Open Data publication in a nutshell

An introduction to the deliverable D5.1 “Methodology for publishing datasets as open data”

Project	Components Supporting the Open Data Exploitation
Acronym	COMSODE
Contract Number	FP7-ICT-611358
Start date of the project	1st October 2013
Duration	24 months, until 30th September 2015

Date of preparation	21st September 2015
Authors	Jan Kučera, Martin Nečaský, Jakub Klímek, Dušan Chlapek
Responsible of the deliverable	CUNI
Email	necasky@ksi.mff.cuni.cz
Dissemination level	PU - Public

Table of contents

1 Introduction

In order to provide publishers of Open Data with guidelines and recommendations how Open Data should be published, in the COMSODE project we developed the Methodology for publishing datasets as open data. It is a comprehensive methodology covering various aspects of Open Data publication ranging from technical and legal issues to organizational issues and user engagement.

The scope of the methodology is broad and even though you can just focus on topics you are interested in, i.e. you do not need to read it whole, it still might be too complex if you just want to get familiar with what Open Data is and what the basic guidelines for its publication are. Therefore we decided to write this short handbook which should provide you with the necessary basics about Open Data, with the overview of the methodology and its contents and with focused step-by-step guides to the phases and activities related to the Open Data publication.

We hope that it will help you to get started with your Open Data initiative. If you ever feel you need a more specific guidelines or recommendations, please keep in mind that you can always refer to the complete methodology.

2 What is Open Data and why is it important

According to Open Knowledge (2015) Open Data is "*data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike.*" Open Definition (Open Knowledge, 2014) provides a more precise definition of what the openness means in Open Data, however there two main dimensions of openness of open data: legal and technical.

Data and databases might be protected by intellectual property rights or other kinds of rights. Therefore it is necessary to ensure that the users have the necessary permission to legally reuse and redistribute open datasets. Common way to achieve this objective is to properly licence open datasets. We provide recommendations on licensing of the open datasets in the Methodology. However you can find other papers that discuss this topic in more detail, for example (Dulong, 2014).

Technical openness ensures that open data are accessible and easily usable. Most notably open data should be available for download from the Internet (or accessible via API) in open and machine-readable formats. The Methodology provides detailed recommendations and guidelines how to publish data in open and machine readable formats.

So use of machine-readable formats is one of the basic requirements for Open Data. However there are different machine-readable formats allowing different ways of data processing. Tim Berners-Lee (2006) suggested a 5 star deployment scheme. Each star represents one level of openness of a dataset based on what formats are used. This 5 star scheme is depicted on the figure 1.

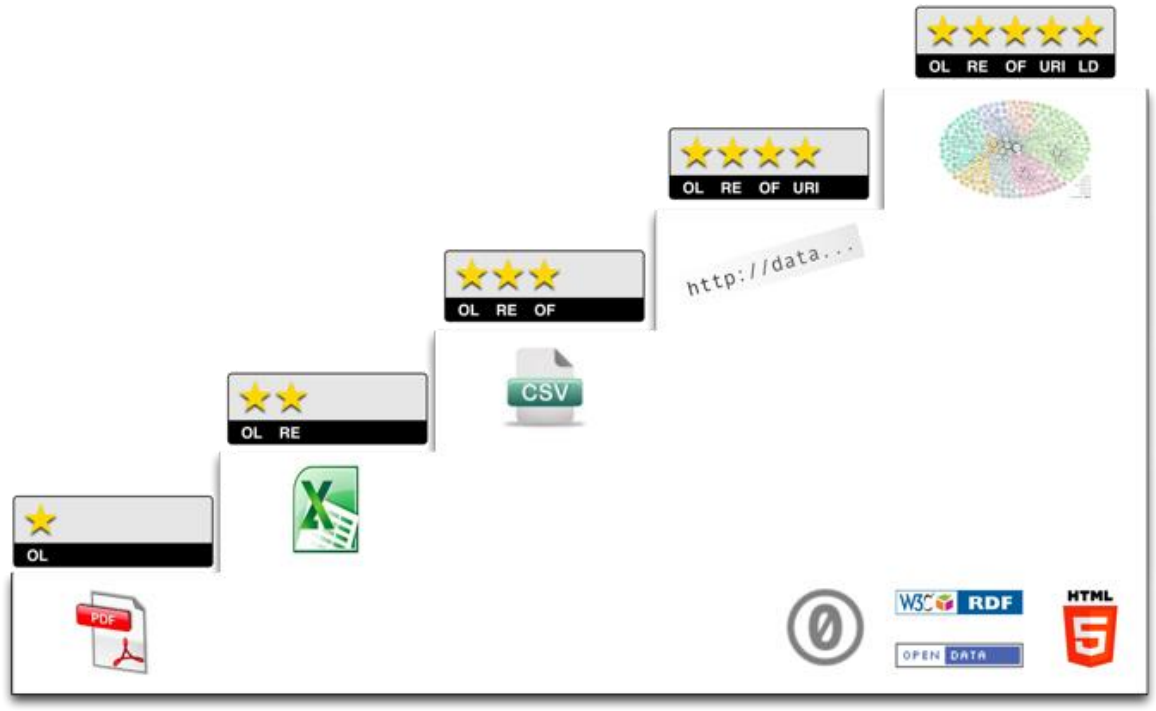

Figure 1: 5 star schema, source: (Hausenblas, 2012)

With the increasing openness level data is more easily processed in software applications and more interlinked with other data providing a rich context. One star data is data in basically any format but it must be openly licensed. Two star data are machine-readable but the format might be proprietary. This affects the end user because s/he might need to purchase some specific piece of software to be able to process the data. Three star data is machine-readable and it uses open formats. This helps to free the end uses from using proprietary applications and allows them to use their favourite software for processing the data. As a publisher you should avoid proprietary formats and provide at least three star open data. Four star data means that URIs[1] are used as object identifiers. This allows pointing to a single object rather than the whole dataset. Five star data is the highest level of openness in the 5 star scheme. Five star data fully complies with the Linked Data principles (see Berners-Lee, 2006) and it is linked to other data to provide context. You can find more detailed guidelines for publication of either three star or five star data in the Methodology.

You can also ask yourself a question, why you should be publishing Open Data. Well, there is no single right answer to that question. In fact it will also depend on your view of the topic and your goals and priorities. It should be noted that your strategic decision and clearly set goals are very important attributes of your Open Data initiative because it will affect what datasets you will publish and how you will measure success.

There are various studies discussing benefits of opening up data that might also represent motivations for Open Data publication. We are going to name some of these studies in order to show what possible benefits Open Data can bring.

[1] Uniform Resource Identifier, http://tools.ietf.org/html/rfc3986

According to an international study by Logica Business Consulting (2012) possible motivations for open data might include:

- Increase transparency;
- Stimulate economic growth;
- Improve government services and responsiveness;
- Encourage reuse;
- Improve public relations and attitudes toward government;
- Improve government data and processes.

Motivations for opening up data were also analysed in Spain as a part of the eEspaña 2014 study. The following motivations were studied among the regional governments and municipalities (Gimeno et al., 2014):

- Improving the quality of citizen services;
- Generating new jobs;
- Improving the perception of transparency;
- Stimulating the economy;
- Improving the efficiency of the institution;
- Reducing costs of the institution;
- Economic development;
- Provide new value-added services;
- Foster innovation and competitiveness.

Although there are some differences, motivations for opening up data named in the abovementioned studies are similar. The motivations are also not mutually exclusive. Encouraging reuse of open data and generating new jobs might both help to stimulate economy.

Open data might be used to inform citizens about actions, activities and policies of government. This might help to increase transparency and trust of citizens. According to Bauer and Kaltenböck (2011) open government data is one of the key enablers of the open government, a movement aimed at establishing modern government which builds on cooperation among politicians, public administration, industry and citizens.

Open data might be also used to better inform citizens and other stakeholders about government services and feedback obtained from the users of the data might be used to improve the published data as well as processes in the public sector bodies (Logica Business Consulting, 2012). Therefore, open data is not only beneficial to the users outside of the public sector but it can also help to increase government efficiency.

Open data might also represent a valuable resource that can be reused to develop new and innovative services and applications. For example the Open Data 500 study analyses reuse of open data by the US companies.[2] Buchholtz, Bukowski and Śniegocki (2014) view open data as complementary to the Big Data trend. According to their study open data by itself can bring benefits to the European economies but even more value can be gained by data-driven decision making which will utilize both open and big data (Buchholtz et al., 2014).

[2] http://www.opendata500.com/

3 Methodology overview

Methodology for publishing datasets as open data (hereafter the Methodology, see Nečaský et al., 2014a) is a generic methodology for publishing open data that provides answers to questions such as: how to identify unique resources in the datasets, how to reuse well known codebooks/vocabularies/ontologies (currencies, NUTS codes, ..), in which formats the data should be published so that they are machine readable, how the data should be transformed (e.g. anonymized) before being published, which descriptive and provenance metadata should be published together with the dataset (such as name, format, location/URL, source, responsible person), etc. The Methodology is intended mainly for data owners and publishers (mainly public bodies) and it is made of 5 main building blocks that are depicted on figure 2.

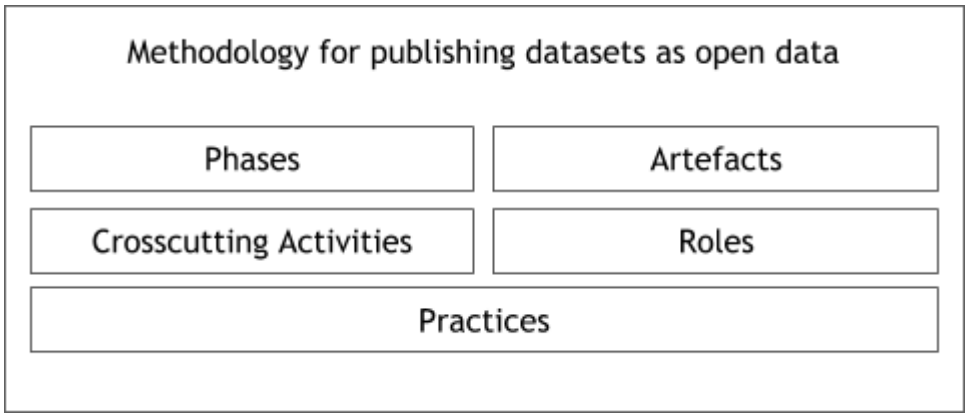

Figure 2: Methodology building blocks

Phases represent the stages of the open data publication process and they reflect the lifecycle of an open dataset. The following phases of the open data publication process were defined:

1. (P01) Development of open data publication plan,
2. (P02) Preparation of publication,
3. (P03) Realization of publication,
4. (P04) Archiving.

There are also some activities that should be performed in every phase of the open data publication process - cross-cutting activities. There are four cross-cutting activities in the methodology:

1. (CA01) Data quality management;
2. (CA02) Communication management;
3. (CA03) Risk management;
4. (CA04) Benefits management.

Relationship between the phases and the cross-cutting activities are depicted on figure 3.

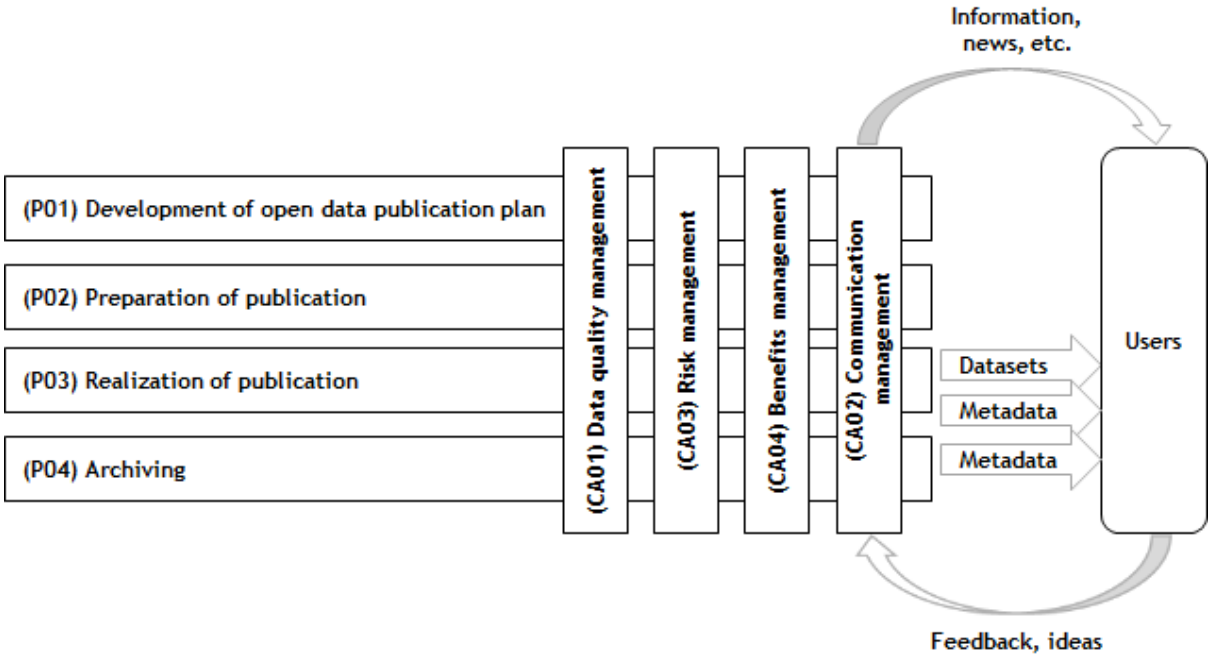

Figure 3: Phases, cross-cutting activities and the feedback loop in the Methodology

The cross-cutting activity (CA02) Communication management ensures that communication between the Open Data publishers and users is a part of the activities related to the Open Data publication. Due to the cross-cutting nature of the activity early feedback of the potential users might be one of the inputs into selection of datasets to be published which in turn can lead to more user-centric release of open datasets.

Both the phases and the cross-cutting activities are further divided into a set of tasks. Tasks represent steps in the publication process. For each of the tasks practices are described that provide more detailed guidelines how the tasks should be performed.

For each of the tasks responsibilities are set. There are 10 roles that are either responsible, accountable, consulted or informed about the result of some task. The following roles are described in the Methodology:

- **Publisher** – A person or an entity that makes a dataset available to the public, e.g. s/he makes the dataset available for download through a website.
- **Owner** – A person or an entity that owns a dataset, i.e. s/he holds rights to the dataset or s/he is legitimate to make decisions about the dataset. Owner is legitimate to decide that a certain dataset will be published as open data and s/he is also legitimate to licence the dataset.
- **Curator** – A person or an entity that curates or maintains a dataset and its catalogue record (metadata). Curator keeps the published datasets accurate and up-to-date. Curator is usually appointed to his or her role by the data owner.
- **OD Catalogue Owner** – OD Catalogue Owner is an organization responsible for the Open Data catalogue. It is responsible for defining policies and rules governing cataloguing and its use. OD Catalogue Owner also selects the data cataloguing software.

- **OD Coordinator** – OD Coordinator is a person in an organization that is responsible for coordination and management of the open data related activities of the organization.
- **OD Catalogue Publisher** – OD Catalogue Publisher is a person or an organization that makes the data catalogue available to the potential users. OD Catalogue Publisher is responsible for operating the data catalog and s/he ensures maintenance of the underlying IT infrastructure.
- **IT Professional** – IT Professional is a person with skills and knowledge in the domain of the information and communication technologies. S/he provides support to other roles, develops and tests the ETL procedures and performs the transformation of the data into the target data formats. If Linked Open Data is published, IT Professional should have the necessary skills and knowledge for publication of LOD, e.g. RDF, ontologies, URI design, linked data software tools.
- **Data Quality Manager** – Data Quality Manager is a person with skills and knowledge in the domain of data quality. S/he is in charge of supervisioning all data quality components and the data quality lifecycle. Typically s/he is one of the supporter of the OD coordinator.
- **Data Quality Expert** – Data Quality Expert is a person with skills and knowledge in the phases and tools of the data quality lifecycle. S/he is in charge of analyzing, applying and in case creating the ETL components related to the data quality. In case of small organization such role can be performed by the data quality manager itself.
- **Legal Expert** – Legal expert is a person with skills and knowledge in the domain of law and legislation. S/he provides his or her expertize about licensing of open datasets and s/he is involved in analysis of datasets where his or her expertize is required in order to ensure that the publication of datasets comply with the legislation.

Artefacts are the last building block of the Methodology. One artefact might be both input and output at the same time. Usually output of one task becomes an input into the subsequent task. However in certain situations an artefact might be both input and output of just one task, i.e. in situation when some artefact is updated by some task.

Methodology for publishing datasets as open data is made up by the following parts:

- methodology overview (Nečaský et al., 2014a) – describes the overall concept of the methodology;
- master spreadsheet (Nečaský et al., 2014c) – provides definition of phases, cross-cutting activities, tasks, roles and their responsibilities and artefacts (inputs/outputs);
- documentation of practices (Nečaský et al., 2014b) – describes practices for each of the tasks.

4 Step-by-step guides to the Open Data publication

As discussed in the previous section tasks of the Open Data publication are organized into phases and cross-cutting activities. In the following sections we provide step-by-step guides for implementation of these phases and cross-cutting activities. The aim is to highlight the most important steps rather than to provide comprehensive guidelines. For details and the recommended practices please refer to the Documentation of practices that is a part of the Methodology (Nečaský et al., 2014b).

4.1 Phases

4.1.1 (P01) Development of open data publication plan

Description: The goal of this phase is to perform identification and analysis of candidate datasets for opening up and to develop an open data publication plan. The open data publication plan mainly describes what datasets will be published as open data, what the target openness level of the datasets is, what roles are involved in the publication of open data and what responsibilities they have. This plan also describes what are the terms and conditions under which the open datasets should be published. Finally, the plan sets a roadmap for publication of the selected datasets.

In order to develop you open data publication plan you should:

- **Analyse your data sources** – identify what data you process and maintain in your organization.
- **Identify datasets for opening up** – select datasets that can be possibly made available as Open Data and describe them with basic metadata such as title, description, responsible person or organizational unit and periodicity
- **Determine the target level of openness for the selected datasets** – using the 5 star scheme proposed by Berners-Lee (2006) and Hausenblas (2012) decide what it the target openness level for your datasets. You should aim at least for three star data, i.e. machine-readable data in open formats such as CSV or XML.
- **Estimate effort** – for each of the candidate datasets estimate effort or costs of making them available as Open Data. Factors that you should take into account include dataset complexity, anonymization of data if necessary, manual operations performed with the dataset, dataset size and target periodicity of publication.
- **Develop your open data publication plan** – set priorities for the candidate datasets taking into account demand for data, benefits and risks resulting from release of the particular datasets and effort/costs needed to release the datasets. Based on the priorities select the final set of datasets to be published and set. Identify other activities needed to enable publication of datasets. Develop an open data publication roadmap including both the planned releases of datasets and other planned activities.

4.1.2 (P02) Preparation of publication

Description: The goal of this phase is to prepare the datasets identified in the open data publication plan for publication according to the specified openness level. The key activities of this phase include configuration of access to the data sources, description of the datasets and creation of the related metadata, selection and implementation of the software tools that are needed in order to perform the open data publication effectively. The design of an approach to publication of the identified datasets as well as the design of the anonymization procedures are performed during this phase as well. This phase also involves design, development and testing of the ETL procedures and selection of the appropriate licence that will ensure legal openness of the published datasets.

In order to prepare your datasets for publication you should:

- **Configure access to the data sources** – data that you are going to release as Open Data are usually stored in some primary data store in your organization. In order to be able to prepare the datasets for publication, primary data sources need to be accessible to those participating in the preparation or to the automated ETL procedures.
- **Define the catalogue record schema and the target data catalogues** – select suitable schema of your catalogue records. To ensure interoperability is recommended to base your schema on DCAT[3] or DCAT-AP[4] vocabulary that can be extended with VoID[5] if necessary. You need to decide whether you will establish your own data catalog and how you can catalogue your datasets in national or other official data catalogues.
- **Describe the datasets with metadata** – describe your datasets with metadata according to the selected metadata schema. Attention should be also paid to the description of the codelists and schemas of your datasets or vocabularies used to represent your data as it helps others to understand the structure and semantics of your data.
- **Select and implement software tools** – publication of Open Data can be supported by software tools such as data portal solution or data extraction, transformation and loading (ETL) software. Determine, acquire and implement software tools to support your Open Data initiative.
- **Define approach to the dataset publication** – in this step you should specify how the data quality is going to be ensure, how the data is going to be anonymised if necessary, how the related datasets are going to be interlinked and your datasets are going to be updated over time.
- **Design and implement ETL procedures** – if possible and feasible we recommend automation of the preparation of datasets with ETL procedures. In this step you

[3] http://www.w3.org/TR/vocab-dcat/

[4] https://joinup.ec.europa.eu/asset/dcat_application_profile/home

[5] http://www.w3.org/TR/void/

should design and implement ETL procedures that perform the necessary transformations of data.

- **Test ETL procedures** – in order to ensure that the ETL procedures provide the desired results it is recommended to test the implemented ETL procedures.
- **License your datasets** – terms of use of an open datasets should be clearly stated. Determine whether your data is a part of the public domain or whether it needs to be licensed. It is recommended to use standard licenses such as the Creative Commons 4.0 licenses. Creative Commons CC0 or the Creative Commons Attribution 4.0 (CC-BY 4.0) are the recommended options.

4.1.3 (P03) Realization of publication

Description: The goal of this phase is to perform the publication of the identified datasets as open data according to the approach designed in the previous phase. Developed ETL procedures will be used to perform the actual transformation of the data into the target open formats. This phase also involves cataloguing of the published open datasets as well as their regular updates and maintenance.

In order to execute publication of your open datasets you should:

- **Publish your datasets** – make your datasets available for download or accessible via API or via dereferencing of URIs. If you implemented ETL procedures to automate some of the steps, execute and monitor the ETL procedures.
- **Catalogue your datasets** – in order to make your datasets discoverable you should also publish the metadata. If necessary update the catalogue records prepared during the Preparation of publication step and make them available in the specified data catalogues, e.g. data catalogue of your organization or the national or other official data catalogue.
- **Maintain your datasets** – unless your data is static and never gets out of data you should regularly updated your datasets. Issues and errors in the data might be revealed after some dataset had been released. If possible issues should be resolved and errors fixed in order to ensure the desired level of the data quality. Feedback provided by the users should be used for improvements of the data quality.

4.1.4 (P04) Archiving

Description: The goal of this phase is to manage end-of-life stage of the dataset lifecycle. Activities of this phase are triggered when it is no longer possible to maintain or even make available some previously published open datasets, e.g. due to the changes in legislation. Regular updates and maintenance of such datasets is terminated during this phase. If it is no longer possible or allowed to make some dataset available its publication is stopped and its availability is cancelled. Users of the datasets should be informed about changes in update policy or availability of the datasets. Guidelines on this topic are provided as a part of the description of the cross-cutting activity CA02 Communication.

In order to manage the end-of-life stage of your datasets lifecycle you might need to:

- **Terminate the dataset maintenance** – if the primary data of the dataset is no longer collected or if the structure or meaning of the data significantly changes so that it is necessary to create a new datasets, maintenance of the dataset should be terminated. If it is possible and feasible the previously published data should be kept

available for those who reuse it, i.e. the dataset should be archived. Update the catalogue record with information that the dataset is no longer maintained and if relevant add links to the catalogue record to the datasets that supersede the unmaintained dataset. Inform users about the change in the datasets status.

- **Terminate the dataset publication** – if it is no longer possible to publish some dataset, its publication should be terminated. Requirement to terminate publication of some dataset might result for example from the changes in legislation or from the ruling of the court. Block the public access to the dataset and its resources. Update the catalogue record with information that the dataset is no longer available and if relevant add links to the catalogue record to the datasets that supersede the unmaintained dataset. Inform users about the change in the datasets availability.

4.2 Cross-cutting Activities

4.2.1 (CA01) Data quality management

Description: The goal of this cross-cutting activity is to ensure quality assessment and improvement of data through a set of phases and decision points. This activity provides a set of guidelines and techniques that, starting from input information describing a given application context, defines a rational process to assess and improve the quality of data.

In order to manage quality of your open datasets you should:

- **Analyse the data quality requirements** – determine what data quality dimensions are important to you and also possibly to the users of your data. Based on the results of this analysis, specify what is the target quality level of relevant data quality metrics.
- **Perform the data quality assessment** – assess what is the current level of quality of the provided data and metadata.
- **Improve the quality of your data** – specify and implement the necessary steps to improve the quality of your data and metadata. This might include correction of the identified error in datasets and reengineering of the processes in which the data is being collected and processed. The latter strategy should be aimed at identification and correction of the root-cause of errors in the data.

4.2.2 (CA02) Communication management

Description: The necessary communication between the owner/publisher of the open datasets and their (potential) users is performed in this cross-cutting activity. The main goal is to ensure that the users are informed about availability of the datasets, that the reuse of the published dataset is promoted and the the users can provide feedback. Identification and classification of the users into different groups might be performed as a part of the communication activities. This classification might be used to better target the communication campaigns. Feedback provided by the users should be analysed and it must be ensured that questions of the users are promptly answered. In case that some dataset reaches its end-of-life, users of such dataset must be informed about the change in the dataset status or availability.

In order to manage your communication activities and to communicate with the users of your data you should:

- **Identify the potential user groups** – the goal of this activity is to identify the potential users of the open datasets and classify them into the user groups. This classification allows tailoring of the communication campaigns for different types of users.
- **Develop your communication strategy** – communication strategy is the main tool for managing open data related communication. Communication strategy should specify appropriate communication channels for the relevant user groups, planned communication campaigns and competencies and responsibilities regarding the communication about Open Data in your organization. Communication must not be one-way only but the feedback loop must be established, i.e. users must be able to provide feedback.
- **Engage users during development of the OD publication plan** – the main objective of this activity is to obtain information about the demand for data which in turn could be used during selection and prioritization of datasets for opening up.
- **Set up the communication channels defined in the communication strategy** – communication channels described in the communication strategy need to be implemented. For example a new service might be added to the publisher's website that will allow users to subscribe and receive news about the publisher's open data.
- **Prepare the communication campaign** – prepare campaigns in which you will inform the users about your Open Data initiative. If the potential users are not properly informed it might take longer for them to notice that some open dataset are available. Therefore communication campaigns should accompany the initial publication of datasets and also every significant addition to your data supply.
- **Inform about progress** – this task should be executed especially when the preparations will take a significant amount of time and its goal is to inform the potential users about important achievements and progress in preparation of datasets for publication. This tasks helps to keep users informed that the open data initiative is still ongoing.
- **Inform about open data** – execute the planned communication campaigns and other communication activities described in the communication strategy in order to inform users about open data related events, especially about the newly published datasets and changes in their publication and availability.
- **Analyse the user feedback** – feedback provided by the user should be systematically collected and evaluated. This task should help you to make your Open Data initiative user-centric which in turn should foster the reuse of the published datasets and thus help you to achieve the desired benefits.
- **Inform about the termination of maintenance or publication of a dataset** – communication activities in the archiving phase should mainly focus on informing users about planned and performed termination of maintenance of datasets or termination of publications of datasets. Users should be informed about these significant changes in advance. If there are any datasets that supersede unmaintained or no longer published datasets users should be informed about such datasets as well.

4.2.3 (CA03) Risk management

Description: The goal of this cross-cutting activity is to ensure that possible risks related to the publication of open data are properly managed. Key risk management activities include risk analysis and development of the risk mitigation plan. Competencies and responsibilities for the identification of risks, monitoring and evaluation of risks, regular actualization of the risk register and for implementation of the risk mitigation actions should be set. Last but not least, it is necessary to ensure cost effective risk response.

In order to manage risks related to the Open Data publication you should:

- **Identify and analyse the potential risks** – identify and analyse potential risks related to the publication of a dataset as open data. Evaluate the probability and impact of each of the identified risks. Identified risks for each of the datasets as well as the results of their evaluation should be recorded in the risk register.
- **Develop your risk mitigation plan** – risk mitigation plan sets competencies and responsibilities for the open data risk management and it contains description of the defined risk mitigation actions. Risk mitigation plan can be a part of the open data publication plan.
- **Update the risk register with the information acquired during the preparation of the datasets** – new information relevant to the open data risk management might be obtained during the preparation of publication phase. Based on new or updated information about datasets the risk register should be updated or extended with newly identified risks.
- **Implement the risk mitigation actions relevant to the preparation of the datasets** – implement the risk mitigation actions that fall into the preparation of publication phase. For example validation of the defined anonymisation procedures or description of datasets with the additional information that is necessary for their correct interpretation.
- **Monitor and manage risks** – monitor risks and ensure proper risk management. Regularly assess the Open Data related risks and update the risk register and the risk mitigation plan.
- **Respond to events** – ensure that the risk related events are handled properly. In case of incidents, take actions to minimize the negative impacts (losses, damage). Every incident and the actions taken to minimize the impact should be recorded. Incident reports should also contain description of the proposed changes to the risk mitigation plan in order to prevent similar incidents in the future.
- **Report about the state of the open data related risks** – inform the relevant stakeholders are about the current state of the identified risk related to open data publication and about the planned mitigation actions. Especially the open data coordinator, owner and the published of the data should be informed about the current state of the risks.

4.2.4 (CA04) Benefits management

Description: The goal of this cross-cutting activity is to ensure that the expected benefits resulting from publication of open data are properly managed across the phases of the open data publication. This cross-cutting activity therefore involves tasks aimed at identification

and evaluation of the potential benefits, continuous monitoring of the benefits and assessment of the achieved results.

In order to manage the expected benefits of the Open Data publication you should:

- **Identify and analyse the potential benefits** – identify potential benefits that can be achieved by publication of Open Data and assess the amount of benefits using quantitative and qualitative methods.
- **Develop your benefits management plan** – develop your open data benefits management plan. This plan should describe the competencies and responsibilities with regard to the management of the open data benefits as well as the actions that need be taken in order to create the necessary prerequisites for achievement of the expected benefits.
- **Update the benefits management plan according to the information acquired during the preparation of the datasets** – new information relevant to the open data benefits management might be obtained during the preparation of publication phase. Based on the new or updated information about datasets the benefits register should be updated.
- **Monitor and manage benefits** – continuously gather and evaluate information relevant for tracking of the identified benefits. Implement the actions needed to achieve the identified benefits. Regularly update the benefits register and the benefits management plan.
- **Report about the open data related benefits** – inform the relevant stakeholders about the current state of the open data benefits and about the planned actions to foster their achievement. Especially the open data coordinator, owner and the published of the data should be informed about the current state of the open data benefits.

References

BAUER, Florian & KALTENBÖCK, Martin, 2011. *Linked Open Data: The Essentials*. Vienna: edition mono/monochrom. pp. 62. ISBN 978-3-902796-05-9.

BERNERS-LEE, Tim, 2006. Linked Data - Design Issues. In: *Design Issues* [online]. 2006-07-27 [cit. 2014-07-10]. Available from: http://www.w3.org/DesignIssues/LinkedData.html.

BUCHHOLTZ, Sonia, BUKOWSKI, Maciej, ŚNIEGOCKI, Aleksander, 2014. Big and open data in Europe. A growth engine or a missed opportunity? In: *Big and open data in Europe* [online]. [cit. 2014-07-10]. Available from: http://www.bigopendata.eu/wp-content/uploads/2014/01/bod_europe_2020_full_report_singlepage.pdf.

DULONG DE ROSNAY, Mélanie, TSIAVOS, Prodromos, ARTUSIO, Claudio, ELLI, Jo, RICOLFI, Marco, SAPPA, Cristiana, VOLLMER, Timothy, TARKOWSKI, Alek, 2014. D5.2. Licensing Guidelines. In: *LAPSI 2.0* [online]. 25th February 2014 [cit. 2015-21-09]. Available from: http://www.lapsi-project.eu/sites/lapsi-project.eu/files/D5.2LicensingGuidelinesPO%20(1).pdf.

GIMENO, Manuel, URIARTE, Blanca Villamía, SAA, Víctor Suárez, 2014. *eEspaña 2014*. Madrid: Fundación Orange. pp. 252. ISSN 2174-3886.

HAUSENBLAS, Michael. 2012. 5 star Open Data. In: *5 star Open Data* [online]. 2012-04-03 [cit. 2014-07-10]. Available from: http://5stardata.info/.

Logica Business Consulting, 2012. Open data and use of standards: Towards a Better Supply and Distribution Process for Open Data. In: *Standardization Forum* [online]. 2012 [cit. 2014-06-20]. Available from: http://www.forumstandaardisatie.nl/fileadmin/os/documenten/Internationale_benchmark_v1_03_final.pdf.

NEČASKÝ, Martin; CHLAPEK, Dušan; KLÍMEK, Jakub; KUČERA, Jan; MAURINO, Andrea; RULA, Anisa; KONEČNÝ, Miroslav a VANOVÁ, Lenka. 2014a. Deliverable D5.1: Methodology for publishing datasets as open data. In: *COMSODE* [online]. 31. 7. 2014 [cit. 2015-09-20]. Available from: http://www.comsode.eu/wp-content/uploads/D5.1-Methodology_for_publishing_datasets_as_open_data.pdf.

NEČASKÝ, Martin; CHLAPEK, Dušan; KLÍMEK, Jakub; KUČERA, Jan; MAURINO, Andrea; RULA, Anisa; KONEČNÝ, Miroslav a VANOVÁ, Lenka. 2014b. Deliverable D5.1: Methodology for publishing datasets as open data, Annex 1: Documentation of Practices. In: *COMSODE* [online]. 31. 7. 2014 [cit. 2015-09-20]. Available from: http://www.comsode.eu/wp-content/uploads/Annex1_D5.1-Documentation_of_practices.pdf.

NEČASKÝ, Martin; CHLAPEK, Dušan; KLÍMEK, Jakub; KUČERA, Jan; MAURINO, Andrea; RULA, Anisa; KONEČNÝ, Miroslav a VANOVÁ, Lenka. 2014c. Deliverable D5.1: Methodology for publishing datasets as open data, Annex 2: Methodology master spreadsheet. In: *COMSODE* [online]. 31. 7. 2014 [cit. 2015-09-20]. Available from: http://www.comsode.eu/wp-content/uploads/Annex2_D5.1-Methodology_Master_Spreadsheet.xlsx.

OPEN KNOWLEDGE, [2014]. Open Definition 2.0. In: *Open Definition* [online]. [cit. 2015-03-10]. Available from: http://opendefinition.org/od/.

OPEN KNOWLEDGE, 2015. What is Open Data? In: *Open Data Handbook* [online]. [cit. 2015-09-21]. Available from: http://opendatahandbook.org/guide/en/what-is-open-data/.