



#AllForJan hackathon

Predstavenie Open Data

FIIT STU, Bratislava, 5.4.2018
Peter Hanečák



Kto som



- **Peter Hanečák:** zaujímam sa o mnohé otvorené veci
 - programátor (od 1990)
 - nadšený používateľ **Open Source** (od 1998)
 - člen iniciatívy **OpenData.sk** (od 2011)
 - člen **SOIT** (od 2012)
 - člen projektu **Open Data Node** (od 2013)
 - “balíčkováč” v projekte **Fedora** (od 2013)
 - prispievateľ **Slovensko.Digital** (od 2015)

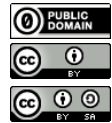


Agenda

- predstavenie Open Data
- prečo je to dobré?
- čo ak údaje nie sú „open“?
- praktické príklady

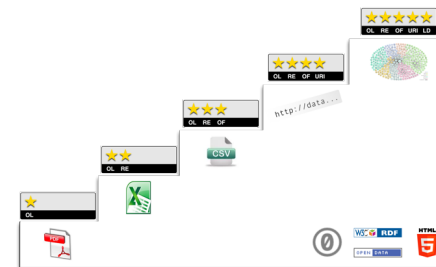


Čo je to Open Data?



hlavné princípy:

- “čo nie je tajné, je verejné” (OpenData.sk)
- “Open means **anyone** can **freely access, use, modify, and share** for **any purpose** (subject, at most, to requirements that preserve provenance and openness).”
(<http://opendefinition.org/>)



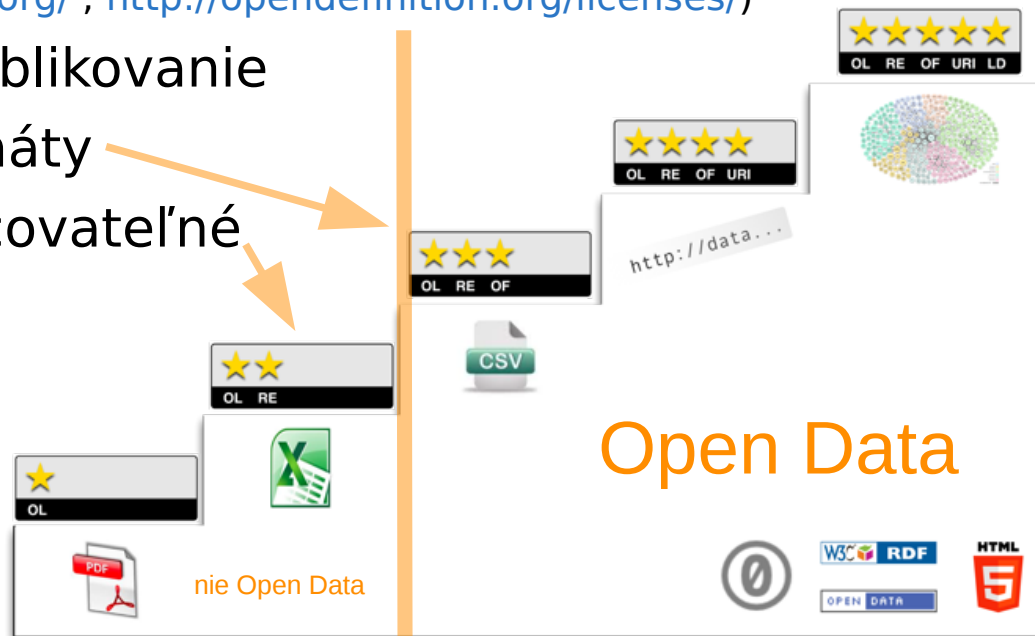
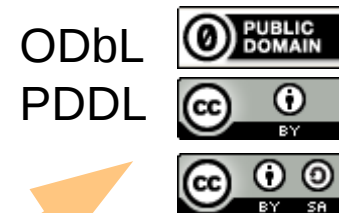


Čo je to Open Data?

hlavné aspekty:

- licencovanie v súlade s Open Definition:
 “Open data and content can be **freely used, modified, and shared** by **anyone** for **any purpose**”
 (<http://opendefinition.org/> , <http://opendefinition.org/licenses/>)

- proaktívne publikovanie
- otvorené formáty
- strojovo spracovateľné



Čo je to Open Data?

hlavné aspekty:

- licencovanie v súlade s Open D
- "Open data and content can be f
- **modified, and shared by anyone**
- (<http://pendamission.org/>), (<https://open.definition.org/licenses/>)
- **proaktívne publikovanie**
- otvorené formáty
- strojovo spracovateľné



šetrí prácu



chráni pred predčasným
vyzradením

<https://www.occrp.org/en/amurderedjournalistslastinvestigation/freedom-of-information-law-reporters-best-friend-or-killer>



Prečo je to dobré?

Open Data je len „enabler“

- vyššia transparentnosť štátu a samosprávy
- lepšia spolupráca občanov a úradníkov
- lepšie alebo nové služby:
 - štátu či samosprávy
 - „3rd party“: firmy, NGO, a pod.
- ekonomické prínosy (z tých služieb)
 - ušetrené výdavky
 - väčší úžitok za tie isté peniaze
 - nové firmy -> vyššia zamestnanosť -> vyšší výber daní, atď.



Prečo je to dobré?

aplikácie, dátová žurnalistika, atď'.





Čo ak údaje nie sú „open“?

scrapre:

- ak sú údaje „na webe“, ale nie na stiahnutie alebo v nevhodnom formáte
- vid' napr.

<https://utopia.sk/wiki/display/opendata/Open+Scraper+Challenge>

infožiadosť (FOIA):

- ak údaje nie sú na webe resp. ak scraper neprichádza do úvahy
- občas pošlú oskenované výtlačky XLS súborov
 - „ go figure ;) “

```

US) | https://github.com/soit-sk/siemens-sk-utilities-scraper/blob/master/scrapper.py
65 # Siemens server too much thus we will try to get data only once a week.
66 last_run = scraperwiki.sqlite_get_var('last_run')
67 if last_run is not None:
68     tmp = datetime.datetime.strptime(last_run, "%Y-%m-%d%H:%M:%S")
69     delta = datetime.datetime.utcnow() - tmp
70     if (delta.days < 7):
71         print '### data is quite fresh (%s), skipping run' % last_run
72         sys.exit()
73
74 # read in a data from their Ajax back-end
75 print '### connecting:',
76 conn = urllib.HTTPConnection(REQUEST_HOST)
77 conn.request('POST', REQUEST_URL, REQUEST_BODY % BBOX_SK, REQUEST_HEADERS)
78 response = conn.getresponse()
79 print 'done'
80
81 # iterate over response and store items
82 print '### parsing response:',
83 tree = lxml.etree.parse(response)
84 tree_root = tree.getroot()
85 item_count = 0
86 for feature_members in tree_root.findall('gml:featureMembers', tree_root.nsmap):
87     for frame in feature_members.findall('slm:frame', tree_root.nsmap):
88         fmr_slm_smid = frame.find('slm:smid', tree_root.nsmap).text
89         fmr_slm_smcislo = frame.find('slm:smcislo', tree_root.nsmap).text
90         fmr_slm_aktivne = frame.find('slm:aktivne', tree_root.nsmap).text
91
92         fmr_slm_smgpoint = frame.find('slm:smgeopoint', tree_root.nsmap)
93         fmr_slm_smgpoint_point = fmr_slm_smgpoint.find('gml:Point', tree_root.nsmap)
94         fmr_slm_smgpoint_point_pos = fmr_slm_smgpoint_point.find('gml:pos', tree_root.nsmap).text
95
96         temp = fmr_slm_smgpoint_point_pos.split(' ')
97         position_lat = temp[0]
98         position_lon = temp[1]
99
100         scraperwiki.sqlite.save(
101             keys=('smid', 'smcislo', 'aktivne', 'position_lat', 'position_lon'),
102             values={
103                 'smid': fmr_slm_smid,
104                 'smcislo': fmr_slm_smcislo,
105                 'aktivne': fmr_slm_aktivne,
106                 'position_lat': position_lat,
107                 'position_lon': position_lon,
108                 'scrap_time': datetime.datetime.utcnow().replace(microsecond=0).isoformat()
109             }
110         )
111         item_count += 1
112
113 print 'done (%d items parsed)' % item_count
114
115 # we're done, close the connection and note run time
116 conn.close()
117 scraperwiki.sqlite.save_var('last_run', datetime.datetime.utcnow().replace(microsecond=0).isoformat())
  
```




Praktické príklady

čo to znamená pre „apky“:

- **dobré:** Open Data = nižšie náklady na apku, „faster to market“
 - napr. RegisterUZ -> FinStat.sk
 - napr. ITMS2014+ -> KamIduEurofondy.sk
- **zlé:** „údaje na webe“ = veľké náklady na samotné získanie údajov
 - napr. justice.gov.sk -> otvorenesudy.sk

◦ <https://github.com/otvorenesudy/otvorenesudy-presentations/blob/master/otvorenesudy-20130822-rubyslava.pdf>

URL	Status	Domain	Size	Remote IP	Timeline
POST Sudne-rozhodni	200 OK	justice.gov.sk	1.7 MB	195.49.189.50:80	24.18s
GET www.justice.gov.	302 Redirect	justice.gov.sk	169 B		
GET www.justice.gov.	302 Redirect	justice.gov.sk	169 B		
GET www.justice.gov.	302 Redirect	justice.gov.sk	169 B		
GET www.justice.gov.	302 Redirect	justice.gov.sk	169 B		
GET default.aspx	200 OK	justice.gov.sk	63.7 KB		
GET default.aspx	200 OK	justice.gov.sk	63.7 KB	195.49.189.50:80	
GET default.aspx	200 OK	justice.gov.sk	63.7 KB	195.49.189.50:80	
GET default.aspx	200 OK	justice.gov.sk	63.7 KB	195.49.189.50:80	
9 requests			1.9 MB		39.1s (onload: 37.69s)

Response Body 1.7 MB (1,730,812 B)

Post Body 2.3 MB (2,426,365 B)

Total Received* 1.7 MB (1,731,187 B)

Total Sent* 2.3 MB (2,427,044 B)

* Including HTTP Headers

„go figure“

Open Data a „IT crowd“

IT crowd sú v Open Data vítaní a potrební!

- **scrapre**

- keď nie sú dodržané niektoré Open Data princípy a dáta treba „dolovať“
- napr. niektoré na <https://morph.io/soit-sk/>

- **analýzy, data journalism, data science**

- nájsť v údajoch informáciu, zaujímavosť, pridanú hodnotu
- využívané aj v slovenskej žurnalistike (Aktuality.sk, DennikN, SME, atď.)
- <https://www.pewe.sk/datalys/2017/01/19/introduction-to-python-for-data-science-at-edx/>
- <https://www.meetup.com/PyData-Bratislava/events/237319355/>
- ...

- **aplikácie, služby**

- niektoré z tých skôr spomenutých, ale aj mnohé ďalšie

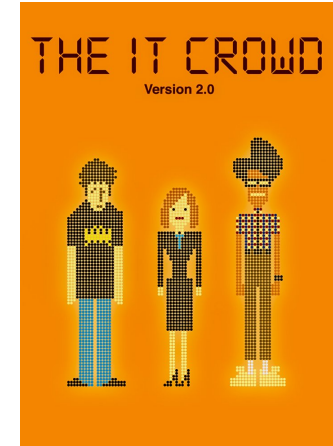


image by IMDb

Open Data a „IT crowd“

Aj IT crowd sú občania!

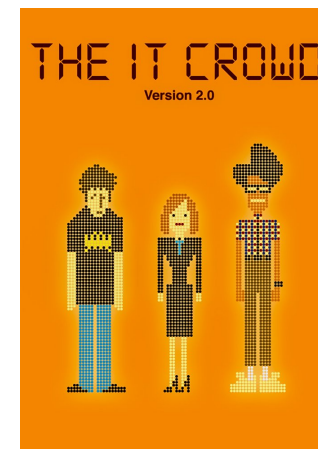


image by IMDb

Zíde sa občan s „IT“ a „data“ zručnosťami:

- #AllForJan hackathon
 - ale samozrejme aj iné
- komunita
 - OpenData.sk, Slovensko.Digital, atď.
- štandardizácia ISVS
 - pracovné skupiny KS a NKIVS, atď.



image by Aktuality.sk



Otázky?

