega gibanja, kar bi lahko preprečili z uporabo in poročanjem o ustrezni kontroli kakovosti izvedbe testov. Zato so leta 2009 Bustin in sodelavci objavili minimalne standarde za objavo rezultatov eksperimentov kvantitativne PCR (Minimum Information for Publication of Quantitative Real-Time PCR Experiments, MIQE, Bustin idr., 2009), ki so sedaj pogoj za objavo rezultatov, dobljenih z metodo qPCR. Gre za t. i. »Standard minimalnih informacij«, ki je nabor priporočil za poročanje o podatkih različnih laboratorijskih metod. Standardi minimalnih informacij sledijo modelu JERM (ang. *Just Enough Results Model*, v prevodu ravno dovolj informacij). S poročanjem, skladnim s tem standardom, je zagotovljena ponovna uporaba in interoperabilnost podatkov. Da dosežemo interoperabilnost, pa je treba dodatno standardizirati tudi podatkovne formate.

Ob vsem naštetem je znanstvena skupnost preplavljena s standardi, hkrati pa se veliko podatkov še vedno producira, objavlja in deli na nestandardiziran način. Uporabo standardov pri objavi priporoča ali celo zahteva veliko znanstvenih revij, vendar pa je zaradi slabe informiranosti recenzentov ta vidik pri recenziji pogosto spregledan. Zato obstaja veliko pobud, ki želijo razmere izboljšati. Med njimi je gotovo vredno omeniti portal FAIRsharing (www.fairsharing.org; Sansone idr., 2019), ki zbira, opisuje in rangira obstoječe standarde. Standardizacije v bioznanostih se je prednostno lotil tudi konzorcij nedavno zaključene evropske COST akcije CHARME (ang. *Harmonising standardisation strategies to increase efficiency and competitiveness of European life-science research*), v okviru katere smo prepoznali veliko ozkih grl pri standardizaciji v vedah o življenju, a hkrati predlagali številne rešitve, predvsem na področju izobraževanja različnih deležnikov (Hollman idr., 2020).

Vse naštete standarde je treba upoštevati tudi pri načrtovanju programske opreme za delo s podatki. Tako so podatkovni standardi pogosto upoštevani pri načrtovanju aparatur in laboratorijskih dnevnikov. Kot primer orodja za analizo podatkov, ki podpira načelo odprte znanosti, lahko navedemo odprtodostopno spletno aplikacijo quantGenius, ki je namenjena kvantifikaciji nukleinskih kislin s qPCR (quantgenius.nib.si; Baebler idr., 2017). Aplikacija sledi minimalnemu standardu poročanja za tovrstne podatke MIQE (Bustin idr., 2009) in omogoča urejanje podatkov v lokalni ali spletni bazi. Lokalna podatkovna baza omogoča pregledovanje rezultatov drugih uporabnikov, kar olajša prenos znanja in metod med sodelavci.

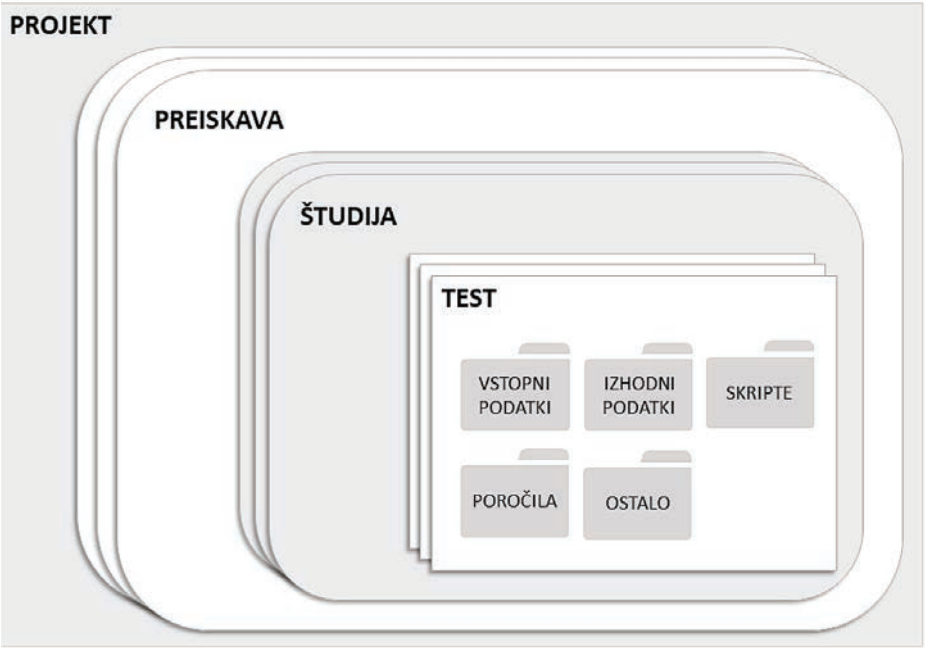
Povezava genskih identifikatorjev s tistimi v datoteki phenodata iz sistema pISA-tree, ki ga podrobneje predstavimo v nadaljevanju, ter povezava z lokalno bazo uporabljenih reagentov in metod pa omogočata popolno sledljivost dobljenih rezultatov.

Upravljanje s podatki v okviru mednarodnih projektov: primer dobre prakse

Evropska komisija se zaveda pomena odprte znanosti, zato v razpise za znanstvene projekte uvaja zahteve po upoštevanju načel FAIR pri podatkovnem upravljanju ter priporočila za objavo rezultatov v znanstvenih revijah z odprtim dostopom. Primer take finančne sheme EU je ERA CoBioTech, ki jo sofinancirajo lokalne agencije oziroma institucije, v Sloveniji je to Ministrstvo za izobraževanje in šport. Na oddelku za biotehnologijo in sistemsko biologijo NIB sodelujemo v dveh ERA CoBioTech projektih, SUSPHIRE (susphire.info) in INDIE (indie.cebitec.uni-bielefeld.de), pri katerih imamo vodilno vlogo pri upravljanju s podatki, sodelujemo pa tudi pri izvajanju bioloških poskusov in analizi podatkov. Že pri prijavi obeh projektov smo morali natančno opredeliti načrt upravljanja z raziskovalnimi podatki (ang. *data management plan*) v skladu z načeli FAIR, določiti osebe, odgovorne za upravljanje, in opredeliti vire, na primer predvidene količine generiranih podatkov, računalniške kapacitete, podatkovne formate in standarde, način lokalne in javne hrambe podatkov ipd. Že na tem mestu smo se odločili, da bomo kot javni centralni repozitorij za rezultate in metapodatke uporabili repozitorij FAIRDOMHub (fairdomhub.org). Surove podatke shranjujemo v za to namenjene področno specifične podatkovne repozitorije, kot določa politika objav večine priznanih znanstvenih revij. V projektu SUSPHIRE so največji nizi podatkov pridobljeni z visokozmogljivim sekvenciranjem RNA za namen spremljanja izražanja genov, za katere uporabljamo podatkovni repozitorij GEO (Gene Expression Omnibus; ncbi.nlm.nih.gov/geo). Metabolomske podatke, pridobljene z bodisi tekočinsko ali plinsko kromatografijo, sklopljeno z masno spektroskopijo (LC-MS ali GC-MS), deponiramo v bazo MetaboLights (ebi.ac.uk/metabolights). Zaporedja DNA bioloških gradnikov deponiramo v centralno bazo podatkov DNA zaporedij GenBank (ncbi.nlm.nih.gov/genbank), saj s tem postanejo del večjih zbirk DNA zaporedij. Načrt upravljanja s podatki smo predstavili na uvodnem sestanku projekta, na katerem smo tudi potrdili skrbnike podatkov (ang. *data stewards*) za vsako od partnerskih organizacij. V načrtu smo opredelili tudi postopek sproščanja podatkov.

Uporaba lokalnega sistema za upravljanje s podatki

Za ta projekta smo nadgradili sistem za lokalno upravljanje s podatki pISA-tree (github.com/NIB-SI/pISA), ki je plod lastnega razvoja. Omogoča enotno organiziranje podatkovnih datotek ter generiranje povezanih metadatotek za opis študij, testov in postopkov. Sistem temelji na modelu ISA (ang. *Investigation/Study/Assay*) oziroma specifikacijah formata ISA-Tab (isa-specs.readthedocs.io) in je implementiran v operacijskem sistemu Windows. Deluje tako, da uporabnika vodi skozi vnaprej pripravljene vprašalnike ter na podlagi odgovorov generira hierarhično strukturo map in tekstovnih metapodatkovnih datotek. V hierarhiji map si sledijo projekt (*project*), preiskava (*Investigation*), študija (*Study*) in test (*Assay*), pri čemer ima vsaka raven svoje metapodatkovne datoteke in mape za različne tipe podatkov. Najzanimivejša je raven testa, na kateri so tipično mape za datoteke z vstopnimi podatki (*input*) in izpeljanimi podatki/rezultati (izhodni podatki – *output*), skripte ali ukazi za računalniško obdelavo podatkov (*scripts*), poročila (*reports*) in drugo (*other*; glej Sliko 1).



Slika 1: Hierarhija strukture pISA-tree za lokalno upravljanje s podatki raziskovalnih projektov. Za raven testov so prikazane mape, ki jih sistem avtomatsko generira.

Na našem oddelku imamo pISA-tree strukturo postavljeno na mrežnem disku, do katerega lahko vedno in hitro dostopamo, vendar ima omejene

kapacitete, zato datotek surovih podatkov (ang. *raw data*) visokozmogljivih metod v velikosti več gigabajtov ne shranjujemo v pISA-tree mape, temveč v drug lokalni repozitorij. V takem primeru pot do surovih podatkov zapišemo v ustrezno pISA-tree metapodatkovno datoteko na ravni testa. Ko surove podatke prenesemo v zgoraj omenjene javne repozitorije (GEO, MetaboLights ipd.), tudi to dopišemo v metapodatkovno datoteko. Sistem je zasnovan tako za poskuse, ki jih izvajamo v molekularnih in biotehnoloških laboratorijih, kot tudi za bioinformatsko in statistično obdelavo podatkov. Dopolnjujeta ga paketa za programsko okolje R *pisar* in *seekr* (oboje dostopno na NIB GitHub strani github.com/NIB-SI), ki omogočata ponovljivo analizo podatkov na osnovi metapodatkovnih datotek in avtomatiziran prenos organizirane lokalne strukture določenega projekta v repozitorij FAIRDOMHub. Za analizo podatkov je ključnega pomena, da jih spremljajo dobro urejeni metapodatki. V biotehnoloških raziskavah teste večinoma izvajamo na t. i. vzorcih, ki lahko predstavljajo posamezne bakterijske kolonije, dele rastlinskega tkiva, manjše količine fermentacijske brozge ipd. Ključno je, da vzorce pred testiranjem ustrezno fizično označimo z nedvoumnimi in enoznačnimi identifikatorji, hkrati pa ustvarimo datoteko, ki natančno opisuje testirane vzorce. V pISA-tree sistemu te datoteke imenujemo »phenodata« (skrajšano za ang. *phenotypic data*, v prevodu fenotipski podatki) in vsebujejo identifikatorje, imena vzorcev, tretiranja vzorcev (na primer tretirani in kontrolni), pogoje gojenja, morfološke značilnosti, datum, čas in pogoje vzorčenja ter druge vzorčno specifične parametre.

Sistem pISA-tree smo partnerjem obeh CoBioTech projektov najprej predstavili na spletnem seminarju, kasneje, ko so vse delovne skupine začele generirati podatke, pa smo organizirali še spletno delavnico za neposredne uporabnike. Dogovorili smo se tudi za fiksno strukturo map na ravni preiskav (*Investigation*), za katero smo pri enem projektu izbrali kar projektne delovne pakete (ang. *work packages*). Na ravni študij (*Study*) pa smo uvedli pravilo za poimenovanje map, ki preprečuje, da bi projektni partnerji iz drugih raziskovalnih skupin nevede enako poimenovali svoje študije. Tako lahko vsak partner lokalno v svoji organizaciji gradi in shranjuje podatke v svojo pISA-tree strukturo. Podatke in analize si lahko na preprost način izmenjujemo in jih neodvisno nalagamo v FAIRDOMHub repozitorij. V skladu z načrtom upravljanja s podatki vsakega pol leta pregledamo podatke in se odločamo o sproščanju v skladu z dogovorjenimi licencami. FAIRDOMHub namreč omogoča, da so podatki nekaj časa dostopni samo projektnim partnerjem.

Upravljanje s sintezno-biološkimi podatki

Uporaba pISA-tree ni primerna za vse tipe podatkov. V zadnjih letih se je zelo uveljavil pristop sintezne biologije, pri katerem iz preprostih gradnikov bioloških molekul sintetiziramo nove gradnike, ki na koncu tvorijo nov organizem. Tako lahko z različnimi kombinacijami gradnikov z zelo podobnimi postopki zgradimo tako rekoč neomejeno število različnih organizmov. Pri teh študijah izvedemo veliko testov, ki pa ne generirajo veliko podatkov in so mnogokrat le vmesni koraki ali koraki v nepravo smer. Zato pristop, pri katerem so podatki organizirani v obliki testov, ki ga uporablja pISA-tree, za tovrstne raziskave ni najbolj praktičen, saj hitro postane nepregleden, pa tudi vnašanje podatkov za nove teste zahteva preveč časa.

Zato smo v okviru obeh prej omenjenih projektov vzpostavili sistem za upravljanje podatkov, ki temelji na standardni skladnji v sintezni biologiji (Patron idr., 2015). Ker so partnerji v projektih vešči standardizacije in tudi sami avtorji standardov, je bila motivacija za vzpostavitev sistema velika. Vendar pa ni bilo enostavno. Partnerji so že imeli vzpostavljene različne sisteme za upravljanje bioloških gradnikov; nekateri niso bili primerni, saj niso bili ustrezno standardizirani ali pa so bili del komercialnih rešitev, kar je onemogočalo njihovo uporabo pri vseh partnerjih projekta, tabele pa so bile nezdružljive. Zato smo zasnovali skupno preglednico, ki vsebuje podatke za različne ravni bioloških gradnikov in je dostopna vsem partnerjem projekta. Zaradi lažje sledljivosti dela med posameznimi partnerji in izognitve neunikatnosti smo posameznim partnerjem dodelili nize identifikatorjev. Pri vnosu smo dopustili možnost t. i. katalogiranja, kar pomeni, da partner v skupno preglednico vnese le identifikator in povezavo na lastno bazo podatkov, kjer so zavedeni vsi metapodatki za posamezen biološki gradnik. Kot drugi podatki projektov bo tudi ta preglednica naložena na FAIRDOMHub.

Sklep

V biotehnoloških raziskavah generiramo velike količine raznovrstnih podatkov, zato je njihovo obvladovanje v skladu s smernicami FAIR velik izziv, dodaten izziv pa je vzpostavitev učinkovitega upravljanja v mednarodnih projektih. Vendar pa opažamo, da se skupnost, tudi ob spodbudi financerjev, vedno bolj zaveda pomena koncepta odprte znanosti, kar bo dolgoročno omogočilo, da bomo do novih biotehnoloških odkritij prišli hitreje in ceneje.

Literatura

Baebler, Špela, Miha Svalina, Marko Petek, Katja Stare, Ana Rotter, Maruša Pompe-Novak in Kristina Gruden (2017): QuantGenius: Implementation of a Decision Support System For Qpcr-Based Gene Quantification. *BMC Bioinformatics* 18(1): 276. Dostopno na DOI: 10.1186/s12859-017-1688-7.

Bustin, Stephen A.(2008): RT-qPCR and Molecular Diagnostics: No Evidence for Measles Virus in the GI Tract of Autistic Children. *European Pharmaceutical Review* 1: 11–17. Dostopno na: https://www.badscience.net/wp-content/uploads/erp_mmr.pdf (23. oktober 2020).

Bustin, Stephen A., Vladimir Benes, Jeremy A. Garson, Jan Hellemans, Jim Huggett, Mikael Kubista, Reinhold Mueller, Tania Nolan, Michael W. Pfaffl, Gregory L. Shipley, idr. (2009): The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clinical Chemistry* 55(4): 611 – 622. Dostopno na DOI: 10.1373/clinchem.2008.112797.

Carroll, Michael W. (2015): Sharing Research Data and Intellectual Property Law: A Primer. *PLoS Biology* 13(8): e1002235. Dostopno na DOI: 10.1371/journal.pbio.1002235.

Evropska komisija (2020): *Data Management – H2020 Online Manual.* Dostopno na: ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm (23. oktober 2020).

Hollmann, Susanne, Andreas Kremer, Špela Baebler, Christophe Trefois, Kristina Gruden, Witold R. Rudnicki, Weida Tong, Aleksandra Gruca, Erik Bongcam-Rudloff, Chris T. Evelo, idr. (2020): The Need for Standardisation in Life Science Research – An Approach to Excellence and Trust. *F1000Research* 9: 1398. Dostopno na DOI: 10.12688/f1000research.27500.1.

Patron, Nicola J., Diego Orzaez, Sylvestre Marillonnet, Heribert Warzecha, Colette Matthewman, Mark Youles, Oleg Raitskin, Aymeric Leveau, Gemma Farré, Christian Rogers, idr. (2015): Standards for Plant Synthetic Biology: A Common Syntax for Exchange of DNA Parts. *New Phytologist* 208(1): 13–19. Dostopno na DOI: 10.1111/nph.13532.

Sansone, Susanna-Assunta, Peter Mcquilton, Philippe Rocca-Serra, Alejandra Gonzalez-Beltran, Massimiliano Izzo, Allyson L. Lister, Milo Thurston in The Fairsharing Community (2019): FAIRsharing as a Community Approach To Standards, Repositories and Policies. *Nature Biotechnology* 37(4): 358–367. Dostopno na DOI: 10.1038/s41587-019-0080-8.

The Gene Ontology Consortium (2019): The Gene Ontology Resource: 20 years and Still GOing Strong. *Nucleic Acids Research* 47(D1): D330–D338. Dostopno na DOI: 10.1093/nar/gky1055.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, idr. (2016): The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* 3:160018. Dostopno na DOI: 10.1038/sdata.2016.18.