

Nadgradnja portala NIO

-

Vzdrževanje in nadgradnja portala NIO NMVNIO-3/2015

Naročnik	Ministrstvo za javno upravo
Oznaka dokumenta:	
Status:	Končna različica
Datum verzije:	14.12.2015
Avtor:	XLAB d.o.o.
Stornja varovanja podatkov:	Javno

Zgodovina dokumenta

Datum	Verzija	Opis	Avtor
10.11.2015	0.1	Prva različica dokumenta	Aleš Černivec
2.12.2015	0.2	Dopolnitev	Aleš Černivec
14.12.2015	0.3	Dodal podrobnosti podpoglavja	Aleš Černivec

Vsebina

1. Uvod.....	4
2. Aktivnosti nadgradnje	4
2.1 Generiranje ustrezne izmenjevalne datoteke v formatu ADMS.....	4
2.2 Zajem metapodatkov interoperabilnostnih izdelkov nad OpenScience.si in GURS	5
2.3 Generiranje RDF datoteke v skladu z ADMS.....	6
2.4 Zajem potrebnih metapodatkov za opis naborov odprtih podatkov v skladu s standardom s standardom DCAT Application Profile for data portals in Europe	6
2.5 Označevanje semantičnih izdelkov za objavo na drugih portalih	6
2.6 Označevanje naborov podatkov za objavo na drugih portalih.....	6
2.7 Podpora NIO za objavljanje zunanjih podatkovnih zbirk	6
2.8 NIFO vizualizacija	7
2.9 Analitika na osnovi odprtokodnih orodij (Piwik).....	8
2.10 Podpora iskanju v NIO na osnovi ElasticSearch.....	8
3. Reference	9

1. Uvod

Namen Portala NIO je celovito predstavljati slovenski nacionalni interoperabilnostni okvir. EU zahteva da je NIO skladen z evropskim interoperabilnostnim okvirom – EIF.

Portal NIO predstavlja osnovno orodje za objavlanje, spremljanje in uporabo različnih interoperabilnostnih izdelkov (tehničnih, semantičnih, organizacijskih). Deluje kot stičišče prijaviteljev in uporabnikov interoperabilnostnih izdelkov, ki so potrebni za usklajen razvoj, vzpostavljanje in vzdrževanje informacijskih sistemov v javni upravi ter širše. Portal predstavlja tudi osrednjo nacionalno točko za odprte podatke javnega sektorja. Povečana uporaba in posledično povečana diskovna obremenitev z datotekami odprtih podatkov na Portalu NIO zahteva reorganizacijo shranjevanja. Pri vzpostavitvi portala, objava odprtih podatkov in storitev še ni bila aktualna, z dodajanjem novih vsebin pa je potrebno zagotoviti tudi ustrezen sistem za hrambo in objavo datotek odprtih podatkov in podatkov o odprtih storitvah.

Namen tega dokumenta je predstavitev aktivnosti nadgradenj projekta NIO po zadnjih specifikacijah, kot opredeljeno v pogodbi št. **C1711-15-320009** – aktivnosti v okviru **nadgradenj in sprememb informacijskega sistema**.

2. Aktivnosti nadgradnje

Tu podjamo seznam glavnih tehničnih posegov v portal, ki so se zgodili v zadnjih mesecih leta 2015. Izpostavljamo glavne aktivnosti, ki so predstavljali bodisi vpeljavo dodatnih knjižnic oz. integracijo z zunanjo storitvijo.

- Nadgradnja metapodatkov za opis semantičnih interoperabilnostnih izdelkov v skladu s standardom Asset description metadata schema (ADMS)
 - Generiranje ustrezne izmenjevalne datoteke v formatu ADMS
 - Primer zajema metapodatkov interoperabilnostnih izdelkov nad OpenScience.si in GURS
 - Generiranje RDF datoteke v skladu z ADMS
- Zajem potrebnih metapodatkov za opis naborov odprtih podatkov v skladu s standardom s standardom DCAT Application Profile for data portals in Europe
 - Generiranje RDF datoteke v skladu z DCAT
- Označevanje semantičnih izdelkov za objavo na drugih portalih
- Označevanje naborov podatkov za objavo na drugih portalih
- NIFO vizualizacija
- Podpora NIO za objavlanje zunanjih podatkovnih zbirk
 - Uporabniški vmesnik za podporo prenosa velikih podatkov
- Analitika na osnovi odprtokodnih orodij (Piwik)
- Podpora iskanju v NIO
 - ElasticSearch

2.1 Generiranje ustrezne izmenjevalne datoteke v formatu ADMS

Izvoz v ADMS obliko zapisa podpiramo po shemi različice 1.01. S portala JoinUp smo prevzeli XML shemo, po kateri izvozimo podatke za označene izdelke.

Nekateri evropski interoperabilnostni portali delujejo le prek RDF oblike zapisa, zato smo omogočili transformacijo podatkov iz XML oblike v RDF. Oba zapisa sta dosegljiva na določenem URL naslovu, nato pa je odvisno od samega izvoznega portala, katerega bo prevzel.

Trenutna arhitektura omogoča izvoz dokumentov v formatu, ki je primeren za zrcaljenje izdelkov na zunanje storitve, kot je npr. Joinup¹ ali European Data Portal². S pravicami za dostop do datoteke ADMS se zagotovi preko URLja do izvozljive datoteke portala, npr. <http://nio.gov.si/testExport.nio?type=<tip izvoza>>, kjer <tip izvoza> lahko zavzame vrednosti

- *adms-xml* za izvoz semantičnih izdelkov v formatu ADMS in XML
- *adms-rdf* za izvoz semantičnih izdelkov v formatu ADMS in RDF
- *dcat* za izvoz izdelkov v formatu DCAT

2.2 Zajem metapodatkov interoperabilnostnih izdelkov nad OpenScience.si in GURS

Zajem podatkov s portala GURS v DCAT poteka preko aplikacijskega vmesnika GURSa, bolj specifično URLjev, ki omogočajo izvoz vsebine GURS v XML

Primer URLja na GURS:

http://prostor3.gov.si/imps/srv/si/csw?SERVICE=CSW&VERSION=2.0.2&REQUEST=GetRecords&constraintLanguage=CQL_TEXT&typeNames=csw:Record&elementSetName=full&resultType=results&maxRecords=110

Vsi metapodatki, ki jih ima GURS na portalu ustrezajo formatu DCAT. Preslikovanje atributov na strani NIO poteka

1. dc:title → title (slo in ang)
2. seznam dc:subject → keywords (slo in ang)
3. dct:abstract → summary (slo in ang)
4. dc:source → če bi shranjevali AssetBean, bi ustrezali authors, a ker mi tega podatka itak ne exportiramo, mogoče tega ni treba.
5. dc:URI → shranjujemo kot povezave. Najprej poteka shranjevanje povezave, nato vezava na produkt, podobno kot pri priponkah. Običajno jim določimo še tip, ki je 7 (ostalo), s tem se izognemo kompleksnosti. Pri tem so opisi lahko tudi prazni (odvisno od izvornih podatkov).

Primer XML dokumenta, ki je zajet s strani GURS:

```
<csw:Record>
  <dc:identifier>56193d0a-b41f-4305-9450-81835e873c01</dc:identifier>
  <dc:title>Geološka karta IŠKEGA VRŠAJA 1: 25 000 - Razpoklinska cona</dc:title>
  <dc:type>dataset</dc:type>
  <dc:subject>geologija, hidrogeologija</dc:subject>
  <dc:subject>geologija</dc:subject>
  <dc:subject>litostratigrafija</dc:subject>
  <dc:subject>strukturna zgradba</dc:subject>
  <dc:subject>geomorfologija</dc:subject>
  <dc:subject>geofizika</dc:subject>
  <dc:subject>lški vršaj</dc:subject>
  <dc:subject>Slovenija</dc:subject>
  <dc:subject>lg</dc:subject>
  <dc:subject>geologija</dc:subject>
  <dc:subject>INSPIRE</dc:subject>
  <dc:subject>Geologija</dc:subject>
```

¹ <https://joinup.ec.europa.eu/>

² <http://www.europeandataportal.eu/>

```

<dc:subject>Hidrografija</dc:subject>
<dc:subject>geoscientificInformation</dc:subject>
<dct:abstract>Sloj prikazuje cone, kjer je kamnina sistematično razpokana.</dct:abstract>
<dc:rights>copyright</dc:rights>
<dc:language>slv</dc:language>
<dc:language>eng</dc:language>
<dc:source>Razpoklinska cona</dc:source>
<ows:BoundingBox crs="urn:ogc:def:crs:EPSG::MGI 1901 / Slovene National Grid (EPSG:3912)">
  <ows:LowerCorner>14.544376691236401 45.92440651042941</ows:LowerCorner>
  <ows:UpperCorner>14.466026742665456 45.980927081150156</ows:UpperCorner>
</ows:BoundingBox>
<dc:URI protocol="image/png" name="thumbnail">resources.get?id=517&fname=GE_karta_s.png&access=public</dc:URI>
</csw:Record>

```

Ko se podatki z GURS uspešno preberejo in preslikajo v notranje podatkovne strukture NIO, omogočimo izvoz v poljubnem podprtem formatu (XML, RDF, text).

2.3 Generiranje RDF datoteke v skladu z ADMS

Slednje je podprto z internimi podatkovnimi strukturami portala NIO – dodana je bila podpora prelikave zajetih podatkov (v XML) v notranje podatkovne strukture. Ko so podatki v tej obliki zajeti, jih lahko izvozimo v poljubnem tekstovnem formatu, za katerega so predvidene preslikave (npr. bolika RDF).

2.4 Zajem potrebnih metapodatkov za opis naborov odprtih podatkov v skladu s standardom s standardom DCAT Application Profile for data portals in Europe

Izvoz podatkov je omogočen tudi po DCAT-AP specifikaciji različice 1.00. Sam postopek izvoza je enak kot pri izvozu v ADMS obliki, s tem da je tu podprta le RDF oblika zapisa.

DCAT je RDF slovar, ki omogoča interoperabilnost med podatkovnimi katalogi, ki so objavljeni na internetu. Document³ predpisuje podatkovno shemo, ki omogoča vidnost in lažjo med-storitveno izmenjavo podatkov. Za podporo slednjem smo pouporabili interne podatkovne strukture, ki so bile razvite za podporo izvoza izmenjevalnih datotek in zajema datotek iz GURS.

2.5 Označevanje semantičnih izdelkov za objavo na drugih portalih

Portal NIO omogoča označevanje posameznih izdelkov za objavo na drugih portalih, ki so podprti za izvoz podatkov v ADMS oz. DCAT formatih. Glavni urednik ima možnost nastavitve izvoza proti podprtimi zunanji portalih.

2.6 Označevanje naborov podatkov za objavo na drugih portalih

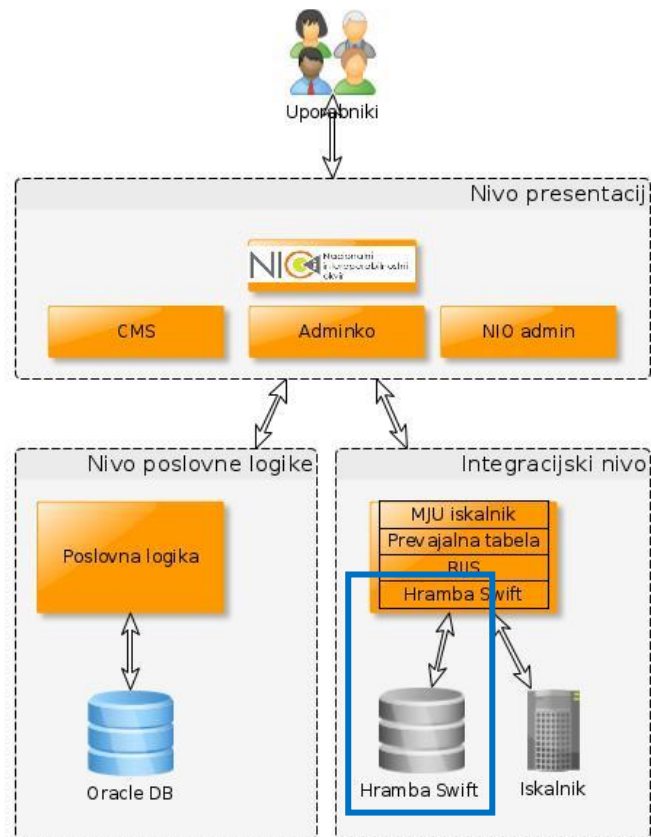
Podobno kakor pri semantičnih izdelkih tudi pri naborih podatkov portal omogoča označevanje le-teh in objavo na podprtih portalih.

2.7 Podpora NIO za objavljanje zunanjih podatkovnih zbirk

NIO podpira uporabo hrambe Swift⁴, arhitekturna slika je vidna na Slika 2-1. Ko uporabnik doda priponko v svoj izdelek, se priponka prenese na podatkovno hrambo Swift. Nastavitev hrambe je narejena preko datoteke nastavitvev, ki se nastavi pred zagonom portala NIO.

³ <http://www.w3.org/TR/vocab-dcat/>

⁴ <http://docs.openstack.org/developer/swift/>



Slika 2-1: Podpora oblaci storitvi za hrambo podatkov.

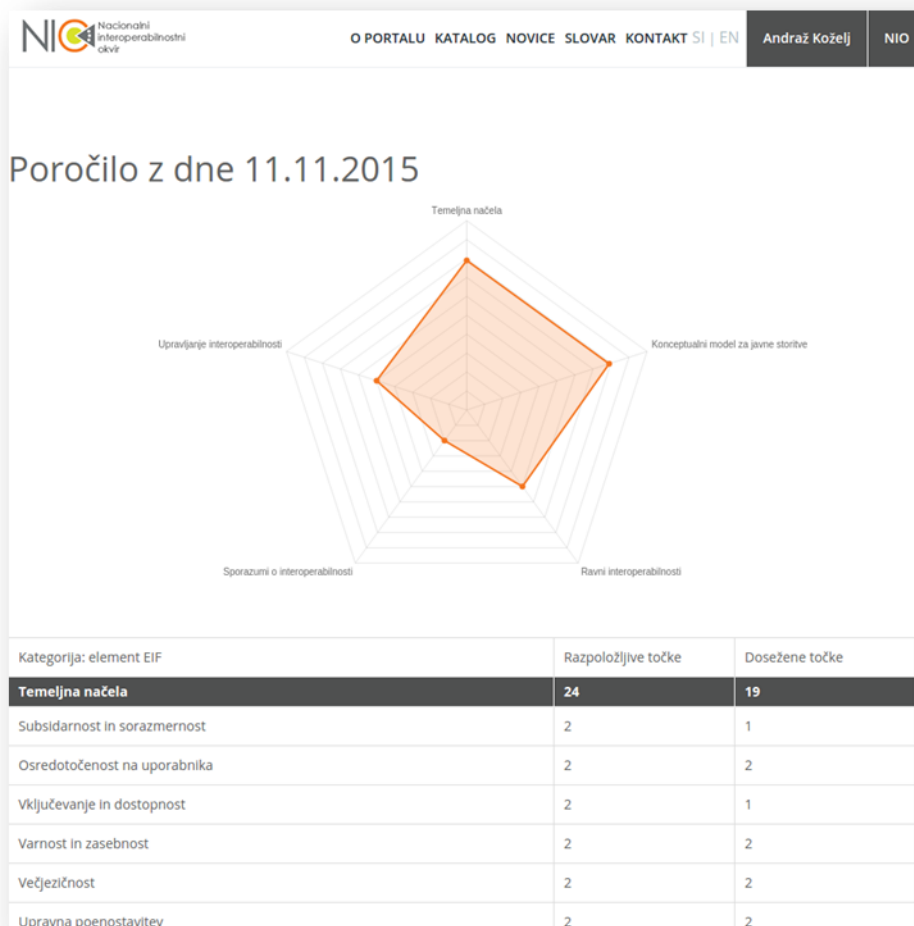
Priponke, ki so hranjene v podatkovni bazi Oracle, se preberejo normalno – na obstoječe priponke nov pristop ne vpliva (združljivost nazaj).

2.8 NIFO vizualizacija

Večkrat letno se za potrebe ocenjevalne lestvice nivoja interoperabilnosti posamične države izpolni anketa, ki poda rezultate doseganja ciljev.

Rezultati se izpolnjujejo ročno, kjer osebe odgovorne za posamične sklope vnašajo ocene. Zbrane ocene se predstavi s tabelo ter »mrežnim grafom«.

Glavni urednik portala lahko pregleduje, popravlja in dodaja rezultate posamičnih anket, na portalu pa se prikaže izris grafa, skupaj z legendo.



Slika 2-2: Primer NIFO vizualizacije.

2.9 Analitika na osnovi odprtokodnih orodij (Piwik)

Za podporo analitiki smo podprli portal s pomočjo orodja Piwik⁵. Gre za odprtokodno orodje, ki je že razmeščeno na infrastrukturi MJU. V portal NIO smo na ključna mesta v CMS vrinili Javascript kodo, ki omogoča sledljivost uporabnikov na portalu.

2.10 Podpora iskanju v NIO na osnovi ElasticSearch

Srce sistema je tekstovni iskalnik Apache Lucene. Dokumenti JSON, ki so v iskalnik poslani s strani portala NIO, so osnovna oblika vsebine za iskanje. Osnovno delovanje iskalnika je hramba, samodejno indeksiranje po vsej atributih dokumentov JSON, in iskanje po le-teh. Iskalnim nam omogoča ustvarjanje specifičnih tipov podatkov, po katerih lahko iščemo, tudi vsebino priponk, če je to potrebno. Ima vmesnik REST za vse operacije (uporaba, konfiguracija, vzdrževanje) in je enostaven za razmestitev in upravljanje/administriranje. Poleg tega ne potrebuje dodatnih zunanjih sistemov (podatkovna zbirka, webserver, ipd). Delovanje v gruči je praktično transparentno za administracijo in uporabo: možna je samodejna povezava več namenskih instanc, kjer lahko izvajamo iskanje/indeksiranje. Glevne značilnosti, ki nam jih ElasticSearch omogoča, so:

- Indeksiranje
- Varnostne kopije
- Porazdelitev bremena

⁵ <http://plugins.piwik.org/>

- Analitiko
- Podpora za močna analičina orodja
- ELK stack (Elastic search – Logstash – Kibana)
- Podpora slovenski lematizaciji (korenjenje, stop-words, ...)
- Enostavno širjenje kapacitet
- Možne kompleksne poizvedbe nad dokumenti
 - <https://www.elastic.co/guide/en/elasticsearch/reference/current/full-text-queries.html>
- Indeksiranje poljubne dodatne vsebine projekta (priponke, novice)

Več podrobnosti o samem ElasticSearch in arhitekturi ELK (ElasticSearch, LogStash, Kibana), lahko najdete tu [2].

3. Reference

[1] NIO 2.0 Arhitekturni dokument, Verzija 1.0

[2] Domača stran Karun Subramanian, <http://karunsubramanian.com/downloads/15-minutes-introduction-to-elk-elasticsearchlogstashkibana/>, dostop 14.12.2015