Czech CKAN Repository as Case Study in Public Sector Data Cataloging

Jan Kučera[1], Dušan Chlapek[1], Jindřich Mynarz[1]
[1] University of Economics, Prague, Czech Republic
{jan.kucera,chlapek}@vse.cz, mynarzjindrich@gmail.com

Abstract. *Over the past few years number of public sector data catalogs started to exist. These data catalogs cover data from different levels of the public sector and they represent results of both volunteer activity and government policy. In this paper we present our experience from cataloging of the Czech public sector data using software Comprehensive Knowledge Archive Network (CKAN) and based on this experience we propose generic model of data cataloging tool features. Content of the Czech CKAN is a result of the academic initiative and thus it is an unofficial catalog of datasets provided by the Czech public sector bodies. Despite the limited scope of the cataloging activity it provided valuable information about available datasets provided by the Czech public sector and the results can help future cataloging initiatives.*

Keywords: Public sector data, data cataloging, data catalog, CKAN, Comprehensive Knowledge Archive Network, data analysis, Open data, Open source for e-govenent, metadata, public information

1. Introduction

Public sector bodies produce large amount of data that can be potentially reused by different subjects in various applications. In his review Vickery [35] estimates that the direct public sector information (PSI) reuse market in 27 EU member states "*was of the order of EUR 28 billion in 2008*" and he adds that "*with easier access, improved infrastructure and lower barriers, aggregate direct and indirect economic benefits for the whole EU27 economy could have been of the order of EUR 200 billion (1.7% of GDP) in 2008*".

Open government is a movement which aims at making governments more transparent and democratic and which tries to establish cooperation between public administration, citizens and industry representatives [2]. Free access and reuse of public sector information, which is at the heart of the open government data movement is seen as a key factor of transparency and open government. Open data is "*data that can be freely used, reused and redistributed by anyone - subject only, at most, to the requirement to attribute and share-alike*" [27].

Public sector information and data are often scattered across different portals which makes them difficult to find [19]. Creating a data catalog of public sector data is a way how to make these datasets more accessible and thus more reusable. Several countries have already started their data cataloging activities. USA and their `data.gov` portal [34] or Great Britain and its `data.gov.uk` portal [11] are examples of such countries and their data catalogs.

According to [5] data catalog is "*a collection of catalog records*". These records consist of metadata describing a dataset which represent "*a collection of information in a machine-readable format*" [5]. Because data catalogs are central point of access to

datasets provided by different public sector bodies they can become the key enablers of open government policy [6].

In this paper we present the experience we have gained during cataloging of datasets published by the Czech public sector bodies. We focus on a software tool Comprehensive Knowledge Archive Network (CKAN) as it powers Czech public sector data catalog. We also introduce a generic model of functionality provided by data cataloging tool. The aim of this paper is to spur a discussion about the methodological approaches to data cataloging and the functionality provided by data cataloging tools. Doing this we would also like to encourage international exchange of experience in this area.

This paper is structured as follows. First we introduce data cataloging activities in the EU member states with a focus on usage of the software CKAN. In the next section we present the Czech data catalog of public sector data. We describe how the data cataloging tool was selected and what the scope of the data cataloging activity was. Next, we introduce a methodology for use of the software CKAN. We provide an overview of the cataloged datasets and we briefly describe use of the models of dataset structure. In the next section we propose a generic model of features that should be provided by data cataloging tool. Concluding remarks are presented at the end of this paper.

2. Data Cataloging Activities

The above mentioned data portals of the USA and Great Britain are not the only data cataloging activities that are being conducted nowadays. There are data catalogs covering public sector data of other countries as well. In the next section we provide details on data cataloging activity that aimed at Czech public sector data. In this section we focus on software Comprehensive Knowledge Archive Network (CKAN) because an instance of this software dedicated to the data of the Czech Republic was used during this cataloging activity [21].

CKAN is an open source web application for online data portals and data catalogs [18]. It is developed by the Open Knowledge Foundation. The CKAN project is governed by the Project Steering Committee [28]. International community of users and contributors has formed around CKAN.

CKAN provides a user interface and an application programming interface (API) [25]. This allows people as well as other applications to use and extend the content of the data catalog.

Content of the CKAN data catalog consists of CKAN packages[1]. Package contains description and other metadata about one or more related datasets (resources) [26]. By default the CKAN package contains only links to the datasets, not the datasets themselves. However with a proper configuration of storage backend it is possible to upload resources directly into CKAN [23]. According to the CKAN documentation the

---

[1] *In CKAN terminology the term package was used for catalog record. In current version of CKAN term CKAN dataset is used instead [22]. But because it can be mistaken for dataset provided by the data sources in this paper we use older term CKAN package or package for CKAN catalog record.*

available resource types are file, uploaded file, list of downloadable files, API and service [24].

CKAN provides predefined set of metadata attributes like author, maintainer, dataset source, homepage of the data sources etc., but the user can use extras fields to add additional attributes. What the data are about is expressed by tags – freely created text strings. Related CKAN packages can be put into a group which can be also used to restrict access to the packages. For example, it can be set that only members of certain groups can edit a particular package.

Basic functionality of CKAN focuses on cataloging datasets and searching in the content of the catalog but functionality can be extended by extensions [8].

Portal Datacatalogs.org [1] is a catalog of open data catalogs and it shows that CKAN powers a number open data catalogs today. Table 1 summarizes the usage of CKAN data catalogs for cataloging of EU member states' public sector data.

Table 1. CKAN data catalogs for EU member states' public sector data

State	Catalog	Packages	As of date
United Kingdom	http://data.gov.uk/data	8647	July 17, 2012
Spain[2]	http://opengov.es/	868	March 9, 2012
Germany	http://offenedaten.de/	472	July 17, 2012
Netherlands	**http://data.overheid.nl** (**http://nl.ckan.net/**)	280	July 17, 2012
Ireland	http://ie.ckan.net/	268	July 17, 2012
Italy	http://it.ckan.net/	242	July 17, 2012
Czech Republic	**http://cz.ckan.net**	171	July 17, 2012
Austria[3]	http://offener.datenkatalog.at/	22	March 9, 2012
Finland	http://fi.thedatahub.org/	17	July 17, 2012

We used Datacatalogs.org and Google search to locate the data catalogs. Data catalog of the Czech Republic is a result of academic activity and it received no support from the public administration officials. Unlike our case the British data.gov.uk portal is funded from the public spending and overseen by the Transparency Board [9] established by the Prime Minister [33].

But CKAN is not only used to power data catalogs at the national level. Berlin Open Data [17] or Helsinki Region Infoshare [10] are examples of CKAN data catalogs at the municipal level.

There are not only national or municipal data catalogs. Federated data catalog of public sector data published by the European Union member states and their

---

[2]   *Portal http://opengov.es/ was inaccessible on 17th July 2012 therefore no updated package statistics was available.*

[3]   *Portal http://offener.datenkatalog.at/ was inaccessible on 17th July 2012 therefore no updated package statistics was available. However, beta version of a new Austrian data catalogue is currently available at http://www.data.gv.at/.*

institutions PublicData.eu is being developed [19]. Data from the Czech CKAN public sector data catalog is now accessible through the PublicData.eu portal [20].

Of course, CKAN software is not the only software tool that is used to power public sector data catalogs. For example State of Oregon in the USA uses Socrata platform [30] and Danish data catalog is available via Digitalisér.dk portal [32].

3. Cataloging Datasets of the Czech Public Sector

The cataloging of datasets provided by the Czech public sector bodies started as an academic activity which was part of an internal research grant of the University of Economics, Prague. The following main goals were set for this activity:

- select a software tool for data catalog and validate maturity of its functionality,
- catalog datasets published by the selected Czech public sector bodies and analyze how many of these datasets comprises of Open data,
- formulate a methodology for using the cataloging tool,
- adapt the cataloging tool to the Czech conditions.

In the following subsections we discuss the Czech public sector data cataloging activity in more detail.

3.1 Selection of the Data Cataloging Tool

We considered three options how to acquire data cataloging software: building custom data catalog application, using online spreadsheet application as data catalog and using some of the available data cataloging software.

The benefit of the first two options is that we would have been able to define the structure of the data catalog as well as the metadata attributes. Custom application could have been fully tailored to our needs but building of such application was considered too costly and time consuming. Online spreadsheet application seemed suitable for basic datasets cataloging and it would have allowed us online collaboration, but we were afraid that with growing number of records in the catalog its maintenance would require significant effort. We also doubted that such data catalog can be easily used by anyone outside of the project team.

We decided to use the CKAN software. It is software that was designed as a data cataloging tool from the start. It was rather fortunate that the Open Knowledge Foundation provides an instance of this tool for the Czech Republic [21]. So we decided to use this available instance of CKAN and add content into this data catalog which was empty at the time we started our cataloging activity.

3.2 Scope of the Cataloging Activity

During the project it was not feasible to catalog all datasets of the entire Czech public sector due to the large number of public sector bodies. We had to restrict the scope of the data cataloging activity to a subset of public sector bodies. During the course of defining of this subset we went from the state level down to the municipal level of public sector. We cataloged datasets of all 14 ministries, all 14 regions and also datasets of 4 cities of the Liberec region. Datasets of the Czech Statistical Office were

catalogued as well because the Czech Statistical Office produces a large number of valuable statistical data.

3.3 Cataloging Approach

We focused our cataloging activity on datasets published on the official web sites of the selected public sector bodies. We decided not to include every dataset that was published on these web sites and we cataloged only datasets that were machine-readable. Therefore we focused on APIs of publicly accessible services and structured data like XML, CSV files and MS Excel spreadsheet files.

We focused on statistical data, budgets and data about grants and other activities of the selected public sector bodies. We did not set any strict inclusion criteria for datasets and every cataloger had to decide whether one particular dataset was appropriate for inclusion into the data catalog or not.

During the cataloging activity we came across statistical datasets or grants results data that were published as HTML tables on the website or as PDF files (see for example [14-15]). We did not include these data sources into the data catalog because we considered these data formats not suitable for direct computer processing without prior conversion into some other formats (relational database, CSV).

The cataloging itself was performed manually and three people were assigned to this task. This allowed individual assessment of each dataset before it was added into the data catalog. On the other hand this approach was naturally error prone and some of the datasets might have been omitted. Thus, the resulting data catalog cannot be considered as a complete and comprehensive catalog of data sources provided by the analyzed public sector bodies.

3.4 Methodology and Adaptation of the Data Cataloging Tool

In order to keep the description of CKAN packages consistent we formulated a methodology for the use of the CKAN software and we published this methodology so that other users and contributors to the data catalog can follow the same guidelines [12]. In the following paragraphs we highlight some of the key methodological guidelines.

We wanted the content of the data catalog to be accessible not only to people from the Czech Republic but also to people from other countries. Although the CKAN software is available in several languages, at the beginning of our cataloging activity it was not possible to create variants of the CKAN package description in different languages. CKAN multilingual metadata features are currently in development [16], [36]. Of course, it was possible to create separate CKAN packages for each language variant but it would have been difficult to keep all language variants consistent. We decided to provide both Czech and English description and we translated also names of the datasets. For this reason description and names of each CKAN package have Czech and English section. We deem this to be a sufficient solution for the bilingual description, however adding more languages would require different approach because texts could become too long, especially the dataset names.

We grouped related datasets into one CKAN package and we used tags for categorization of the packages. We tried to tag the packages with appropriate

concepts from Integrated Public Sector Vocabulary (IPSV) [4] and classification of economic activities CZ-NACE [7]. IPSV concept's name (preferred label) and CZ-NACE concept's name were used as tags. Because CKAN tags are simple text strings direct linking to the concepts by providing their URIs was not possible. We rejected the possibility to add concept's full URIs as tags because they are difficult for users to understand. Because we wanted to make future computer processing of the CKAN packages easier we added specially formatted tags in the format of "`IPSV_ID`" where `ID` stands for an identifier of the IPSV concept and "`CZ-NACE_Code`" where `Code` stands for an identifier of CZ-NACE concept. If needed, these tags can be parsed and the concepts can be looked up by their identifiers. A rule was set that whenever a concept from IPSV or CZ-NACE is used that the CKAN package must be tagged with both the concept's name and the special tag containing identifier of the concept.

In the methodology we also recommend creating superior CKAN packages or meta-packages[4] for related CKAN packages or packages covering the same topic or issue. We defined CKAN meta-package as a package that contains no links to data sources but only links to other related CKAN packages. This can help user to navigate between related packages without the need to use search capabilities of CKAN. Labor costs regional statistics meta-package [13] is one such example because it provides links to other packages each containing data about one particular region.

In addition to the methodology we created a short quick start guide aimed at the new users of CKAN [31]. We also translated the CKAN user interface into Czech because there was no Czech localization available at the beginning of the project.

3.5 Cataloged Datasets

We cataloged 1470 datasets which were grouped into 141 CKAN packages. 358 tags were used to categorize these packages. The most common data format was MS Excel 2003 spreadsheet (xls) which comprised 93% of the datasets.

Largest number of datasets with the largest total size was published by the Ministry of Labour and Social Affairs: 814 files, 542.2 MB total. Out of this number, 780 files were comprised by the labor cost regional statistics data. Czech Statistical Office published 218 cataloged files and the Ministry of Interior of the Czech Republic published 124 files. Other analyzed public sector bodies published less than one hundred cataloged data files.

Table 2 summarizes what amount of cataloged data each of the analyzed Czech public sector bodies published.

---

[4]   Superior CKAN package or CKAN meta-package are no official CKAN terms. These terms we defined solely for the purpose of our methodology.

Table 2. Cataloged datasets (as of May 2011)

Author	Packages	Datasets	Size (MB)
Ministry of Labour and Social Affairs	23	814	542.2
Ministry of Finance of the Czech Republic	22	58	8.2
Ministry of the Interior of the Czech Republic	14	124	363
Czech Statistical Office	13	218	39.1
Ministry of Agriculture of the Czech Republic	13	82	36.5
Olomouc region	7	12	9.5
Ministry of the Environment of the Czech Republic	5	16	0.4
Ministry of Defense and Armed Forces of the Czech Republic	4	19	2.3
Pilsen region	4	6	1.3
Central bohemian region	5	5	6.4
Ministry of Transportation	3	28	0.9
Ministry of Justice	3	24	2.9
Pardubice region	3	10	2
Ustí region	3	9	2.3
Ministry of Culture	3	7	1.5
Karlovy Vary region	3	3	0.9
South Moravian region	3	3	0.4
Ministry of regional development	2	10	0.4
Hradec Králove region	2	6	0.6
Vysočina region	1	9	0.4
Liberec region	1	3	3.9
Czech office for standards, metrology and testing	1	1	3.6
General Financial Directorate	1	1	0.1
Ministry of education, youth and sports	1	1	0.1
South Bohemian Region	1	1	0.04
Total	**141**	**1470**	**1028.7**

One of the goals of the project was to analyze what amount of published datasets is open. However, we will not discuss the legal aspects of open data in the context of Czech legislation. We found out that datasets published by the Czech public sector bodies are not clearly licensed. Therefore, we marked those datasets in the Czech CKAN data catalog with the label "`License not specified`".

3.6 Modeling the Dataset Structure

As a part of the analysis of the cataloged datasets we created a conceptual UML class diagrams describing structure of a small subset of the cataloged data. This provided additional insight into what data and about what entities are contained in the analyzed data files. Knowledge of the dataset structure helped us later in the project when we experimentally transformed some of the cataloged tabular data into RDF data format

as linked data format (for additional information about linked data technologies see for example [2-3]).
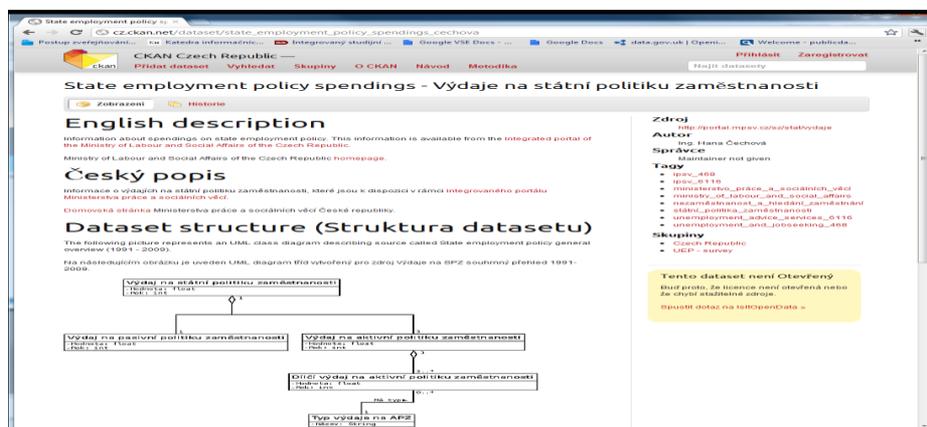
Fig. 1. CKAN package with model of the dataset structure

3.7 Discussion of Results

Our experience shows that CKAN is quite mature data cataloging tool. It provides necessary functionality to add and maintain data catalog records and to search in the contents of the data catalog. None of the problems we faced during cataloging significantly hindered our work.

Today, citizen who needs data from more than one public sector body usually has to search through the web site of each institution he or she is interested in. Despite the fact that our cataloging activity was restricted to a defined subset of the Czech public sector bodies, it shows that data catalog can provide unified access to the public sector data sources from different institutions and thus make reuse of public sector information easier.

None of the cataloged datasets was marked as open data due to the missing licensing information. Having clear statement of rights and responsibilities governing the use of the published data can help reduce legal uncertainty in this issue.

UML class diagrams describing datasets structure were successfully used during our experimental transformation of tabular data into RDF data format. Knowledge of the dataset structure might be needed in other transformations of cataloged datasets or during the development of applications that would use the cataloged data. Another scenario is ontology engineering based on the commonly used schemata. We suggest the effective use of conceptual models in data catalogs as a future research topic.

CKAN usage methodology was a key element of our cataloging activity. In the methodology we established a common approach to the package creation and description which helped us to keep the packages consistent. Methodological guidelines also deal with some important questions such as whether to create separate packages for every dataset or to group related datasets into one CKAN package (as was stated above, we decided to group related datasets in order to avoid unmanageable number of CKAN packages). We therefore suggest formulation of

methodology as a step that should be considered during any data cataloging activity. Exchange of experience in this area can help establish best practices in data cataloging.

4. Overview of Features of Data Cataloging Tool

In this section we propose a generic model of features that should be provided by the data cataloging software tools. This model was created as a generalization of features provided by the CKAN software and it reflects our own experience with cataloging the Czech public sector data sources. Some of the features were also inspired by the Socrata platform [29]. We grouped the proposed features as shown on figure 2. Each of the groups is described in more detail in the following subsections.
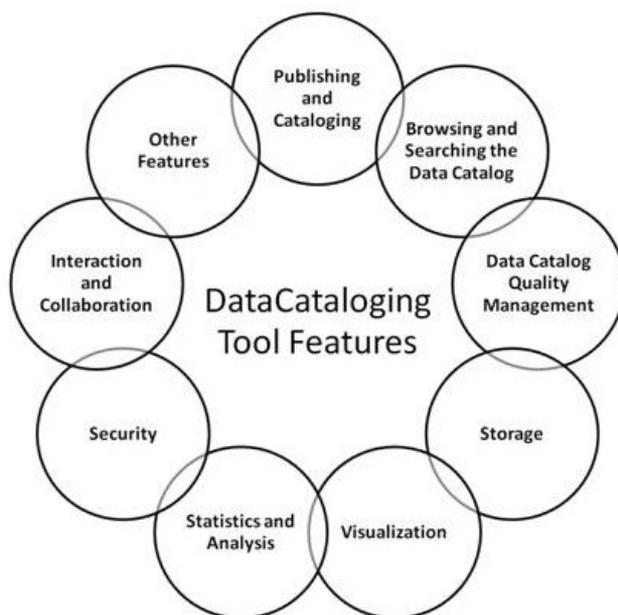

Fig. 2. Data cataloging tool features

4.1 Publishing and Cataloging

Cataloging features accompanied with search capabilities constitute the core features of the data cataloging tool. Essential features enable creating, storing and editing of the data catalog records. Tags or other means how to express what the data are about should be available in addition to the more detailed descriptions.

Based on our experience with the IPSV vocabulary we suggest that data catalog should provide functionality to tag the catalog records with concepts from selected vocabulary or predefined sets of terms. We also suggest adding graphical objects like UML class models to the package or dataset description.

In collaborative environment where more users can edit the same catalog record revision history might be necessary. Policies with strict requirements on data catalog quality might need the functionality of approved revision which allows the package maintainer to accept or reject revisions made to the packages he or she oversees. In scenarios where large amount of data needs to be cataloged automatic collection of available dataset metadata like for example file size, name or format can reduce the amount of manual work and it can increase metadata accuracy.

4.2 Browsing and Searching the Data Catalog

In order to effectively use the data catalog users must be able to find datasets they need. Thus the data cataloging tool should provide full-text search as well as search by tags or categories of datasets and catalog records. Advanced search by metadata attributes (for example by the author of the datasets) or filtering of search results should be also available.

Users might look for data that are related to some geographic location. If the metadata contain necessary geo coordinates it is possible to look for the data by selecting appropriate area in the map or by providing a set of geo coordinates describing this area.

4.3 Data Catalog Quality Management

Catalog records require maintenance and quality assurance because without it dataset metadata can be incomplete and it can become obsolete over time. Data cataloging tools should therefore provide a function to check whether the links to the data sources are still valid and provide report of invalid data sources. Checking which catalog record have missing metadata attributes should be also possible. This can for example help to identify records with missing description.

4.4 Storage

Data cataloging tool must be able to store metadata descriptions of datasets but it can be capable of storing the datasets themselves. The latter allows users not only to catalog datasets published somewhere else but it allows them to upload datasets directly into the data catalog and publish them in this way.

4.5 Visualization

Visualization can help users to better understand data. If the data cataloging tool provides access not only to the dataset metadata but to the raw data contained in the datasets charts and graph might be used to visualize this data. Data that relate to some geographic location can be visualized using a map.

4.6 Statistics and Analysis

The data cataloging tool should provide a basic set of statistics about the content of the data catalog, for example number of packages, number of resources or datasets, number of users, activity by days/months etc. Advanced users might want to perform custom ad-hoc queries about the content of the data catalog.

4.7 Security

Data cataloging tool must provide at least some basic security features. It must be possible to perform backup and recovery of the data catalog contents in order to protect the catalog from accidental loss of its entire contents.

Depending on the policy or approach it might be desirable that not every user can edit or view all of the available packages or datasets. Thus the data cataloging tools should provide some basic access control features to restrict access based on users, roles or package groups.

4.8 Interaction and Collaboration

Users of the data catalog can provide valuable feedback to the catalog maintainers. The data cataloging tool should therefore allow users to add comments to the packages or datasets and rate their quality. Features for submitting proposals or requesting for data can help track the demand for new datasets.

4.9 Other Features

If the data cataloging tool is aimed at international community of users and contributors it should provide multilingual user interface and multilingual description of packages and datasets. Quality documentation and tutorials are needed as well.

Application programming interface (API) is needed to utilize the data catalog in other applications.

5. Conclusion

The number of data cataloging activities that aim at public sector data is being conducted in Europe and other parts of the world. In this paper we presented our experience from cataloging public sector data in the Czech Republic using open source web application CKAN. We cataloged total number of 1470 datasets published by 24 public sector bodies. 93% of these datasets comprised of MS Excel 2003 spreadsheet files. We defined a methodology for use of the CKAN software which helped us to keep the CKAN packages consistent. We consider a methodology that guides a cataloging activity to be one of its key success factors.

Generic model of features provided by the data cataloging tool was also introduced in this paper. Although this paper focused mainly on the CKAN software other data cataloging tools are used as well. This generic model of data cataloging tool's features can be used as a reference during selection and evaluation data cataloging tools. By presenting this model we also want to spur a discussion about features that the data cataloging tools should provide.

We are going to keep on cataloging the public sector data sources and further developing the methodology. We will translate the methodology into English because it proved to be very helpful during our cataloging activity and we hope that English version will attract more catalogers outside of the Czech Republic. We see it also as a way how to receive valuable feedback and comments on the methodology. Our ongoing CKAN activities will contribute to the EU ICT FP7 project "LOD2 – Creating Knowledge out of Interlinked Data".

References

1. About datacatalogs.org, http://datacatalogs.org/about
2. Bauer, F., Kaltenböck, M., 2011. *Linked Open Data: The Essentials.* Edition mono/monochrom, Vienna
3. Bizer, C., Heath, T., Berners-Lee, T., 2009. Linked Data - The Story So Far. In: Special Issue on Linked Data, *International Journal on Semantic Web and Information Systems*
4. Cabinet Office e-Government Unit, 2006. Integrated Public Sector Vocabulary (IPSV2_00) , http://doc.esd.org.uk/IPSV/2.00.html
5. Cyganiak, R., Maali, F., 2012. *Use Cases and Requirements for the Data Catalog Vocabulary*, http://dvcs.w3.org/hg/gld/raw-file/default/dcat-ucr/index.html
6. Cyganiak, R., Maali, F., Peristeras, V., 2010. Self-Service Linked Government Data with dcat and Gridworks. In: *Proceedings of the 6th International Conference on Semantic Systems*, pp. 37:1–37:3. ACM, New York
7. Czech Statistical Office: *Classification of Economic Activities* (CZ-NACE) (2011), http://apl.czso.cz/iSMS/en/klasstru.jsp?kodcis=80004&cisjaz=8260
8. Extensions, 2012, http://wiki.ckan.org/List_of_Extensions
9. FAQ, http://data.gov.uk/faq
10. Helsinki Region Infoshare: About, http://www.hri.fi/en/about/
11. HM Government: data.gov.uk - Opening Up Government, http://data.gov.uk/
12. Kučera, J., Tajtl, M., Mynarz, J., Chlapek, D., 2011. Metodika pro práci se software CKAN [CKAN usage methodology] , http://www.opendata.cz/cs/node/20
13. Labour cost regional statistics (2011), http://cz.ckan.net/dataset/labour_cost_regional_statistics
14. Ministry of Finance of the Czech Republic: Veřejné rozpočty 2010 - TABULKA 2 – Příjmy [Public budgets 2010 - Table 2 - Incomes], http://www.mfcr.cz/cps/rde/xchg/mfcr/xsl/vlad_fin_stat_62902.html
15. Ministry of the Environment of the Czech Republic: Konečné výsledky výběrového řízení na projekty NNO pro rok 2011 [Final results of tenders for NGO 2011 projects], http://www.mzp.cz/cz/vysledky_projekty_nno_2011
16. Multilingual Metadata, http://trac.ckan.org/ticket/1736
17. Offene Daten Berlin [Berlin Open Data], http://daten.berlin.de/
18. Open Knowledge Foundation: About CKAN, 2011, http://ckan.org/about/
19. Open Knowledge Foundation: About publicdata.eu, 2011, http://publicdata.eu/about
20. Open Knowledge Foundation: Availability map, 2011, http://publicdata.eu/map
21. Open Knowledge Foundation: CKAN Czech Republic, 2011, http://cz.ckan.net/

22. Open Knowledge Foundation: CKAN Data Management System Documentation - CKAN CHANGELOG (2012), http://docs.ckan.org/en/latest/CHANGELOG.html

23. Open Knowledge Foundation: CKAN Data Management System Documentation - FileStore and File Uploads (2012), http://docs.ckan.org/en/latest/filestore.html

24. Open Knowledge Foundation: CKAN Data Management System Documentation - Resource (2012), http://docs.ckan.org/en/latest/domain-model-resource.html

25. Open Knowledge Foundation: Features (2011), http://ckan.org/features/

26. Open Knowledge Foundation: Metadata (2011), http://ckan.org/features/metadata/

27. Open Knowledge Foundation: The Open Data Manual (2011), http://opendatamanual.org/index.html

28. Project Steering Committee, http://wiki.ckan.org/Project_Steering_Committee

29. Socrata: The Socrata Open Data Platform (2012), http://www.socrata.com/solutions/solution-tour/

30. Socrata: State of Oregon, 2012, http://www.socrata.com/customer-spotlight/state-of-oregon/

31. Tajtl, M., 2011. Návod na práci s katalogem CKAN [CKAN catalog quick start guide], http://opendata.cz/en/node/30

32. The Danish National IT and Telecom Agency: About Digitalisér.dk (2009), http://digitaliser.dk/resource/432461

33. Transparency Board: New Public Sector Transparency Board and Public Data Transparency Principles (2010), http://data.gov.uk/blog/new-public-sector-transparency-board-and-public-data-transparency-principles

34. United States Government: Data.gov, http://www.data.gov/

35. Vickery, G. 2011. Review of recent studies on PSI re-use and related market developments

36. Wainwright, M., 2012. News from the CKAN team, http://www.ckan.org/2012/02/07/news-from-the-ckan-team/

JEL: M10, O10