



WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS



Towards assessing the quality evolution of Open Data portals

Jürgen Umbrich, Sebastian Neumaier, Axel Polleres

ODQ Workshop, 30.03.2014, Munich, Germany

SUCCESS OF OPEN DATA ...

The White House

Office of the Press Secretary



For Immediate Release

May 09, 2013

Executive Order -- Making Open and Machine Readable the New Default for Government Information



HORIZONS NEW TECHNOLOGIES March 6, 2015

We can use open data to make smart cities “searchable”

We can use open data to make smart cities “searchable”

HORIZONS2 NEW TECHNOLOGIES2 March 6, 2015

CKAN instances around the world

The CKAN data management platform is in use by numerous governments, organisations and communities around the world. Being [open source](#), we don't know of all the instances, but here is a showcase of the ones we are aware of. We are proud and excited to see all the different uses being made of CKAN – if you would like to see your CKAN instance here [add it to the CKAN Census](#)!

To view instances by region or type, just choose the correct facet below.

Show: [All](#) [Region](#) [Type](#)

116 instances

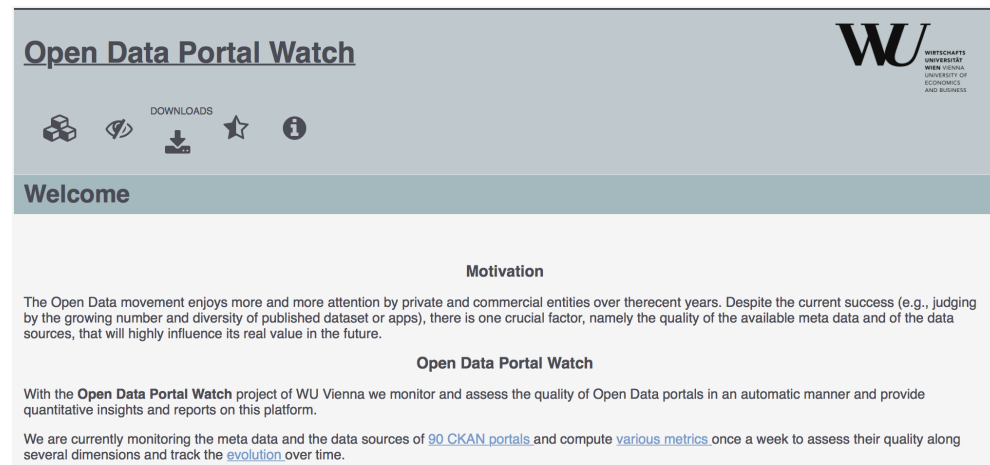
... FACES DATA QUALITY ISSUES

- Meta data
 - Missing values
 - Incorrect
 - Missing tags
 - No contact info
- Data
 - Changing URLs
 - Formats
 - Encoding



OPEN DATA PORTAL WATCH

- Periodically monitoring a list of Open Data Portals
 - 90 CKAN powered ODPs
- Quality assessment
- Evolution tracking
 - Meta data
 - Data



<http://data.wu.ac.at/portalwatch/>

QUALITY DIMENSIONS

DIMENSION	DESCRIPTION
Retrievability	The extent to which meta data and resources can be retrieved.
Usage	The extent to which available meta data keys are used to describe a dataset.
Completeness	The extent to which the used meta data keys are non empty.
Accuracy	The extent to which certain meta data values accurately describe the resources.
Openness	The extent to which licenses and file formats conform to the open definition.
Contactability	The extent to which the data publisher provide contact information.

Objective measures which can be automatically computed in a scalable way

CONCRETE METRICS (1/2)

- RETRIEVABILITY:
 - HTTP GET lookup for **dataset** (via API) or **resources**
- USAGE:
 - Ratio of used keys and all defined keys
 - Metrics also for subsets of keys
- COMPLETENESS:
 - Ratio of non-empty keys in a dataset

CONCRETE METRICS (2/2)

- Openness
 - Licence: opendefinition.org list
 - Format: manually selected based on articles, etc...
- Contactability:
 - Availability of contact information (text, url, email)
 - No check if email address bounces
- Accuracy (resource descriptions)
 - Currently based on HTTP RESPONSE HEADERS
 - Check: format, file size, mime-type

ODP LIST

Open Data Portal Watch



Brief overview of 89 Open Data CKAN portals

Sort by [Domain](#) [Country](#) [Datasets](#) [Resources](#) Filter: [Tile view](#) [Table View](#)

annuario.comune.fi.it Italy 358 DATASETS 1363 RESOURCES	catalogue.datalocale.fr France 303 DATASETS 751 RESOURCES	dados.gov.br Brazil 501 DATASETS 4344 RESOURCES	data.buenosaires.gob.ar Argentina 123 DATASETS 626 RESOURCES
data.edostate.gov.ng Nigeria 164 DATASETS 207 RESOURCES	data.glasgow.gov.uk United Kingdom (common practice) 384 DATASETS 1943 RESOURCES	datagm.org.uk United Kingdom (common practice) 360 DATASETS 506 RESOURCES	data.gov.sk Slovakia 216 DATASETS 556 RESOURCES
ckan.data.graz.gv.at Austria 151 DATASETS 341 RESOURCES	data.kk.dk Denmark 102 DATASETS 346 RESOURCES	data.lexingtonky.gov government 93 DATASETS 186 RESOURCES	data.nsw.gov.au Australia 311 DATASETS 458 RESOURCES
data.ohouston.org non-commercial 227 DATASETS 361 RESOURCES	data.ottawa.ca Canada 119 DATASETS 493 RESOURCES	data.cityofsantacruz.com commercial 52 DATASETS 72 RESOURCES	dados.recife.pe.gov.br Brazil 43 DATASETS 318 RESOURCES

Portal Overview

Open Data Portal Watch



Portal: GovData | Datenportal für Deutschland - GovData

OVERVIEW

DETAILS

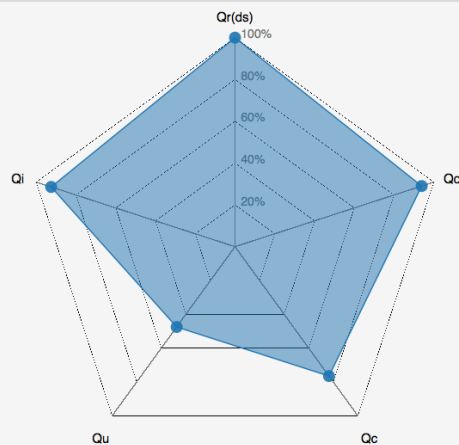
EVOLUTION

Available Snapshots



Snapshot: Sun Feb 22 2015 23:52:47 GMT+0100 (CET)

QUALITY



SIZE

DATASETS

13195

RESOURCES

37256

OPENNESS

LICENSE

AVG.
0.17

FORMAT

AVG.
0.94

RETRIEVABILITY

DATASETS

AVG.
1.00

RESOURCES

AVG.
0.79

CONTACTABILITY

EMAIL

AVG.
0.92

URL

AVG.
0.00

Portal Details

Available Snapshots



Snapshot: Sun Feb 22 2015 23:52:47 GMT+0100 (CET)

Age

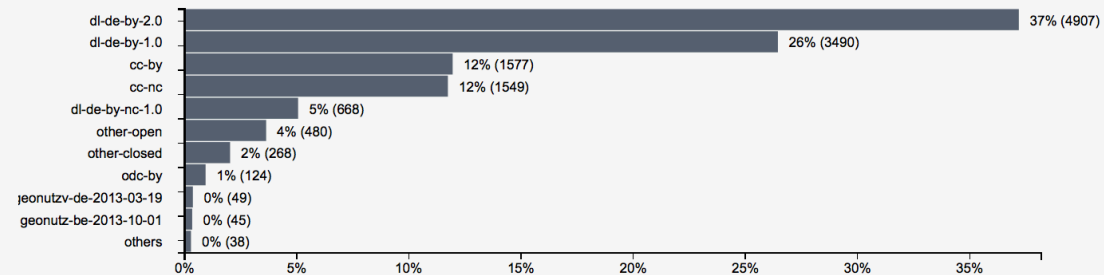
	oldest	Datasets average	newest	oldest	Resources average	newest
created:	2013-2-17	2014-7-24	2015-2-22	2012-7-9	2014-11-3	2015-2-23
modified:	2013-2-17	2014-11-22	2015-2-23	2012-3-12	2014-2-26	2015-2-23

Retrievability

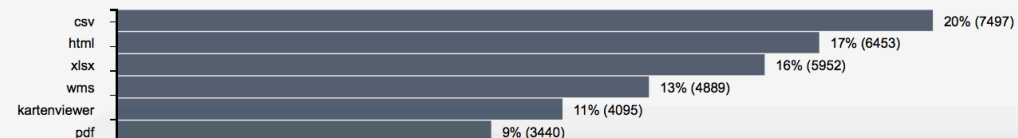
	Total	200 OK	403 Forbidden	404 Not Found	Server Timeout	Others
Datasets	13195	100%	0%	0%	0%	0%
Resources	22734	79%	0%	0%	18%	3%

Openness

Available licenses

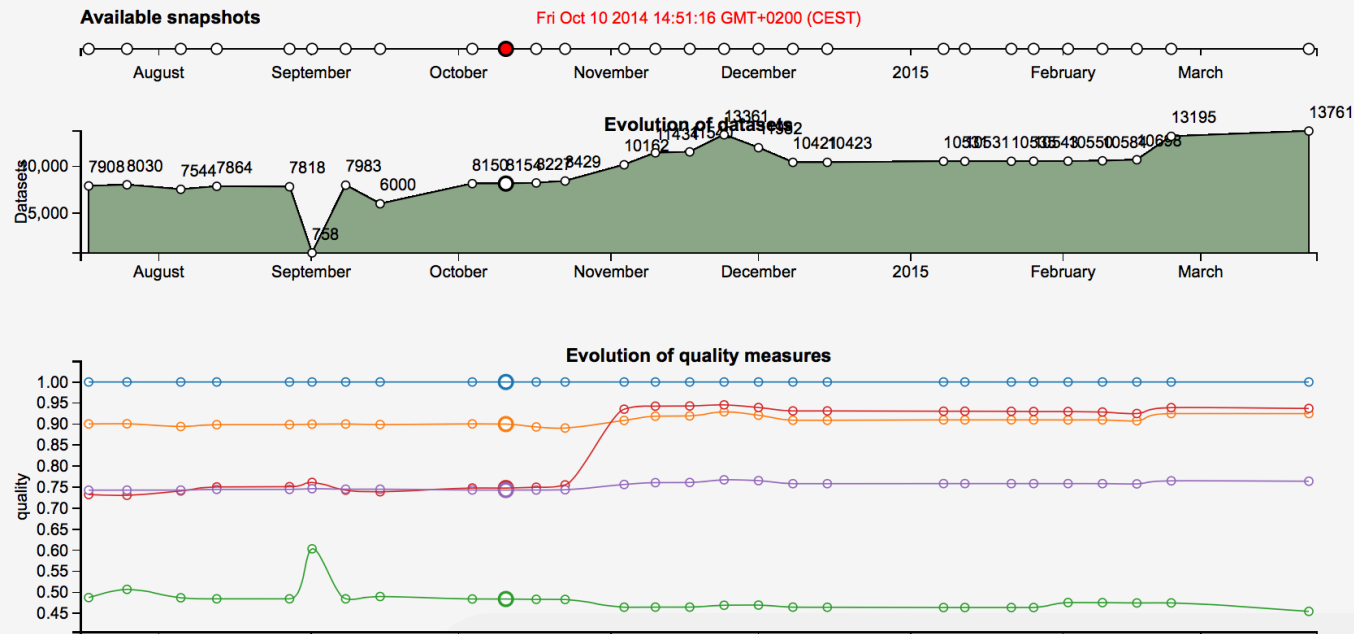



Available formats



WU
WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

WU



ODP CHANGES

Changes between the first and last snapshots

dataset changes

70 PORTALS WITH DATASET CHANGES

- Avg. increase by 87.05% for 60 portals
- Avg. decrease by -64.16% for 10 portals








Show entries

Search:










↑ PORTAL	↑ FROM	↑ TO	↑ CHANGE	↓ CHANGE PERCENTAGE
data.sa.gov.au (2014-07-17)→(2015-03-15)	484	5721	5237	1082.02%
datos.codeandomexico.org (2014-07-17)→(2015-03-15)	94	715	621	660.64%
data.opendataportal.at (2014-07-17)→(2015-03-16)	46	323	277	602.17%
annuario.comune.fi.it (2014-08-07)→(2015-03-15)	50	351	301	602.00%
udct-data.aigid.jp (2014-08-07)→(2015-03-16)	431	2110	1679	389.56%
catalogo.datos.gob.mx (2014-08-08)→(2015-03-15)	111	360	249	224.32%

Data Dumps

Open Data Portal Watch Dumps

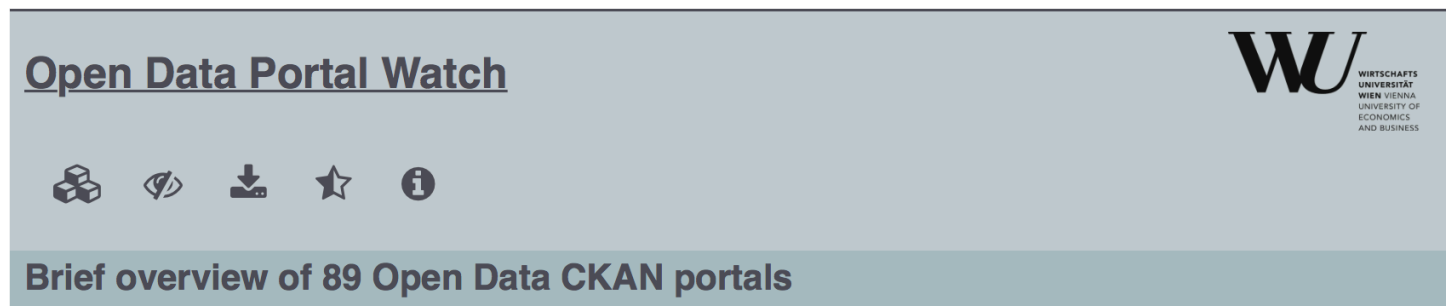
Name	Last modified	Size
 Parent Directory		-
 africaopendata.org/	16-Mar-2015 13:03	-
 annuario.comune.fi.it/	16-Mar-2015 13:03	-
 bermuda.io/	16-Mar-2015 13:14	-
 catalog.data.gov/	05-Feb-2015 15:28	-
 catalog.data.ug/	16-Mar-2015 13:07	-
 catalogo.datos.gob.mx/	16-Mar-2015 13:08	-
 catalogodatos.gub.uy/	16-Mar-2015 13:15	-

Open Data Portal Watch Dumps

Name	Last modified	Size
 Parent Directory		-
 2014-07-17.gz	05-Feb-2015 15:13	2.2M
 2014-07-25.gz	05-Feb-2015 15:13	2.2M
 2014-08-05.gz	05-Feb-2015 15:13	2.2M
 2014-08-12.gz	05-Feb-2015 15:13	2.2M
 2014-08-27.gz	05-Feb-2015 15:13	2.2M
 2014-09-01.gz	05-Feb-2015 15:14	2.2M
 2014-09-07.gz	05-Feb-2015 15:14	2.2M
 2014-09-14.gz	05-Feb-2015 15:14	2.2M

Open Data Portal Watch

- A framework to periodically monitor and assess the quality evolution of Open Data portals
 - 6 Quality dimensions
 - Weekly snapshots
 - Evolution analysis
 - Portal Meta Data as data dumps



<http://data.wu.ac.at/portalwatch/>

Open calls

Calls for Research & Innovation Papers, Industry Presentations and Poster/Demos are now open



Call for Research & Innovation Papers

The Research & Innovation track at SEMANTICS welcomes the submission of papers on novel scientific research and/or innovations relevant to the topics of the conference. Submissions must be original and must not have been submitted for publication elsewhere. Papers should follow the ACM ICPS guidelines for formatting (<http://www.acm.org/sigs/publications/proceedings-templates>) and must not exceed 8 pages in length for full papers and 4 pages for short papers, including references and optional appendices.

Abstract Submission Deadline	May 22, 2015
------------------------------	--------------



Call for Industry & Use Case Presentations

To address the needs and interests of industry SEMANTICS presents enterprise solutions that deal with semantic processing of data and/or information in areas like like Linked Data, Data Publishing, Semantic Search, Recommendation Services, Sentiment Detection, Search Engine Add-Ons, Thesaurus and/or Ontology Management, Text Mining, Data Mining and any related fields. All submissions have a strong focus on real world applications beyond the prototypical status and demonstrate the power of semantic systems!

Submission Deadline	July 1, 2015
---------------------	--------------



Call for Posters and Demos

The Posters & Demonstrations Track invites innovative work in progress, late-breaking research and innovation results, and smaller contributions (including pieces of code) in all fields related to the broadly understood Semantic Web. The informal setting of the Posters & Demonstrations Track encourages participants to present innovations to business users and find new partners or clients. In addition to the business stream, SEMANTICS 2015 welcomes developer-oriented posters and demos to the new technical stream.

Submission Deadline	June 17, 2015
---------------------	---------------