

Plan upravljanja istraživačkim podacima - HRZZ-2020-02-9666 - iTMD

Popčević, Petar

Data management plan / Plan upravljanja istraživačkim podacima

Publication year / Godina izdavanja: **2023**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:253:539292>

Rights / Prava: [Attribution-NonCommercial-ShareAlike 4.0 International](#)/[Imenovanje-Nekomercijalno-Dijeli pod istim uvjetima 4.0 međunarodna](#)

Download date / Datum preuzimanja: **2024-12-18**



Repository / Repozitorij:

[IF Repository](#)

Plan upravljanja istraživačkim podacima

Opće informacije		
	Ime i prezime predlagatelja	Petar Popčević
	Matična organizacija	Institut za fiziku
	Naziv projekta	Interkalirani dihalogenidi prijelaznih metala
	Upravitelj podacima	Petar Popčević, ppopcevic@ifs.hr
1.	Prikupljanje podataka i dokumentacija	
	Koje ćete podatke prikupljati, obrađivati, stvarati ili se ponovno njima koristiti? (navedite format, vrstu i opseg podataka)	<p>Podaci izmjereni, analizirani i objavljeni unutar ovog projekta odnose se na materijale, uzorke, instrumentalne tehnike i slične alate u području fizike kondenzirane tvari. Većina podataka opisuje specifičnosti pojedinačnih uzoraka određenih materijala, kao što je, na primjer, $\text{Co}_{1/3}\text{NbS}_2$. "Uzorak" ovdje označava određenu (malu) količinu materijala, koja može biti jedan komad kristala ili, primjerice, određena količina praha. Dodatni podaci obuhvaćaju softverski kod za analizu podataka namijenjen za ponovnu uporabu, podatke generirane simulacijama zajedno s kodom simulacije, CAD modele posebno dizajniranog mjernog instrumenta, dnevnik s detaljima postupaka i slično.</p> <p>Dijelovi podataka prikupljaju se mjerenjem specifičnih karakteristika materijala, kao što su njegova električna otpornost, magnetska susceptibilnost, toplinska vodljivost, specifična toplina i dr., u raznim uvjetima. Ostali podaci vezani za pojedini uzorak mogu uključivati težinu uzorka, slike uzoraka u različitim fazama pripreme za eksperiment. Također, podaci o mjernim sekvencama, koje definiraju što mjerni softver treba raditi, bit će pohranjeni za odgovarajuća mjerenja.</p> <p>Većina podataka specifičnih za uzorak su ASCII tekstualne datoteke koje sadrže numeričke podatke, prikupljene i spremljene odgovarajućim mjernim softverom. To uključuje PPMS i MPMS sustave od QuantumDesign-a, sa softverom "MultiVu", kao i postave razvijene u laboratorije, s mjernim programima napisanima korištenjem softvera LabViewu ili Basic. Ekstenzije imena datoteka koje se koriste za ASCII datoteke su TXT, CSV, DAT, jer su to one koje MultiVu i drugi komercijalni softver proizvode - također, to su općenito vrste datoteka za mjerenje podataka u znanstvenoj zajednici fizike kondenzirane tvari, gdje se podaci gotovo uvijek odnose na jedan određeni uzorak. Ovi formati su otvoreno čitljivi i obično sadrže metapodatke eksperimenta (npr. verzija softvera, ime uzorka, masa uzorka ...). Također, lako se uvoze i koriste s mnogim analitičkim programima poput "Origin", "Mathematica" i "MATLAB". Dodatne ekstenzije ASCII tekstualnih datoteka mogu biti SEQ i JSON za definiciju mjernih sekvenci. CAD modeli bit će arhivirani koristeći otvoreni format STEP.</p> <p>Slike uzoraka u pripremi snimaju se kamerama mikroskopska u laboratorijima, uglavnom kao PNG ili JPEG datoteke, ili kamerama pametnih telefona, u formatu u kojem ih snimaju.</p>

		<p>Neće postojati ograničenja za ponovnu uporabu postojećih podataka osim znanstvenih razloga koji se odnose na integritet odgovarajućih postojećih podataka ili izvedivost njihove uporabe u prvom redu, tj. ako su primjenjivi postojeći podaci iz literature dostupni, mogli bi se koristiti.</p> <p>Većina eksperimenata (1 mjerenje 1 uzorka) generira samo male količine numeričkih podataka (ispod 50 MB)</p> <p>Također, na jednom uzorku može se izvršiti više mjerenja. Dodatno, mjerenja s višedimenzionalnom prostornom/lokalnom/lateralnom/momentalnom rezolucijom (spektroskopski podaci, 2D površinski ili 3D volumetrijski podaci uzoraka izmjereni primjerice pomoću ARPES-a, RIXS-a, STM-a, TEM-a, AFM-a) generiraju puno veću količinu podataka, obično množeći broj u rasponu od 50 MB s odgovarajućom rezolucijom.</p> <p>Ovisno o tome koliko je uzoraka dostupno, očekujemo oko 100-1000 mjerenja. Za jednostavna mjerenja ---- otpornosti---- to bi generiralo ukupno oko ----30 GB---- sirovih podataka iz eksperimentalnih tehnika. ----5-6---- posjeta korisničkim centrima (sinkrotron, neutronske reaktore), svaki od kojih bi mogao generirati podatke na razini od ----40 GB----, dodalo bi oko ----200 GB----. Simulacije koje se izvode s blago varirajućim parametrima sigurno su izvor velikih skupova podataka, gdje bi jedan projekt simulacije mogao dodati od ----100 do 400 GB----. To rezultira ukupno oko ----700 GB---- sirovih podataka.</p>
	<p>Kako će se podaci prikupljati, obrađivati ili stvarati? (ukratko navedite metodologiju i procese osiguranja kvalitete te načine organiziranja podataka)</p>	<p>Općenito, pridržavamo se dobre znanstvene prakse kako bi se osigurala visoka kvaliteta istraživačkih podataka - gdje god je to moguće, mjeri se više uzoraka za određeni efekt koji će biti objavljen, itd. Programi za prikupljanje podataka napisani od korisnika treba prekontrolirati barem još jedna osoba osim onoga tko je program napisao. Kontrola prikupljenih podataka provest će se prema principu četvero očiju: mlađi suradnik koji obavlja mjerenja i stariji član tima koji pregledava cijeli proces, uključujući prikupljanje i analizu podataka.</p> <p>Podaci su kategorizirani prema načinu generiranja: kod, rezultati simulacije, eksperimenti malih razmjera u laboratoriju, posjete korisničkim postrojenjima. Općenito, struktura mapa bit će organizirana na sljedeći način, ako je primjenjivo: ciljna publikacija, materijal, tehnika mjerenja, datum provedbe (npr. određeni tjedan/dan mjerenja), podaci vezani za određenu sekvencu, uzorak. Na primjer, svi podaci vezani za posjetu korisničkim objektima općenito će se pohranjivati i arhivirati zajedno (na razini općeg datuma provedbe), a za rezultate simulacija ista razina može se koristiti za različite verzije koda, dok razina uzoraka može predstavljati drugi skup parametara.</p> <p>Dokumentacija za ponovnu uporabu uključuje skeniranje ručno pisanih dnevnika mjerenja gdje postoje, a dodaju se podacima na razini "općeg datuma provedbe". Gdje ručno pisani dnevnici ne postoje vodit će se digitalno (npr. tablice programa Excel koje su uređene po datumu u redovima ili tablet za digitalno pohranjivanje ručno pisanih bilješki). Digitalni dnevnici trebali bi se dodati podacima na kraju "izvođenja". Podaci uneseni u takve dnevnike trebali bi slijediti dobru znanstvenu praksu, gdje sve što je potrebno za razumijevanje onoga što je učinjeno treba biti prisutno.</p> <p>Meta podaci i ključne riječi bit će dodane nakon završetka skupa mjerenja (obično jednog "izvođenja" ili nekoliko, za određenu tehniku mjerenja), najkasnije u vrijeme objave. Za prilagođene razvijene postavke podaci će se pohraniti na PC koji kontrolira odgovarajući postav, organizirani po direktorijima prema uzorcima i nazvani tako da je datum izvođenja očit, a svaki naziv je jedinstven i detalji su prepoznatljivi iz dnevnika. Za instrumente poput PPMS-a,</p>

	<p>MPMS-a, podaci su organizirani u mapama za svakog korisnika, i podmapama za svaku kategoriju i grupu uzoraka. Ime datoteke je jedinstveno i spomenuto je u dnevniku koji se vodi za uređaj Općenito, treba se pridržavati dobre znanstvene prakse kako bi se osigurala visoka kvaliteta istraživačkih podataka - gdje god je to moguće, trebalo bi mjeriti više uzoraka za određeni pothvat koji će biti objavljen, itd. Postupak prikupljanja podataka pomoću alata za softversko pisanje, koji su prilagođeni, trebao bi pregledati barem jedna osoba osim prvobitnog razvojnog programera softvera, a recenzija podataka provest će se prema principu četiri očiju između člana juniorskog tima koji obavlja mjerenja i člana starijeg tima koji pregledava cijeli proces, uključujući prikupljanje i analizu podataka. Podaci su kategorizirani prema načinu generiranja: kod, rezultati simulacije, eksperimenti malih razmjera u laboratoriju, posjete objektima. Općenito, struktura mapa bit će organizirana na sljedećim razinama, ako je primjenjivo: ciljna publikacija, materijal, tehnika mjerenja, opći datum provedbe (npr. određeni tjedan mjerenja), podaci naspram sekvenci, uzorak. Na primjer, svi podaci o posjetima objektima općenito će se pohranjivati i arhivirati zajedno (na razini općeg datuma provedbe), a za rezultate simulacija ista razina može se koristiti za različite verzije koda, dok razina uzoraka može predstavljati drugi skup parametara.</p> <p>Dokumentacija za ponovnu uporabu uključuje skeniranje ručno pisanih dnevnika mjerenja gdje postoje, a dodaju se podacima na razini "općeg datuma provedbe". Gdje ručno pisani dnevnici već ne postoje i treba ih zadržati, to bi trebalo napraviti digitalno (npr. tablice programa Excel koje su uređene po datumu u redovima ili tablet za digitalno pohranjivanje ručno pisanih bilješki). Digitalni dnevnici trebali bi se dodati podacima na kraju "izvođenja". Podaci uneseni u takve dnevnike trebali bi slijediti dobru znanstvenu praksu, gdje sve što je potrebno za razumijevanje onoga što je učinjeno treba biti prisutno.</p> <p>Metapodaci i ključne riječi bit će dodane nakon završetka skupa mjerenja (obično jednog "izvođenja" ili nekoliko, za određenu tehniku mjerenja), najkasnije u vrijeme objave. Za prilagođene razvijene postavke podaci će se pohraniti na PC koji kontrolira odgovarajuću postavku, organizirani po direktorijima prema uzorcima i nazvani tako da je datum izvođenja očit, a svaki naziv je jedinstven i detalji su prepoznatljivi iz dnevnika. Za instrumente poput PPMS-a, MPMS-a, podaci su organizirani u mapama za svakog korisnika, i podmapama za svaku kategoriju i grupu uzoraka.</p> <p>Podaci o sintezi prikupljaju se ručno, za vrijeme sinteze, i unose u SQL bazu podataka putem prilagođenog HTML/PHP obrasca koji je smješten lokalno na računaru u laboratoriju za sintezu, osim u analogni dnevnik u koji se unose ručno pisane bilješke. Ulazi su prepoznatljivi po serijskom broju za svaku sintezu, a ovaj serijski broj uvijek se spominje uz bilo koji drugi eksperiment na uzorcima iz ove serije.</p>
<p>Koju ćete dokumentaciju i metapodatke ustupiti osim podataka? (navedite koje su informacije potrebne korisnicima kako bi mogli čitati i interpretirati podatke u budućnosti te koji će se</p>	<p>U ručno pisani dnevnik bit će unesene informacije o vrsti uzorka, seriji, broju, metodi pripreme, kontaktiranju, masi i postupku mjerenja.</p> <p>Slike uzoraka u različitim fazama pripreme snimaju se kamerama mikroskopskih postava u laboratorijima i koristit će za određivanje geometrije uzorka.</p> <p>Što se tiče sinteze materijala, informacije o uzorku materijala koji je sintetiziran unutar projekta uvijek bi trebale uključivati ID rasta/serije iz kojeg dolazi, a koji pak treba sadržavati informacije o početnom sastavu sastojcima (npr.</p>

	standardi koristiti pri tumačenju podataka)	brojevi serije korištenih prahova, ID prekursora), zatim proces korištene sinteze, vrijeme trajanja i sve relevantne detalje potrebne za reprodukciju te iste sinteze. Metapodaci prikupljeni softverom MultiVu definirani su od strane QuantumDesing-a za svaku specifičnu opciju i uključuju sve relevantne informacije za to mjerenje kao zaglavlje u mjernoj datoteci.
2.	Pravna i sigurnosna pitanja	
	Jeste li ograničeni sporazumom o povjerljivosti? Imate li potrebna dopuštenja za prikupljanje, obradu, čuvanje i dijeljenje podataka? Jesu li osobe čiji se podaci pohranjuju informirani o tome i jesu li dali privolu? Kojim ćete se metodama koristiti u svrhu zaštite osjetljivih podataka (GDPR - posebne kategorije osobnih podataka)?	Pri izvedbi ovog projekta neće se kršiti etička načela.
	Kako će se regulirati pristup podacima i njihova sigurnost? Koji su potencijalni rizici koje treba uzeti u obzir? Kako ćete osigurati sigurnost pohrane osjetljivih podataka?	Podaci će se pohraniti u centraliziranom sustavu za pohranu kojim upravlja Odjel za informatiku Instituta za fiziku. Pristup podacima upravlja se preko identiteta ustanove koji je siguran sustav i slijedi najbolje prakse u pogledu upravljanja identitetom. Naš centralni sustav pohranjivanja podataka ima dostatnu zalihost, vrši se zrcaljenje i stalno se nadzire.
	Kako ćete upravljati zaštitom autorskih prava i intelektualnog vlasništva? Tko će biti vlasnik podataka? Koje će se licencije primjenjivati na podatke? Koja će se ograničenja primjenjivati na ponovnu uporabu osobnih podataka?	Ne očekuje se da će rezultati istraživanja dovesti do patenta. Ostali problemi intelektualnog vlasništva će se rješavati prema preporukama Instituta za fiziku. Budući da podaci nisu podvrgnuti ugovoru, te se neće patentirati, objavit će se kao otvoreni podaci pod licencijom Creative Commons-BY.
3.	Pohrana i čuvanje podataka	

<p>Kako će podaci biti pohranjeni i kako će biti napravljena sigurnosna kopija podataka (<i>backup</i>) tijekom istraživanja? Koji su kapaciteti čuvanja podataka kojim raspolazete? Kojim se procedurama koristite za sigurnosnu kopiju (<i>backup</i>)?</p>	<p>Laboratorijski dnevnicu čuvaju se u laboratoriju glavnog istraživača. Elektronički podaci pohranjeni su na računalu odgovarajućeg eksperimentalnog postava, kao i na osobnom računalu odgovarajućeg člana projektnog tima koji ih prikuplja i obrađuje. Osim toga, istraživači se koriste prostorom za pohranu i arhiviranje na institucijskom repozitoriju za sekundarnu pohranu podataka. Grupa glavnog istražitelja ima na raspolaganju 1 TB prostora za pohranu podataka koji se može i dodatno proširiti. Institucionalni repozitorij redovito izrađuje sigurnosne kopije</p>
<p>Koji je vaš plan čuvanja podataka? U kojim će se formatima čuvati?</p>	<p>Podatke ćemo čuvati trajno na institucijskom repozitoriju Instituta za fiziku uspostavljenom na sustavu Dabar. Tablične podatke čuvat ćemo u CSV, TXT ili DAT obliku, a tekstualne u DOCX (Office Open XML), TXT te PDF-A obliku. Slike će se čuvati u JPG, PNG ili BMP obliku.</p>
<p>4. Dijeljenje i ponovna uporaba podataka</p>	
<p>Kako i gdje će se podaci dijeliti? Na kojem repozitoriju planirate dijeliti podatke? Kako će potencijalni korisnici doznati za podatke?</p>	<p>Podaci istraživanja koji su objavljeni u radu putem slika bit će objavljeni kao sirovi podaci, s otvorenim pristupom na Zenodu, u trenutku objavljivanja. To će biti sirovi izmjereni/simulirani podaci, iz kojih su slike sačinjene, a NE samo digitalne reprezentacije određenih točaka/linija na grafu, gdje se više varijabli može međusobno isprepletati. Podaci koji nisu izravno korišteni za generiranje slika za publikacije bit će objavljeni s ograničenim pristupom. To uključuje, primjerice, sirove podatke mjerenja za dodatne uzorke koji nisu uključeni u konačne slike, na primjer ako izgleda da mjerenje određenog uzorka ima neku sustavnu grešku, dok su podaci za uzorke bez ove pogreške, koji predstavljaju intrinzično ponašanje materijala, dostupni. Podaci će biti objavljeni na Zenodu, uz odgovarajuće ključne riječi i sažetak – DOI i referenca na njega bit će navedeni/citirani u odgovarajućoj publikaciji, ako se određena serija podataka odnosi na određenu publikaciju. Licenca za to bit će CC-BY. Za pristup podacima neće biti potreban drugi softver osim standardnih uređivača teksta.</p>
<p>Ako postoje podaci koji se ne smiju dijeliti (prijavitelji vezani zakonskim, etičkim, autorskim pravila, povjerljivošću i sl.), pojasnite razloge ograničenja.</p>	<p>Not all research data will be published openly immediately because a) this would generate a lot of “noise” in regards to available data, not all data which we capture for one material might represent the true nature of this material, and making all of it available might increase the workload of other researchers who try to find relevant data, and because b) we need to keep a certain advantage over competitors to be able to publish our own research findings and attract funding. If collaborations are agreed with external researchers, access to previously restricted data might be granted per individual decisions.</p> <p>Svi istraživački podaci neće odmah biti javno dostupni, jer a) to bi generiralo puno "šuma" u vezi s dostupnim podacima, svi podaci koje bilježimo za jedan materijal možda ne predstavljaju pravu prirodu tog materijala, te bi objavljivanje svih tih podataka moglo povećati radno opterećenje drugih istraživača koji pokušavaju pronaći relevantne podatke, i jer b) moramo zadržati određenu prednost nad konkurentima kako bismo mogli objaviti vlastita istraživačka otkrića i privući financiranje. Ako se dogovore suradnje s vanjskim istraživačima, pristup prethodno ograničenim podacima može biti dodijeljen na temelju pojedinačnih odluka.</p>

	Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima <i>FAIR-a</i> .	Zenodo repozitorij je u skladu s FAIR načelima.
	Potvrdite da ćete se koristiti digitalnim repozitorijem koji održava neprofitna organizacija (ako ne, objasnite zašto ne možete dijeliti podatke na digitalnom repozitoriju koji nije komercijalan).	Zenodo održava neprofitna organizacija.