

Portal otvorenih podataka

Preporuke o prilagodbi skupova podataka za javnu objavu i ponovno korištenje

Verzija 0.6

Naslov:

Preporuke o prilagodbi skupova podataka za javnu objavu i ponovno korištenje

Opis:

Dokument opisuje preporuke o prilagodbi skupova podataka za javnu objavu i ponovno korištenje na Portalu otvorenih podataka data.gov.hr. Definira poželjne formate za objavu, i potrebne metapodatke.

Ključne riječi:

objavljivanje informacija, javna služba, javnost uprave, pristup informacijama, otvoreni podaci, formati otvorenih podataka

Jezik:

hrvatski (HR)

Stvaratelj:

Ministarstvo uprave, Uprava za e-Hrvatsku

Izdavač:

Ministarstvo uprave, Uprava za e-Hrvatsku

Mjesto i datum nastanka:

Zagreb, 9.11.2015.

Izvor:

Ministarstvo uprave, Uprava za e-Hrvatsku

Pripremio:

Zoran Luša – Ministarstvo uprave

Odobrio:

Tomislav Vračić – Ministarstvo uprave

Što su otvoreni podaci?

Otvoreni podaci po samoj definiciji označavaju one koje bilo tko može slobodno koristiti, upotrijebiti za ponovno korištenje i ponovno distribuirati, uz uvjet da se navede izvor podataka i omogući dijeljenje pod sličnim uvjetima. Njihova je bitna značajka da ne sadrže osobne podatke niti druge zakonom zaštićene podatke. **Ukoliko imate dvojbu da li objaviti određeni skup podataka molimo posavjetujte se s Agencijom za zaštitu osobnih podataka.**

Informacije koje se nalaze u javnoj upravi i javnom sektoru u cjelini imaju veliki ekonomski potencijal - direktne i indirektno mogućnosti koju pružaju te informacije procjenjuju se u EU na 40 mlrd EUR godišnje.

Ponovnom uporabom informacije kao što su npr. geografske informacije, statistički podaci, meteorološki podaci, podaci podaci o vremenu, javni registri i slično, mogu se kombinirati iz različitih izvora, kreirati složene i nove aplikacije, kako za komercijalnu tako i za nekomercijalnu uporabu. Time se stvara dodana vrijednost koja može pomoći u stvaranju novih poslova i radnih mjesta. Istodobno, korisnicima omogućava veći izbor i veću vrijednost za novac.

Zašto otvoriti podatke?

Otvaranje podataka, pogotovo onih koji su vezani uz tijela javne vlasti, bitno je jer je riječ o podacima koji su već zakonski definirani kao javni. Istodobno, postoje brojne fizičke i pravne osobe te udruge kojima bi ti podaci olakšali obavljanje određenih zadataka i pomogli u stvaranju novih proizvoda i usluga, novih radnih mjesta te, posredno, jačanju gospodarstva, poboljšanju kvalitete života i novim prihodima koje ostvaruje država od veće gospodarske aktivnosti.

Također, osim transparentnosti i demokratske kontrole, poboljšanih ili novih proizvoda i usluga, poticanja inovativnosti, nužno je napomenuti i da se povećava efikasnost javnih usluga, a kombiniranjem podatkovnih izvora može se stvoriti i novo znanje o određenim činjenicama.

Otvoreni podaci su od velike vrijednosti i samim tijelima javne vlasti. Mogu povećati efikasnost pojedinih tijela te istovremeno im smanjiti troškove.

Ujedno, prava na pristup informacijama i ponovnu uporabu informacija uređena su pravnim regulativama, ponajprije Zakonom o pravu na pristup informacijama – NN 25/13, Direktivom 2003/98/EZ o ponovnoj uporabi informacija javnog sektora¹ te Direktivom 2013/37/EU Europskog parlamenta i Vijeća od 26. lipnja 2013. godine o izmjeni Direktive 2003/98/EZ o ponovnoj uporabi informacija javnog sektora².

Odabir i priprema skupova podataka

¹ <http://ccvista.taiaex.be/show.asp?link=/Fulcrum/CCVista/HR/32003L0098-HR.doc>

² <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:EN:PDF>

Pod kvalitetnim podacima podrazumijevamo one koji su prikladni za uporabu pri odlučivanju i planiranju. Prvo je potrebno odabrati skupove podataka koje planirate učiniti dostupnima. Imajte na umu da biste trebali imati i kontakt s potencijalnim korisnicima podataka. Oni će zasigurno imati dobre ideje o tome koji bi skupovi podataka mogli biti prioritetni.

Da bi skup podataka bio **kvalitetan**, podaci u njemu moraju biti:

- točni, odnosno da predstavljaju točne i stvarne podatke
- dostupni, odnosno da se njima može pristupiti bez dojava o grešci
- potpuni, odnosno da sadrži sve moguće podatke koji su potrebni za označavanje skupa podataka
- konformirani, odnosno da slijede prihvaćene i najkorištenije standarde
- konzistentni, odnosno da ne posjeduju proturječne podatke koji bi otežavali njihovo korištenje
- kredibilni, odnosno da dolaze iz pouzdanih i nadležnih izvora
- procesibilni, odnosno računalno čitljivi
- relevantni, odnosno da uključuju prikladnu količinu podataka
- pravovremeni, odnosno da su objavljeni u dovoljno prikladnom vremenu

Uobičajeni elementi skupova podataka:

- pružiti prikladan opis podataka (metapodatke)
- korištenje standardnih rječnika za metapodatke i podatke gdje god takvi rječnici postoje (uz W3C preporuke)
- određivanje licence za ponovno korištenje podataka
- pridržavanje zakonskih zahtjeva koji se odnose na zaštitu osobnih i ostalih osjetljivih podataka
- metapodatke i podatke predstaviti po mogućnosti po principima povezanih podataka koristeći URI-eve za identifikaciju
- pružiti informacije o izvoru podataka

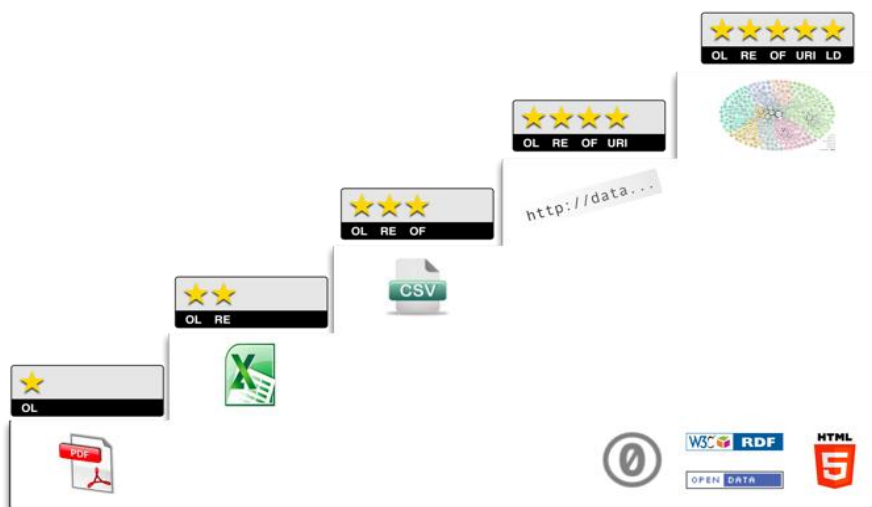
Najbolje prakse pri objavi povezanih podataka dostupni su na <https://dvcs.w3.org/hg/gld/raw-file/default/bp/index.html>

Odabir formata skupa podataka za ponovno korištenje i računalnu obradu

Skupovi podataka, odnosno dokumenti za objavu trebali bi se objaviti u formatima koji opisuju tip podataka koji se objavljuje. Poželjno je da format podataka zadovoljava **barem razinu ocjene od 3**

zvjezdice, sukladno Tim Berners-Lee-jevoj skali '5 zvjezdica otvorenih podataka' - <http://5stardata.info/hr/>

- **Za pisane izvještaje**, HTML format, pored čistog teksta (TXT) bi trebao biti prva i najbolja opcija. PDF/A može biti izvrstan format za prikaz, ali može biti neprikladan za korisnike koji koriste npr. softver za čitanje teksta. Ukoliko baš morate objaviti PDF dokument, svakako razmislite i o alternativnim formatima za objavu. Dokument u DOC formatu ne preporučamo za objavu jer je riječ o zatvorenom formatu. Ukoliko ga baš morate objaviti, uz njega svakako objavite i verziju u formatu otvorenog standarda - HTML, TXT ili XML.
- **Za podatke** preporučamo korištenje formata otvorenog standarda CSV (kodiranje Unicode UTF-8), TSV, a ukoliko je to moguće i JSON, XML te RDF, odnosno neki od oblika koji podržavaju povezljive (linked) podatke uz pomoć URI-ja. Za objavu strukturiranih podataka ne preporučamo objavu isključivo i jedino PDF/A formatu. Dokument u XLS formatu ne preporučamo za objavu jer je riječ o zatvorenom formatu. Ukoliko ga baš morate objaviti ili mislite da će biti prikladan za neke korisnike, uz njega svakako objavite i verziju u CSV, TSV, JSON, XML ili RDF formatu.
- Ukoliko koristite formate JSON i RDF/XML, metapodatke u tim skupovima podataka opisujte **koristeći W3C standarde uz pomoć DCAT-AP** koji se standardizirano koristi za opisivanje skupova podataka javnog sektora u Europi.
- Ukoliko objavljujete pisani izvještaj koji sadrži statističke tablice, preporučamo da se one objave u dodatku izvještaja u prikladnom formatu za strukturirane podatke.
- Kad god je moguće, koristiti otvoreni format - http://en.wikipedia.org/wiki/Open_format
- Ukoliko je to moguće, skupovi podataka mogu se objavljivati i u više dostupnih formata



Slika 1 - Primjeri za svaku razinu ocjene 5 zvjezdica otvorenih podataka; izvor: <http://5stardata.info/hr/>

Nužni metapodaci potrebni za objavu na Portalu otvorenih podataka Republike Hrvatske

Pored samih podataka koji se objavljuju na vlastitim internetskim stranicama, za katalogiziranje u Portal otvorenih podataka Republike Hrvatske potrebno je pripremiti metapodatke, odnosno opise svakog skupa podataka. U metapodacima preporučujemo da se nađu sljedeće informacije:

- Ime skupa podataka (npr. Plan dodjele frekvencija u Republici Hrvatskoj)
- Identifikator (npr. plan_dodjele_frekvencija_u_rh)
- Naziv datoteke – URL – format
- Opis skupa podataka – dodati opis skupa podataka, što sadrže podaci, zašto su nastali, kako su strukturirani (što sadrže pojedini stupci itd...). Poželjno je da opis počne s kratkom rečenica koja opisuje skup podataka, ponajviše jer prvih nekoliko riječi se može prikazivati u skraćenim pregledima skupova podataka.
- Vrsta licence – (preporučena je Creative Commons Attribution licenca)
- Tema skupa podataka (prema pojmovniku [EUROVOC](#), dostupnim na Središnjem katalogu Digitalnog informacijsko-dokumentacijskog ureda)
- 'tagovi', odnosno ključne riječi za skup podataka
- Dodatni resursi – moguće je dodati poveznice na dodatne resurse, primjerice, može biti riječ o PDF datoteci koja dodatno opisuje podatke, ili poveznici na mrežno mjesto na kojem se može naći više informacija.
- Vremenski period na koji se odnosi skup podataka (ili stanje na dan...)
- Područje na koje se odnosi skup podataka (cijela HR, dio HR – do razine županije, EU, van EU – ostatak svijeta).