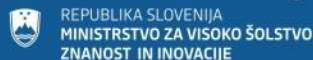


Objavljanje drugih rezultatov raziskovalnega dela

dr. Uroš KUNAVER, Centralna tehniška knjižnica Univerze v Ljubljani

mag. Miro PUŠNIK, Centralna tehniška knjižnica Univerze v Ljubljani

Ljubljana in online, Centralna tehniška knjižnica Univerze v Ljubljani,
Usposabljanje podatkovnih strokovnjakov, 14. – 17. 10. 2024





Kaj so drugi rezultati raziskovalnega dela?

Digitalni:

- programska oprema,
- delovni postopki in protokoli,
- modeli,
- formati, metapodatkovne sheme, geslovniki, šifranti, tezavri, ontologije...

Fizični:

- novi materiali,
- protitelesa, reagenti, vzorci,
- standardi (v kemiji).



Zakaj jih je potrebno objavljati?

- Ker so v skladu s politikami OZ izenačeni z raziskovalnimi podatki.
- Ker financerji tako zahtevajo (bodo zahtevali).

Na primer: projekti Obzorje Evropa v navodilih za pripravo NRRP zahtevajo, da navedemo, katere druge rezultate raziskovalnega dela bomo (poleg podatkov) ustvarili in v kakšni obliki (v skladu z načeli FAIR) jih bomo objavili:

<https://enspire.science/wp-content/uploads/2021/09/Horizon-Europe-Data-Management-Plan-Template.pdf>

3. Other research outputs

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.



Vzporednice z ravnanjem z raziskovalnimi podatki

- Za nekatere vrste drugih rezultatov raziskav (predvsem za programsko opremo) se lahko zahteva priprava načrta ravnanja z njimi.
- Zaželeno je, da so tudi drugi rezultati raziskav opremljeni s PID-om (npr. z DOI).
- Tudi druge rezultate raziskav moramo opremiti z ustreznimi metapodatki in opisati z ustreznimi izrazi.
- Tudi drugim rezultatom raziskav je potrebno dodati opis v obliki posebne priloge - v obliki datoteke »preberi me« oziroma opisa provenience.
- Tudi drugi rezultati raziskav morajo biti opremljeni z ustrezno licenco.
- Tudi za druge rezultate raziskave veljajo enaki varnostni in etični vidiki kot za raziskovalne podatke.



Načrt ravnanja s programsko opremo

Če je programska oprema pomemben ali celo glavni rezultat raziskave, je smiselno razmisliti o izdelavi načrta ravnanja s programsko opremo (ang. software management plan). Trenutno je to zelo redko zahteva financerjev, lahko pa se v prihodnosti to spremeni. Kako pripraviti načrt ravnanja s programsko opremo je lepo opisano na strani **Software Sustainability Institute (SSI): Writing and using a software management plan** ().

Načrt ravnanja s programsko opremo je smiselno izdelati predvsem, če:

- pri projektu sodeluje več ljudi z različnimi vlogami/zadolžitvami,
- pri projektu sodeluje več različnih raziskovalnih inštitucij,
- je projekt časovno razdeljen na več podprojektov, ki so odvisni en od drugega,
- gre za del večjega projekta, ki se bo nadaljeval tudi po koncu naše raziskave.



Trajni identifikatorji za druge rezultate raziskav

Programska oprema - pregled trajnih identifikatorjev za programsko opremo (in repozitorijev zanjo) najdemo na straneh organizacije **Better Scientific Software (BSSw)**: <https://bssw.io/items/persistent-identifiers-for-software-in-scientific-computing>:

- **GitHub** nudi integracijo z repozitorijem **Zenodo**, ki omogoča trajno hrambo določene verzije programske opreme opremljene z DOI (<https://guides.github.com/activities/citable-code/>) – več o tem in **GitLab** repozitoriju v tretjem sklopu predavanj.
- **FigShare** je splošen repozitorij za raziskovalne podatke in programsko kodo, ki ponuja možnost dodeljevanja DOI (<https://figshare.com/>).
- **Software heritage** (<https://www.softwareheritage.org/>) je repozitorij za arhiviranje programske kode. Uporablja poseben PID imenovan SoftWare Heritage persistent Identifier (SWHID) <https://docs.softwareheritage.org/devel/swh-model/persistent-identifiers.html>.
- **DOECode** je repozitorij zasnovan za programsko opremo v okviru DOE projektov (<https://www.osti.gov/doecode/>). DOECode uporablja DOI.



Trajni identifikatorji za druge rezultate raziskav 2

- **Fizični vzorci in primerki: International Generic Sample Number (IGSN)** (<https://igsn.github.io/overview/>) – gre v bistvu za posebno obliko DOI, ki ga pridobimo s pomočjo agenta (npr. DataCite) – več o tem: <https://support.datacite.org/docs/igsn-id-registration-guide>.
- **Laboratorijski material v biomedicini** (npr. plazmidi, celice, protitelesa, organizmi ipd.): **Research Resource Identifier (RRID)** (<https://www.rrids.org/>).
- **Merilni inštrumenti: Persistent Identifier of Instruments (PIDINST)** <https://docs.pidinst.org/en/latest/>, ki ga upravlja delovna skupina RDA za PIDe inštrumentov (<https://www.rd-alliance.org/groups/persistent-identification-instruments-wg/>). Več o tem v članku <https://datascience.codata.org/articles/10.5334/dsj-2020-018>.



Metapodatki za druge rezultate raziskav

- Večinoma nam minimalni/obvezni metapodatkovni standard določi kar **repozitorij**, v katerem želimo objaviti druge rezultate raziskav.

Zato je dobro že pred začetkom raziskovanja ali vsaj pred začetkom meritev načrtovati, katere metapodatke bomo potrebovali.

- Nekateri **splošne metapodatkovne sheme** nudijo podporo (možnost vnosa) metapodatkov za druge rezultate raziskav.

Primer: metapodatkovna shema **DataCite** (<https://datacite-metadata-schema.readthedocs.io/en/4.5/>) v polju ResourceType podpira programsko opremo, modele, fizične objekte, storitve, standarde, postopke...

DataCite Metadata Properties

Overview

1. Identifier

2. Creator

3. Title

4. Publisher

5. PublicationYear

6. Subject

7. Contributor

8. Date

9. Language

10. ResourceType

11. AlternateIdentifier

12. RelatedIdentifier

13. Size

14. Format

15. Version

16. Rights

17. Description

18. GeoLocation

19. FundingReference

20. RelatedItem

Appendices

Mappings

Guidance

XML Schema and Examples

- [Audiovisual](#)
- [Book](#)
- [BookChapter](#)
- [Collection](#)
- [ComputationalNotebook](#)
- [ConferencePaper](#)
- [ConferenceProceeding](#)
- [DataPaper](#)
- [Dataset](#)
- [Dissertation](#)
- [Event](#)
- [Image](#)
- [InteractiveResource](#)
- [Instrument](#)
- [Journal](#)
- [JournalArticle](#)
- [Model](#)
- [OutputManagementPlan](#)
- [PeerReview](#)
- [PhysicalObject](#)
- [Preprint](#)
- [Report](#)
- [Service](#)
- [Software](#)
- [Sound](#)
- [Standard](#)
- [StudyRegistration](#)
- [Text](#)
- [Workflow](#)
- [Other](#)



Metapodatki za programsko opremo

Metapodatki poleg splošnih podatkov kot so avtor, ime programske opreme, datum izdelave/objave, licenca vsebujejo še podatke kot so: programski jezik, ciljni operacijski sistem, vrsta uporabniškega vmesnika, združljivost, verzija, uporabljeni podatkovni modeli, način namestitve, informacije o varnosti, varnostnih kopijah, performančnih lastnostih...

Primeri metapodatkovnih shem za nekatere prej opisane repozitorije programske opreme:

- Software Heritage: <https://docs.softwareheritage.org/devel/architecture/metadata.html>.
- DOECode: <https://github.com/doecode/software-metadata>.

Nabor možnih metapodatkov za opisovanje programske opreme je opisan na straneh podjetja LeanIX: <https://www.leanix.net/en/wiki/trm/software-metadata>.



Razširjen nabor metapodatkov in provenienca

Kadar metapodatkovna shema repozitorija, kamor odložimo rezultate raziskave ne zadošča za celovit opis naših rezultatov, moramo te opisati posebej. Običajno ta opis shranimo v obliki „preberi me/read me“ datoteke ali pa sledimo kakšnemu od standardov za provenienco.

Pomembno je, da bo lahko nekdo, ki je **strokovnjak na določenem področju**, iz naših metapodatkov (iz repozitorija in/ali iz „preberi me“ datoteke) **povsem razumel**, kako smo do objavljenih rezultatov prišli.

- Pri programski opremi se lahko pri izbiri potrebnih metapodatkov zgledujemo po metapodatkovnih shemah, omenjenih na prejšnjih prosojnicah.
- Mnogokrat je pomembno, da opišemo zgodovino razvoja in poreklo programske opreme, metapodatkovne sheme, ontologije, delovnega postopka...
- Pri fizičnih rezultatih je pomembno opisati celoten postopek izdelave/pridobivanja (surovine, naprave, pogoji...).



Klasifikacijski sistemi, terminologija in ontologije

Programska oprema – za razliko od ostalih področij in vrst rezultatov je za področje programske opreme na voljo ogromno različnih **geslovnikov** – vprašanje je, katerega bomo izbrali za opisovanje naše programske opreme (velja pogledati, če kakšen geslovník ponuja/zahteva izbrani repozitorij). Nekaj primerov:

- **Wikipedija** ponuja sezname geslovnikov, akronimov... in lastnih strani na temo računalništva (https://en.wikipedia.org/wiki/Category:Computing_terminology), med drugim tudi lasten **Glossary of computer science** (https://en.wikipedia.org/wiki/Glossary_of_computer_science).
- Precej obsežen geslovník je na straneh **Dixons Fazakerley Academy** <https://www.dixonsfa.com/uploads/images/J277-Glossary-Booklet.pdf>.
- V Sloveniji imamo obsežen, vendar rahlo zastarel **Leksikon računalništva in informatike** iz leta 2002 (zapis v COBISS: <https://plus.cobiss.net/cobiss/si/sl/bib/117254144>).

Computer Science Ontology (CSO) je avtomatsko generirana ontologija na podlagi 16 milijonov publikacij. Rezultate prikaže v grafični obliki (<https://cso.kmi.open.ac.uk>).



Klasifikacijski sistemi, terminologija in ontologije 2

Za **fizične objekte** (materiali, vzorci, reagenti...) uporabljamo enake klasifikacijske sisteme, terminologije in/ali ontologije, kot za opisovanje raziskovalnih podatkov na področjih iz katerih izhajajo - npr. na področju kemije IUPAC nomenklatura, na področju biologije biološko taksonomijo, na področju medicine tezaver MASH in podobno.

Podobno velja tudi za **postopke, modele** in podobne rezultate raziskav.



Licenciranje drugih rezultatov raziskav

Drugi rezultati raziskav, ki so avtorska dela (npr. programska oprema) morajo biti opremljeni z ustrezno licenco.

- Programska oprema je glede licenciranja zelo specifična (o tem je bilo v četrtem dnevu posebno predavanje), zato zanjo ne uporabljamo Creative Commons licenc, ampak posebne licence, namenjene prav programski opremi.
- Preostale rezultate raziskav (če so avtorska dela – npr. geslovniki, slovarji, ontologije...) licenciramo z CC BY 4.0 ali CC0. CC0 je najprimernejša za rezultate kot so metapodatkovne sheme, ontologije, geslovniki, postopki... saj bistveno olajša kasnejši razvoj in dopolnjevanje teh objektov. Če bi uporabili CC BY licenco, bi morali ob uporabi teh objektov navesti vse, ki so kadarkoli kaj prispevali k njihovem razvoju.
- Metapodatke objavljamo pod licenco CC0.



Centralna tehniška knjižnica Univerze v Ljubljani
Central Technical Library at the University of Ljubljana
Trg republike 3, SI-1000 Ljubljana
Slovenija / *Slovenia*



Univerza v Ljubljani

