



Government of Flanders in Belgium

Update - Q4 2014

English Translation – Q1 2015

Open Data Manual

Practice-oriented manual for the publication and management of Open Data using the Flemish Open Data Platform

Contact: Noël Van Herreweghe

Open Data Program Manager - Government of Flanders in Belgium



<http://www.vlaanderen.be/opendata>

info@opendataforum.info – noel.vanherreweghe@bz.vlaanderen.be

+32 473 881872

Skype: noelvan

License: CC-BY

PREFACE

December 2010, the Flemish Minister for Public Governance, Geert Bourgeois, organised the “Flanders in Action round table *i-Vlaanderen*” on open government and ICT. One of the main conclusions was that the participants were very strongly in favour of opening up more public sector data.

Open Data is data or information collected by the government as part of its public task, with minimal or no constraints, electronically available and using open standards.

13 June 2013, the revised directive on the re-use of public sector information was adopted by the European Parliament. The new directive makes it easier to open up public sector information and emphasises that such information contributes to economic growth and job creation in the EU Member States.

Open Data is also the standard within the Flemish public administration. The Flemish public administration has a wealth of information at its disposal in a large variety of areas. Although this information is currently publicly available, re-use is not always allowed. Re-use implies that the data can be used for both non-commercial and commercial purposes, either free of charge or at a reasonable cost. When talking about re-use, we refer to the use of data in smartphone applications (Apps), websites, research, etc. These types of re-use already exist today. However, re-use gets a huge impulse when public sector information becomes freely available as Open Data.

However, opening up information as Open Data can be challenging, raising issues such as copyright, intellectual property and liability. Therefore, information, consultation and communication are essential.

A first step in the right direction was the creation of the Open Data Forum. This forum is a knowledge exchange platform and gives access to the Open Data Portal of the Flemish public administration (<http://www.vlaanderen.be/opendata>). The knowledge platform aims to promote knowledge and best practice sharing on Open Data between partners from within and outside the Flemish public administration.

The present manual serves as a practical guide, offering you background information on Open Data in addition to a number of guidelines on how to actually make Open Data available.

Noël Van Herreweghe,
Government of Flanders Open Data Programme Manager

TABLE OF CONTENTS

1/ In general	6
1.1 / Open Data	6
1.2 / Purpose of this manual	6
1.3 / Policy and legislation	6
1.4 / The Flemish Open Data Forum and the Flemish Open Data Platform	7
2/ Choose your dataset(s)	8
2.1 / General guidelines	8
2.2 / Prior legal review	8
2.2.1 / Intellectual property rights	8
2.2.2 / Exceptions to open government	9
2.2.3 / Protection of personal data	10
2.3 / Criteria and priorities	10
2.3.1 / Legal requirements	11
2.3.2 / Previously published data	11
2.3.3 / Value of the data	11
2.3.4 / Reach of the data	12
2.3.5 / Priorities	12
3/ Choose a model licence	14
3.1 / Prior decision on charges	14
3.2 / Prior decision on potential use categories	14
3.3 / Choice of model licence	15
4/ Open up your source data	16
4.1 / Underlying idea: use of ETL techniques	16
4.2 / Scenarios for opening up source data	17
4.2.1 / Scenario 1: Starting from an existing publication	18
4.2.2 / Scenario 2: Starting from an existing dataset	20
4.2.3 / Scenario 3: Starting from a database	21
4.2.4 / Scenario 4: Starting from an existing source system	23
4.2.5 / Scenario 5: Starting from different source systems	25
4.3/ Opening up geographical and location-based information in Flanders	26
4.3.1/ Positioning	26
4.3.2/ Guideline for the publication of geographical data	27
4.3.3/ Technical impact	28
4.4/ How to get started with Open Data in your projects?	29
5/ Structure your dataset(s) and choose an open format or API	31
5.1 / Minimum requirements	31
5.1.1 / Quality	31
5.1.2 / Consistency	31
5.2 / Standards, open formats and APIs	32
5.3 / Linked Open Data	34
5.4 / Maturity model for Open Data	34
6/ Publish your dataset(s)	35
6.1 / Prior technical review	35
6.2 / Through in-house website (data dump or API) or upload to CKAN	35
7/ Document your dataset(s)	36
7.1 / Open Data contact point	36
7.2 / Contact address for information and feedback	36
7.3 / Terms and conditions of use	37
7.4 / Charges	37
7.5 / Guarantees as to availability	38
7.6 / Instructions for use	38
7.6.1 / The data itself	39
7.6.2 / The process	39
7.7 / Website language	39

8/ Make your dataset(s) discoverable	40
8.1 / Metadata	40
8.1.1 / What is metadata?	40
8.1.2 / How to produce metadata?	40
8.1.3 / Metadata guidelines	41
8.2 / Adding metadata in CKAN	42
8.2.1 / Creating an account	42
8.2.2 / Creating a dataset	42
8.2.3 / Adding metadata through the CKAN API	49
9/ Evaluate your Open Data practice	50
Annex 1: Master Data Management	51
Annex 2: Adding metadata through the CKAN API	55
Annex 3: Metadata DCAT-CKAN	59
Annex 4: Overview of recommendations	62
Glossary	6

1 IN GENERAL

1.1 OPEN DATA

Open Data is data or information collected by the government as part of its public task, with minimal or no constraints, electronically available and using open standards.

Open Data ensures greater transparency on government's activities, promotes efficiency, both within and outside, the government and leads to an innovative re-use of public data by citizens, businesses and organisations.

Open Data also fosters entrepreneurship, results in the development of innovative products and services, offers instruments for alternative decision-making in the field of management, planning and science, and contributes to the creation of a Flemish knowledge economy.

In addition, Open Data generates added value for the government itself, including better public service, administrative burden reduction and increased interaction and cooperation with citizens, businesses and organisations. The centralised provision of Open Data will also result in efficiency gains for the public sector bodies themselves and help improve data quality.

1.2 PURPOSE OF THIS MANUAL

In terms of content, this manual aims to provide guidance to anyone who wants to open up data or information as Open Data and thus has a need for a step-by-step plan which may serve as a guideline for introducing and shaping an Open Data strategy within their organisation. Such a strategy describes the whole process from selecting data to being operational with Open Data, from the source right through to the publication of Open Data on the Open Data Platform.

However, this document is not exhaustive. Updates of this document will be announced on the Open Data Forum (www.vlaanderen.be/opendata). This knowledge platform aims to promote knowledge and best practice sharing on Open Data between partners from within and outside the Flemish public administration. This forum provides among other things blogs, news, events and interesting links. The menu button "Data" on this forum gives access to the Flemish Open Data Portal which, in its turn, grants access to a large number of datasets that are available for re-use as Open Data, also for commercial purposes, either free of charge or against a reasonable charge.

The content of the manual as well as the Open Data Licences - approved by the Board of Senior Administration Officials on 12 December 2013, see <http://www.bestuurszaken.be/modellicenties-bij-het-aanbieden-van-open-data> - is included in this manual.

1.3 POLICY AND LEGISLATION

23 September 2011, the ministers of the Government of Flanders signed a "concept agreement" on Open Data - <http://www.bestuurszaken.be/conceptnota>. This paper outlines the policy context and regulatory framework for Open Data in Flanders and contains a vision and the main strategic lines for Open Data.

The following six main strategic principals are defined in this paper:

1/ Open Data is the standard within the Flemish public administration.

Open Data is the standard. Closed data will only be allowed under the condition that an explicit justification is given.

2/ Re-use of Open Data is allowed.

Open Data can be re-used, also for commercial purposes, either free of charge or against a reasonable cost. To that end, Open Data will use simple, standardised licence models.

3/ Open Data applies open standards.

Open Data is not really open unless open standards are applied. Open Data implies open formats and open interfaces.

4/ Open Data should originate from authentic data sources, whenever possible.

The development of Flemish authentic data sources will lead to reliable and high-quality public sector data and information.

5/ Open Data adopts an integrated approach.

The local authorities in Flanders are also major suppliers of data. Moreover, the link with the federal level should not be forgotten. Cooperation across all levels of government offers high added value.

6/ Information on the management processes within the Flemish public administration in a central directory.

Data with regards to management processes can be made available as Open Data following a decision by the Government of Flanders.

The concept paper resulted in a "Flemish Action Plan Open Data", adopted by the Government of Flanders on 19 July 2013 - <http://www.bestuurszaken.be/vlaams-actieplan-open-data> - and includes a number of actions. The action plan seeks to establish a Flemish Open Data policy, building on the main strategic lines from the concept paper and highlighting a number of aspects: the added value for the government itself, the socio-economic added value and the interaction with other projects and organisations.

The overall Government Coalition Agreement 2014-2019¹ also defines the ambition to see Open Data as the standard within the Flemish public administration and to implement the concept at an accelerated pace.

When opening up data for re-use by third parties, the entities of the Flemish public administration must comply with existing legislation, including:

- The Flemish Parliament Act on Open Government
- The Flemish Parliament Act on Re-use of Public Sector Information
- The Copyright Act and the Act on the Legal Protection of Databases
- The Act on the Protection of Privacy on Processing of Personal Data (Privacy Act)

On 13 June 2013, the revised directive on re-use of public sector information was adopted by the European Parliament. In accordance with the new directive, data with no or minimal constraints regarding privacy, security, patents, copyright, deadlines or other, should become the standard within the EU Member States. Following this revised Directive, the Flemish legal framework for Open Data will be adapted in the course of 2014.

1.4 FLEMISH OPEN DATA FORUM AND FLEMISH OPEN DATA PLATFORM

The Open Data Forum (<http://www.vlaanderen.be/opendata>) is foremost a knowledge platform. It is designed for sharing knowledge and best practices on Open Data between public sector bodies and stakeholders within and outside the Flemish public administration. On this forum the latest news can be found, as well as blogs, upcoming events and interesting links regarding Open Data.

The menu button "Data" - operational since February 2014 - on this forum gives access to the Flemish Open Data Platform. This platform gives access to a large number of datasets available for re-use. At this moment (October 2014), more than 1,200 datasets are available as Open Data. The Flemish public administration has been using the CKAN software, an open source tool which is used worldwide by various public sector bodies. This portal is integrated with the DATATANK, an Open Data management system which will substantially extend the functionalities of the Flemish Open Data Platform.

¹ <http://www.vlaanderen.be/publicaties/detail/het-regeerakkoord-van-de-vlaamse-regering-2014-2019>

2 CHOOSE YOUR DATASET(S)

One of the points often made is that one does not really know which datasets should be opened up (first) and in what order. In this chapter we give a number of tips for helping you make a selection.

2.1 GENERAL GUIDELINES

How to get started with Open Data? The Open Data Handbook by OKFN, the Open Knowledge Foundation (http://opendatahandbook.org/nl_BE/), may provide directions in this respect.

There are four key steps to be followed when opening up data.

- 1/ Choose your dataset(s).** Choose the dataset(s) you plan to make available as Open Data.
- 2/ Apply an open licence.** Determine what intellectual property rights exist in the data. Apply a suitable 'open' licence that grants the necessary rights and supports the definition of openness. The Flemish public administration has defined generic Open Data licences: (<http://www.bestuurszaken.be/modelllicenties-bij-het-aanbieden-van-open-data>).
- 3/ Make the data available.** In bulk and in a useful format. You may also wish to consider alternative ways of making data available, like, for instance, via an API. The use of an API is recommended when two-way and/or real-time communication is required. In this case, the API is provided on the Open Dataset.
- 4/ Make the data discoverable.** Post a reference to the dataset or the dataset itself on the Flemish Open Data Platform.

From a policy-based focus, the following step-by-step plan may be used:

- 1/ Create support.** Convince the management, record the policy decision and incorporate Open Data into your information policy plan.
- 2/ Analyse.** Convince co-workers and examine and describe added-value. Take legal implications into account.
- 3/ Test.** Set up a pilot.
- 4/ Secure.** Communicate your experiences.
- 5/ Evaluate** the process and adjust it, if necessary.

2.2 PRIOR LEGAL REVIEW

Before opening up data to the general public, check whether there are any legal constraints. In this context the three categories of constraints to consider are intellectual property rights, legislation regarding the processing of personal data, and possible exceptions on open government, such as state security.

2.2.1 INTELLECTUAL PROPERTY RIGHTS

When intellectual property rights could be an issue, the public sector organisation must check whether it is the rightsholder of these data or whether it has in some way acquired the right to publish the data. If this is not the case, the necessary rights must first be obtained by means of an agreement with the rightsholder. For the rights held by civil servants or employees this can be arranged through the staff regulations or the labour agreement.²

For datasets, documents or other material which is created by a private party under contract with the public sector organisation in question, the transfer of the intellectual property rights must be regulated in an agreement. We

² This is already the case for civil servants who come under the Flemish Staff Regulations. See Government of Flanders Decree of 13 January 2006 determining the legal position of the personnel of the services of the Flemish public administration, Belgian Official Journal of 27 March 2006.

recommended the inclusion of a provision granting the necessary rights to the public sector organisation to make the material available as Open Data in any agreement or public contract with a third party for the creation of copyright-protected material. This may involve a full transfer of rights, which results in the public sector organisation becoming the rightsholder. Alternatively, it may be an option to insist on a licence for the provision of the data, for instance because the full transfer would be too expensive or because the private contractor wants to use the material as well. In the first case, the public sector organisation becomes the 'owner' of the material, whereas in the second case the public sector organisation obtains the right to distribute the data, but the contractor continues to be the 'owner'.

Recommendation 1: Verify whether intellectual property rights exist on the data you wish to make available. If the public sector organisation is not the intellectual property rights holder, insist on an agreement with the current rightsholder. Include a provision giving the public sector organisation the necessary rights to make the results available as Open Data in any future agreements or public contracts with third parties or for the creation of datasets or documents.

2.2.2 EXCEPTIONS TO OPEN GOVERNMENT

Even if the public sector organisation is the full owner of the data, this does not automatically mean that it can make these data available to the general public.

The protection of other legitimate interests may in some cases require that the data are not made public. These interests are listed in the Flemish Parliament Act of 26 March 2004 on Open Government. For every dataset which is considered to be made available as Open Data, it must therefore first be checked whether the exceptions of the Flemish Parliament Act on Open Government apply. In the following cases the data must not be made available:

- 1/ If their disclosure is not without prejudice to the public sector organisation's obligation to maintain confidentiality.
- 2/ If their disclosure is not without prejudice to the protection of privacy (unless the person concerned consents to the disclosure).³
- 3/ If their disclosure is not without prejudice to the secrecy of the deliberations of the Government of Flanders and the responsible authorities that depend on it, the organs of the Flemish Parliament or a public sector organisation in Flanders.
- 4/ If it concerns administrative documents compiled exclusively for criminal or administrative penalty proceedings.
- 5/ If it concerns administrative documents compiled exclusively for the possible implementation of disciplinary measures, for as long as the possibility of a disciplinary measure continues to exist.
- 6/ If it concerns administrative documents which contain information provided by a third party without this party being obliged to do so, and which the said party has explicitly designated as confidential (unless this person consents to the disclosure).
- 7/ If the importance of the disclosure does not outweigh the economic, financial or commercial interests of a public sector organisation in Flanders.
- 8/ If the importance of the disclosure does not outweigh the confidential nature of Flanders' international relations or Flanders' relations with supranational institutions, the federal government and the other Communities and Regions.
- 9/ If the importance of the disclosure does not outweigh the confidential nature of commercial and industrial information, when this information is protected to safeguard a legitimate economic interest (unless the party from which the information originates consents to the disclosure).
- 10/ If the importance of the disclosure does not outweigh the administration of justice in civil or administrative proceedings and the possibility of receiving a fair trial.
- 11/ If the importance of the disclosure does not outweigh the confidentiality of the actions of a public sector

³ This exception partially overlaps with the regulation on the processing of personal data (cf below).

organisation, insofar as this confidentiality is required for administrative enforcement, the performance of an internal audit or the political decision-making process.

12/ If the importance of the disclosure does not outweigh the public order and safety.

Separate exceptions apply to environmental information. However, they are comparable to the aforementioned interests.⁴

Recommendation 2: Check whether making the data available is not in violation of any interests protected by the Flemish Parliament Act of 26 March 2004 on Open Government.

2.2.3 PROTECTION OF PERSONAL DATA

In accordance with the concept paper of the Government of Flanders, Open Data is about making non-personal public sector data available. Therefore, check for each dataset that is made available if this dataset contains any personal data. If this is indeed the case, the data will in principle not be made available as Open Data.⁵ The concept 'personal data' is defined as follows: "any information relating to an identified or identifiable natural person".⁶ An identifiable natural person is "one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity".

Recommendation 3: Check whether the data contain personal data before they are made available.

2.3 CRITERIA AND PRIORITIES

Keep the following parameters in mind during the selection process when you want to publish datasets as

⁴ See Article 15 of the Flemish Parliament Act of 26 March 2004 on Open Government.

§1. The environmental bodies referred to in Article 4 shall reject an application for disclosure, insofar as it relates to environmental information, if they are of the opinion that the importance of the open nature does not outweigh the protection of one of the following interests:

the protection of individual privacy, unless the person concerned consents to the disclosure;

the secrecy of the deliberations of the Government of Flanders and of the responsible bodies that depend on it, the secrecy of the deliberations of the organs of the Flemish Parliament, as well as the secrecy specified by law or Flemish Parliament Act of the deliberations of the organs of the bodies referred to in Article 4, §1, 3° to 10°;

the confidential nature of administrative documents compiled exclusively for criminal or administrative penalty proceedings;

the confidential nature of administrative documents compiled exclusively for the possible implementation of disciplinary measures, for as long as the possibility of a disciplinary measure continues to exist;

the protection of information provided by a third party without this party being obliged to do so, and which the said party has explicitly designated as confidential, unless this person consents to the disclosure;

the confidential nature of the international relations of the Flemish Region or the Flemish Community and of the relations between the Flemish Region or the Flemish Community and the supranational bodies, the federal government and other Communities and Regions;

the confidential nature of commercial and industrial information, when this information is protected to safeguard a legitimate economic interest, unless the party which the information originates from consents to the disclosure;

the administration of justice in civil or administrative proceedings and the possibility to obtain a fair trial;

the confidentiality of the actions of an environmental authority, insofar as this confidentiality is required for administrative enforcement, the performance of an internal audit or the political decision-making process;

public order and safety;

the protection of the environment the information relates to.

§2. In so far as the requested information concerns emissions to the environment, the grounds for exception referred to in §1, 1°, 2°, 5°, 7°, 9° and 11° shall not apply.

For the grounds for exception referred to in §1, 3°, 4°, 6°, 8° and 10°, the fact whether the requested information relates to emissions to the environment shall be taken into account."

§3. The exceptions mentioned in §1, 9° and 11° shall not apply to information, specified in the Cooperation Agreement of 21 June 1999 between the federal State, the Flemish Region, the Walloon Region and the Brussels-Capital Region concerning the control of major-accident hazards involving dangerous substances.

⁵ The relation between Open Data and the protection of personal data is highly complex. Therefore, the Government of Flanders has decided in principle to leave personal data out of consideration and not to make them available as Open Data. In this way the protection of personal data is guaranteed to the highest possible extent.

Open Data:

- **Legal requirements:** Is there a legal obligation to make data available as Open Data? What legal constraints may exist?
- **Previously published data:** Is the information already publicly available or is it yet to be opened up?
- **Value of the data:** Are the data useful for social commitment and/or do they have any commercial value?
- **Reach of the data:** Are the data intended for the general public or for specific target groups?

Each parameter, including examples, is discussed in greater detail below.

2.3.1 LEGAL REQUIREMENTS

See paragraph 2.2 regarding the legislative framework and prior legal review.

2.3.2 PREVIOUSLY PUBLISHED DATA

Some data are already electronically available and can therefore be published relatively quickly and easily as an Open Dataset. Examples include:

- Cadastral information
- Topographic maps
- Traffic information
- Weather forecasts

2.3.3 VALUE OF THE DATA

Some data have mainly societal value. Examples include:

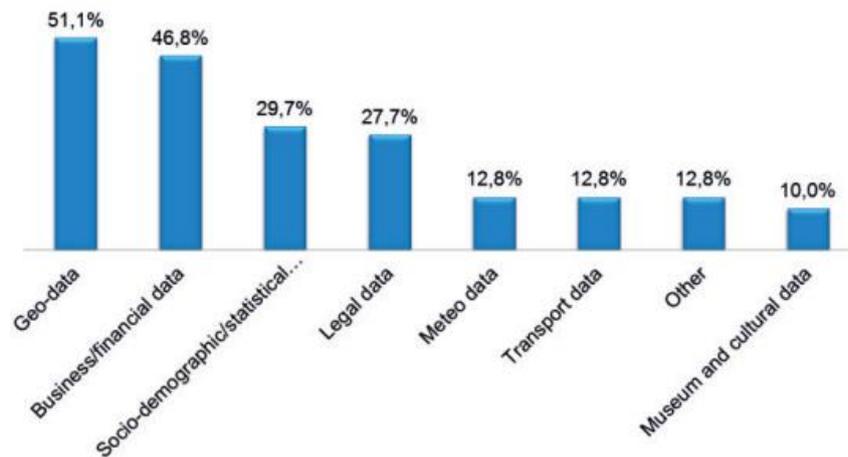
- Acts and parliamentary data (like data on votes by representatives)
- Pre-election data such as political party programmes
- Data from e-government and e-participation campaigns (like public consultations, crowdsourcing)

Other data may have more commercial value:

- Road maps
- Real-time traffic information
- Real-time weather information

A recent study⁷ shows that there is foremost an interest in the following data:

⁷ http://datos.gob.es/datos/sites/default/files/files/Estudio_infomediario/121001%20RED%20007%20Final%20Report_2012%20Edition_vF_en.pdf



This shows that people are especially looking for data that can be displayed on a map, in other words, geographical data. Anything that can be visualised on a map is particularly interesting, because it can be easily adapted to a smartphone or tablet application. It also means that a large part of the general public is mostly looking for convenient applications which links the many information sources with each other and present them graphically.

Information on businesses (business/financial data) also scores high. Again, it demonstrates the added value of being able to find business information through Open Data, obviously respecting the company's national.

In third place comes socio-demographic/statistical information. It goes without saying that information related to a particular area or neighbourhood is interesting for everyone. Data regarding schools, traffic, pollution, crime, accessibility, etc., may be important when looking for a new house for instance.

These first three categories together indicate that people are above all looking for applications that generate life-enhancing added value or for information which can facilitate interaction and communication with government. Gathering information from different sources and thereby creating added value is in many instances the task of businesses. Therefore, it is crucial for these data to be easily accessible.

The statistics regarding cultural information are somewhat surprising. Either information is available through other channels or the public underestimates its added value. As a matter of fact, this is what cities are most actively engaged in at present.

2.3.4 DATA SCOPE

Some data are specifically intended for the general public and therefore even more interesting:

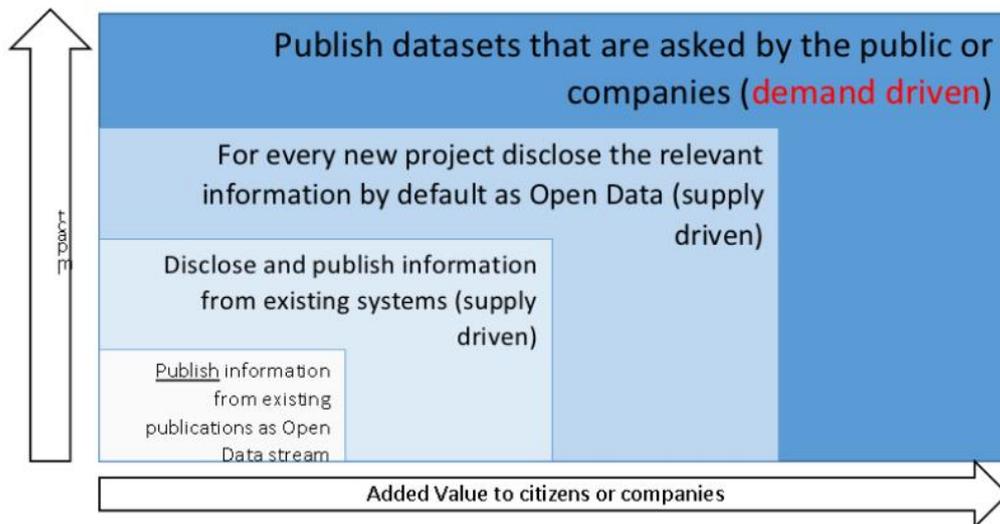
- Traffic information
- Public transport
- Election data

Other data are essential for small groups of people and niche markets.

- Information about facilities and financial support for people with special needs
- Economic statistics
- Court decisions

2.3.5 PRIORITIES

We can rightfully claim that a great deal of information is suitable for publication as Open Data and is usually already available through publications or documents. Provided the legal requirements are met, we propose using the model below to determine what may be relevant and of decisive importance for each public sector public sector organisation:



The basic principle is that information and data from existing publications lend themselves most easily to being published as Open Data. In most publications a lot of data have indeed been processed which are displayed in the form of a graph or table. Therefore, we propose to first transform these data into an Open Dataset and publish this. Each publication may be accompanied by an Open Dataset as an addition or for a detailed analysis.

We recommend to look also at the systems that are the source of the information and to see which data from those systems can be opened up. Creating Open Datasets from the source information may involve extra effort, but the result will be that also that source data is published in a structured way as Open Data. The necessary effort may be limited if all or part of that source data has already been published somewhere else.

Examples include:

- Lists of car parks
- Lists of public toilets
- Bus stops
- Train stations
- Financial data (like annual figures)
- Culture
- Tourism
- Opening hours
- Addresses
- ...

The next step is to include an additional step for each new project of information collection or processing which examines how these data can also be published as an Open Dataset. This additional step will automatically be built-in into each project and will immediately make sure that these data are also consistently provided as Open Data. By providing data as Open Data in a structured manner, they can also easily be (re-)used in in-house applications and publications.

Until now, we have mainly adopted a supply-driven approach. Every time, the government has taken the initiative and has decided which dataset is opened up and when.

It becomes more interesting when the government can also open up datasets at the request of citizens or businesses. This is a demand-driven scenario that allows the government to flexibly respond to the needs of the users.

The social value and impact of Open Data will indeed increase as the supply of data grows, but especially if the demand can be met as soon as possible. We know that there is already a large demand for data, especially in the following sectors:

- Socio-economic information
- GEO data
- Business information
- Statistical information
- Legal data

The point is that it is best to start basically by focusing on data that is already available in existing publications. However, the (societal) added value of Open Data increases for citizens and/or the government if you publish all data that you have at your disposal automatically as Open Data, especially if you adopt a demand-driven approach.

Recommendation 4: First publish the data that are already available and may generate added value for citizens, businesses or organisations.

3 CHOOSE A MODEL LICENCE

In order to promote the use of Open Data as much as possible and to simultaneously support the public administrations in making their data available, the Flemish public administration has drawn up model licences that can be used by all. Combining several datasets may be more difficult if different licence conditions apply to the individual datasets. The use of uniform licence conditions enables optimal use of Open Data, which may result in innovative and valuable applications. In addition, the bodies save time and effort by using the transparent licences drawn up by the Flemish public administration. The administration has formulated these model licences in such a way that they can be used by all administrations in Flanders, also on a local level, to make their data available. By applying these licences, the bodies help build an open government which supports participation and innovation.

3.1 PRIOR DECISION ON CHARGES

The Flemish Open Data concept paper starts from the basic principle that it should be possible to re-use Open Data free of charge or against a reasonable charge. This means that the public sector organisation must decide for each dataset whether it wants to make it available free of charge or impose a reasonable charge.

When opting for a reasonable charge, the public sector organisation must determine the criteria that will be used to define this charge. Within the current regulatory framework, the total income from this charge should not exceed "the cost of collection, production, reproduction and dissemination, together with a reasonable return on investment".⁸ However, the European Directive on the re-use of public sector information currently recommends to not let the charges exceed the marginal costs for reproducing and disseminating the documents.⁹ The future revision of this Directive will in all probability make the principle of these marginal costs mandatory (bar some exceptions).

When a public sector organisation opts for a reasonable charge, it must also establish a collecting procedure. This implies, among other things, that the method of payment must be decided, an account number must be made available, etc.¹⁰

Recommendation 5: Determine whether a charge will be made for re-using the data. Establish criteria for the calculation of this charge and organise the procedure for payment of the charge.

3.2 PRIOR DECISION ON POTENTIAL USE W.R.T. CATEGORIES

In keeping with the Flemish Parliament Act of 27 April 2007 on the re-use of public sector information and the Government of Flanders concept paper, it must be possible to re-use Open Data for non-commercial as well as for commercial purposes. However, it is still possible to make a distinction between commercial and non-commercial use with regard to the charge. It is not always easy to specify commercial and non-commercial use and the general principle of Open Data is that no difference is made between the different types of use. For this reason, administrations are advised in the first instance not to distinguish between the different types of use.

If, for specific reasons, the public sector organisation needs to make a distinction between commercial and non-commercial use of Open Data, it is essential that a clear definition of the concept 'commercial' is drawn up and communicated to potential users. For this definition, concepts of 'profit-making' and 'trader' can be considered. Each natural or legal person engaging professionally in commercial transactions, is deemed to use the data for commercial purposes (unless he can prove otherwise).¹¹

Recommendation 6: If a distinction is to be made between charges for commercial and non-commercial re-use, give a clear definition of the concept 'commercial', which implies the use by traders for purposes of profit.

⁸ Article 7 of the Flemish Parliament Act of 27 April 2007 on the re-use of public sector information, Belgian Official Journal of 5 November 2007.

⁹ Recital 14 Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information OJ L 345 90.

¹⁰ See below.

¹¹ See Government of Flanders Circular 2012/31. Circular on Access/Use and Re-Use, <http://vademecum.vandenbroele.be/entity.aspx?id=173>.

3.3 CHOICE OF MODEL LICENCE

On the basis of the decisions regarding the types of use and the possible charges, administrations can choose from five licence models (licences 4a and 4b must always be used together):

- 1/ A Creative Commons Zero¹² (CC0) license, by means of which the public sector organisation waives its intellectual property rights, insofar as this is legally possible¹³. As a result, the data can be re-used for any purpose whatsoever, without any obligation of attribution.
- 2/ Free Open Data Licence: Under this licence the public sector organisation does not waive its intellectual rights, but allows the data to be re-used for any purpose whatsoever, free of charge and with minimal constraints.
- 3/ Open Data Licence against a Reasonable Charge: Under this licence the public sector organisation still makes its data available for whatever re-use. However, it wants to receive a reasonable charge for all types of re-use.
- 4a/ Free Open Data Licence for Non-Commercial Re-Use: To comply with the principle of Open Data, the data should be available with minimal constraints for both non-commercial and commercial re-use. However, if necessary, a distinction can be made between the charges, if so desired by the public sector organisation concerned. For instance, a reasonable charge may be imposed for commercial re-use, whereas non-commercial re-use can be made free of charge. This licence concerns the free non-commercial re-use. Licence 4b will then be applied for commercial re-use.
- 4b/ Open Data Licence against a Reasonable Charge for Commercial Re-Use: When a distinction is made on the basis of the commercial nature of the re-use when making a charge, this licence is the counterpart of the Free Licence for Non-Commercial Re-Use.

The following combinations are thus possible for a specific dataset:

- CC0 for any possible re-use
- A free licence for any possible re-use
- A licence against a reasonable charge for all types of re-use
- The combination of a free licence for non-commercial re-use and a licence against a reasonable charge for commercial re-use¹⁴

All these licences are non-transactional and therefore do not have to be signed by the licensee. By using the data the licensee agrees to the terms and conditions. As a consequence, the data can be re-used entirely anonymously. The user's identity may be asked within the framework of the payment procedure for the use of data against a reasonable charge. However, in this case the future re-use of the data will still be anonymous.

The choice of one single licence for any possible re-use is the best solution for preserving the simple character of the licence models. The combination of two licences should only be applied if this is deemed absolutely necessary by the bodies concerned.

Recommendation 7: Use the model licences of the Flemish public administration for making Open Data available. Preferably choose one licence for all types of re-use, without distinguishing between commercial and non-commercial purposes.

¹² <http://creativecommons.org/about/cc0>

¹³ See Point 3.2 of the legal note with regard to the validity of the CC0 deed

¹⁴ In theory, the combination of a free licence for commercial use and a licence against a reasonable charge for non-commercial use is possible. However, in practice this is very unlikely to occur.

4 OPEN UP YOUR SOURCE DATA

This chapter describes a number of scenarios on how to open up data from a source system, transform them into Open Data and provide them on the Open Data Platform.

Nowadays, public sector bodies already use a lot of data for the publication of (annual) reports or other documentation. These data can also be published as Open Data.

The ultimate goal is to define a limited number of scenarios that bodies can use to create an Open Data stream with minimal extra steps and/or efforts, in addition to the existing data stream.

The advantage of the scenarios is that they are recognisable and broadly applicable. By developing these scenarios in combination with existing standard technologies, we apply a 'best practice' as a standard approach.

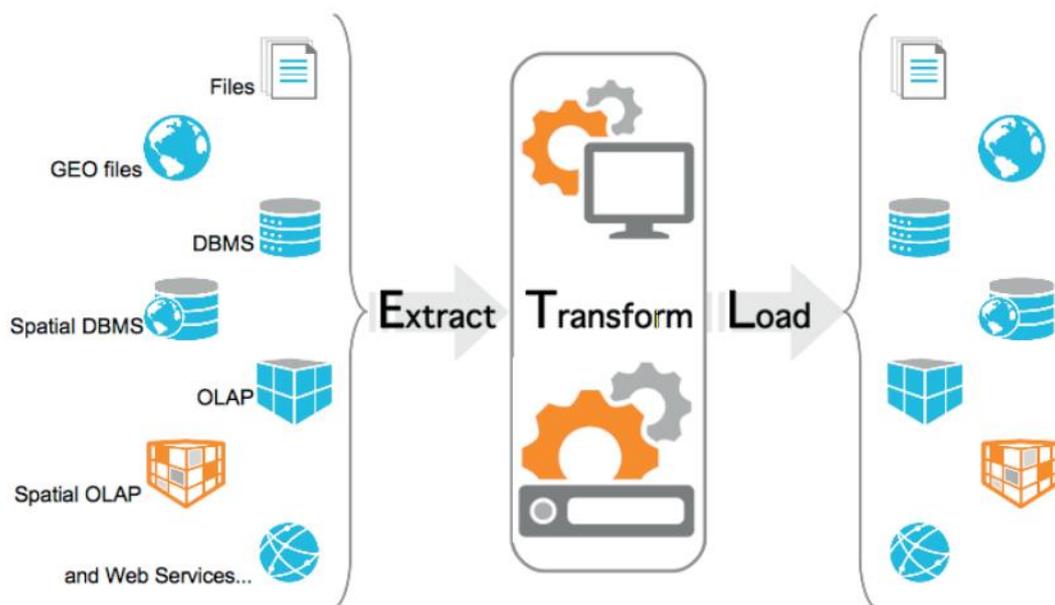
These scenarios are a starting point for transforming most data into Open Data.

In the last but one paragraph further explanation is given on opening up geographical and location-based information in Flanders.

Finally, several tips are given that may help to identify a need for Open Data within existing or new projects and to open up these data within the scope of the project.

4.1 UNDERLYING IDEA: USE OF ETL TECHNIQUES

We have drawn our inspiration for these scenarios from the Business Intelligence / Data Warehousing (BI / DWH) area of practice. For many years, techniques have been applied in that area to extract information from source systems or databases, to (partially) clean it, to make it consistent and to publish it. This process is often called the ETL (Extract, Transform, Load) process:

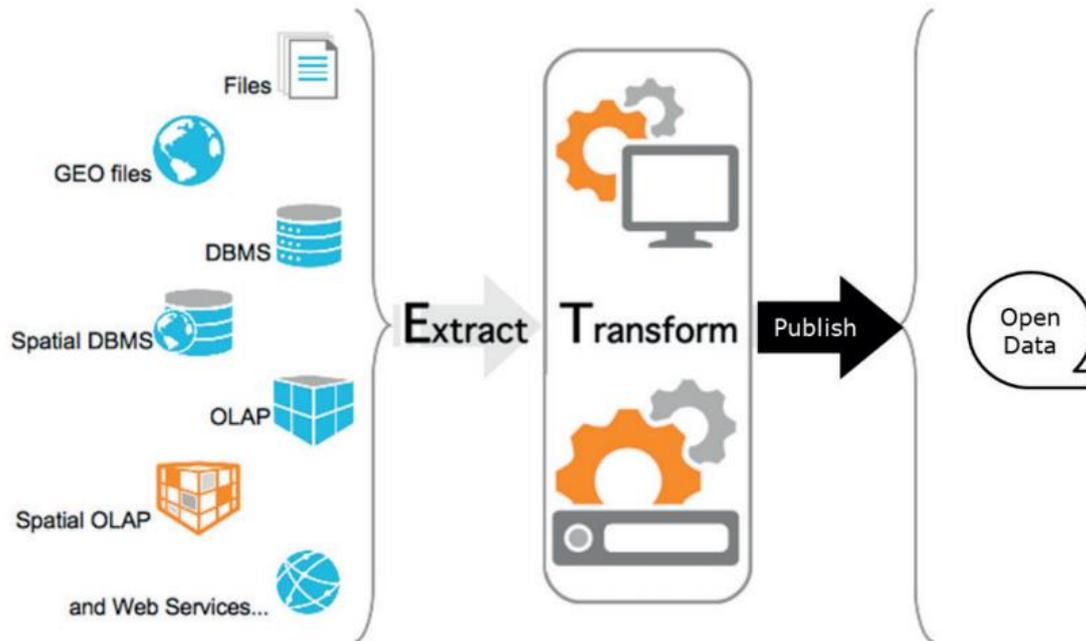


In a nutshell:

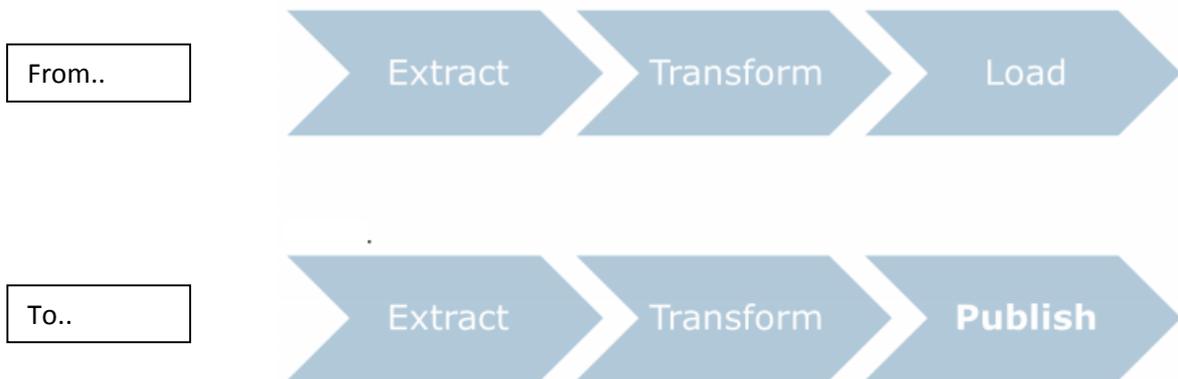


The technology (i.e. ETL software) used for traditional data warehouses can also be used for Open Data. These tools offer possibilities to prepare information for analysis or processing.

To a large extent, the same steps are required for publishing data in an Open Data stream. Only the final step, i.e. loading the information in a data warehouse environment, is replaced by another step, specifically uploading the data (and associated metadata) to the Open Data Platform, which we therefore call 'Publish'.



Summarised as a process diagram:



This means that for an Open Data stream we can largely re-use or apply the same ETL technology. In other words, no new techniques or tools are required for creating an Open Data stream.

If a data warehouse already exists, this is often also a suitable candidate to build on to (partially) provide Open Data. This will allow for easier monitoring of the consistency between the data that is used internally and the data that is provided externally (i.e. Open Data). In addition, a report (through BI tool on a DWH) can also be a suitable source for publication as Open Data.

It is therefore recommended to build as much as possible on existing technology, adding an extra step to publish the data concerned in an Open Data stream.

4.2 SCENARIOS FOR OPENING UP SOURCE DATA

All this is brought together in a number of scenarios that are further described below and that may serve as a starting point for administrations.

Currently, the following scenarios are in place:

- 1/ Starting from an existing publication:** The public sector organisation collects data from an existing publication (e.g. in graphs or tables) and publishes them as Open Data.
- 2/ Starting from an existing dataset:** The public sector organisation has already created certain data that can now also be published as Open Data.
- 3/ Starting from a source database:** In this scenario we extract data directly from a source database that is maintained by the public sector organisation and publish certain data as Open Data, after the necessary transformations and checks have been performed.
- 4/ Starting from an existing source system or package:** Bodies often use a commercial package that has its own database. These data can often not be accessed directly or they may be stored in a proprietary format that is determined by the supplier. In this scenario we use techniques for extracting these data from the package, transforming them and publishing them as Open Data. The difference with Scenario 3 is that packages often require the data to be opened up through other channels (like API or package-specific tools). However, these data can often also be opened up via traditional ETL tools.
- 5/ Starting from different sources and consolidating data:** This scenario is very similar to a data warehouse technique, where data are often already collected from different sources and uploaded to temporary tables from which the actual data warehouses are then often uploaded. The uploading itself may then be replaced by the publication of an Open Data stream.

The scenarios are described in order of increasing complexity. We believe Scenario 1 is the most accessible for all. The complexity increases for every subsequent scenario, because it is increasingly difficult to open up the data because additional steps are required. The scenarios are also based on the elements described further down in the present document, and in particular considering what is needed to make the data compliant with Open Data principles and to ensure they are opened up with sufficient quality.

Each of these scenarios is discussed in greater detail below. In the next chapter we elaborate on the technical tools which may support or automate this process.

4.2.1 SCENARIO 1: STARTING FROM AN EXISTING PUBLICATION

What?

In this scenario we start from an existing process in which data are collected and processed before they are included in a publication.

Just think of all the publications made available by a public sector public sector organisation (see for instance <http://www.vlaanderen.be/nl/publicaties>) and the amount of data that are collected and included in the publication. These data are the result of a number of process steps that are carried out by one or more administrations. One example is VRIND (<http://www.vlaanderen.be/nl/overheid/werking-vlaamse-overheid/hoe-werkt-de-vlaamse-overheid/vrind-2012-cijfergegevens-en-indicatoren-over-de-vlaamse-samenleving>). The wealth of figures in this publication are also interesting as Open Data for further analysis and/or processing.

When?

Because the government will continue to issue publications and administrations will continue to collect a lot of data for publication, this seems to us the easiest scenario for Open Data. It is indeed these published data that we can also transform into an Open Data stream, if relevant of course.

Therefore, this scenario can be used:

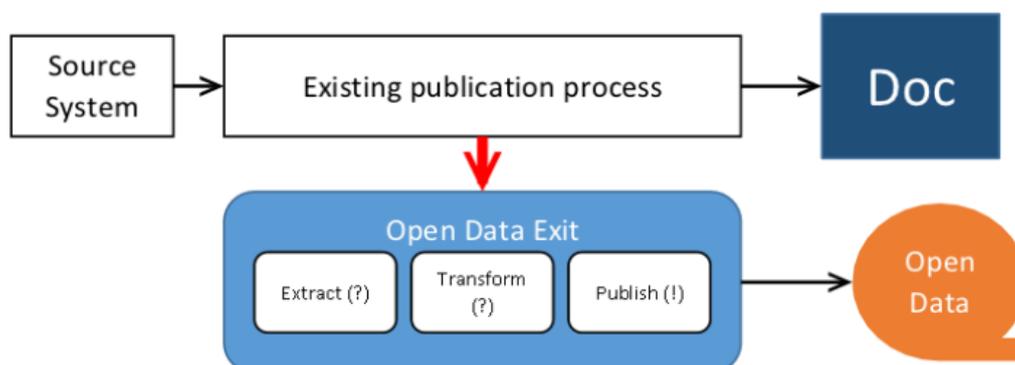
- For new publications; in this case, an additional step will be included in the process to make the data also available as Open Data.

- For existing publications; in this case, the existing process will be adapted to create the Open Data stream and to keep it up to date.

Once you have set up the process for extracting an Open Data stream from the publication and publishing it, you can apply this repeatedly. This way, the Open Data stream is kept synchronised with every update of the publication.

How?

The essence is that an additional step is included in the process to publish the data also as Open Data¹⁵. See figure below:



In the existing process, the IT department will have to identify the moment when the data are ready to be taken to the Open Data publication step. In this case, we only have a number of 'Publish' activities to perform.

Below an overview is given of the steps to be executed:

Extract

Not applicable in this scenario.

Transform

Not applicable in this scenario.

Publish

Once the data have been selected, they must be processed to meet the Open Data criteria:

- Collecting metadata.
- Publishing dataset (preferably automatically).
- Choosing licence model.
- Offering necessary conversions on the platform, and possibly also an API.
- Setting up a feedback loop, making sure the public sector organisation can be contacted in case of comments.
- Ensuring regular updates.

The last step is to publish the data as an Open Data stream on the Open Data Platform, in keeping with the guidelines drawn up earlier in the present document.

¹⁵ In an Open by Default approach the Open Data Exit will be part of the processing process.

Example

An example of this scenario is the data from the VRIND publication. See website <http://www.vlaanderen.be/nl/overheid/werking-vlaamse-overheid/hoewerkt-de-vlaamse-overheid/vrind-2012-cijfergegevens-en-indicatoren-over-de-vlaamse-samenleving> and text:

"VRIND, i.e. Flemish Regional Indicators, is an annual publication by the Research Centre of the Government of Flanders on the results of Flemish policy and their impact on society and the environment.

On the basis of figures, VRIND offers insight into the activities of the Flemish public administration, the policy areas in which it operates, and the results it achieves. For each area, VRIND also outlines the recent socio-cultural, economic, ecological and demographic developments in Flanders."

The data are collected and compiled typically as part of the annual publication cycle. The data bundled together in VRIND are (almost) all being opened up already via local statistics/VOBIP Cognos publiek (see http://aps.vlaanderen.be/lokaal/lokale_statistieken.htm).

However, this was done without considering the Open Data criteria. For this reason, an additional step has been added to the last stage of the process, during which the final datasets were first isolated. After that, a script was written that complements the metadata of each dataset and uploads them to the CKAN platform. It concerned a total of more than 1,100 datasets.

This method may be repeated for every publication. The data can be uploaded either manually (if relatively few datasets) or automatically. In the latter case the Open Datasets can be uploaded repeatedly when information is changed or new information is added.

4.2.2 SCENARIO 2: STARTING FROM AN EXISTING DATASET

What?

A lot of bodies currently already publish a great deal of information on websites in a downloadable format like XLS (note: if PDF, see Scenario 1). We also notice that many investments have already been made to build viewers which allow this information to be consulted or visualised online (like BIVO, VOBIP publiek, see: <http://vobippubliek.vlaanderen.be/cognos10>). This implies that the public sector organisation already has a process or IT support to prepare, publish and view this information.

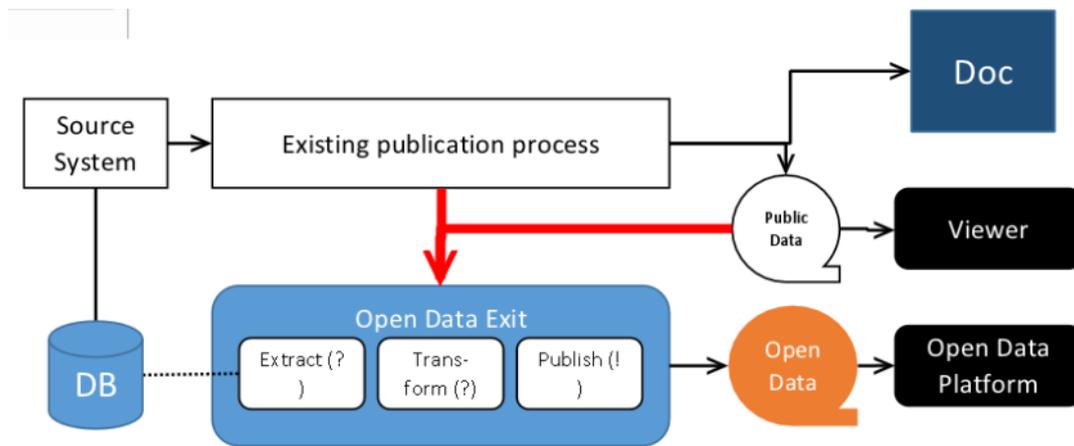
Within this scenario, we want to add an additional step to this process, which also prepares the data for publication as Open Data. This means that some extra things need to be done as a minimum, like providing metadata and publishing them on the Open Data Platform.

When?

We presume that the government will keep making information available via downloads or viewers and that Open Data is an additional channel for releasing this information. This means that the underlying process will continue to exist and may be extended by an additional step.

How?

As stated, we presume that a process is in place during which the data are now already being prepared. We propose to build in an additional step into this process to also transform the dataset into an Open Data stream.



In some cases we find that public data are already being created and made available through a specific (web) viewer application in addition to the publication. Usually, these public data do not entirely meet the Open Data criteria. Still, with minimal efforts and additional steps this could prove a sound basis for quickly turning them into an Open Dataset.

This scenario takes into account a number of additional steps, since the existing process deals with the data in a totally different manner.

These steps are:

Extract

Isolate data and filter them from the database in a uniform dataset. This may require an additional step to extract data directly from the database. We may also select different data for the Open Dataset (e.g. fewer fields, anonymised) than in the existing process. Of course, this step will not be necessary if the data from the publication already satisfy the Open Data criteria.

Transform

This encompasses a thorough quality check of the data, as is the case in every data warehouse environment. For instance, using uniform names for fields and content -- no cryptic abbreviations, not 0 or 1 for gender but M = Man, storing addresses in a consistent manner, writing names in full and in the same format, etc. We presume, however, that these checks and actions are already included in the existing process. Therefore, these steps may have to be harmonised. The idea is by no means to create a separate process alongside the existing one. However, experience shows that most of the time specific data are created for a selected audience (e.g. with access restrictions, non-public). Of course, this is not allowed for Open Data, which is why extra steps will be required to prepare the data. It goes without saying that this step is not necessary if the data from the publication have already gone through a similar process and satisfy the Open Data criteria.

Publish

The following steps are always required in this scenario, once the Open Dataset has been created:

- Collecting metadata.
- Publishing dataset (preferably automatically).
- Choosing licence model.
- Offering necessary conversions on the platform, and possibly also an API.
- Setting up a feedback loop, making sure the public sector organisation can be contacted in case of comments.
- Ensuring regular updates.

Example

A nice illustration of this scenario is given by the unemployment figures published on <http://www.werk.be/cijfers> which can also be downloaded as XLS. The basic idea is that these datasets are also made available on the Open Data Platform and are automatically updated every month (i.e. publication of the latest data). Through the website people have access to the data using the viewer that is built-in into the website and offers limited download options (from an Open Data perspective).

Another example is (the public version of) BIVO (dashboard company information) where a visualisation component is already provided on the basis of Cognos Viewer. This viewer too offers a number of possibilities that can be used directly from the website, as well as limited download options (from an Open Data perspective).

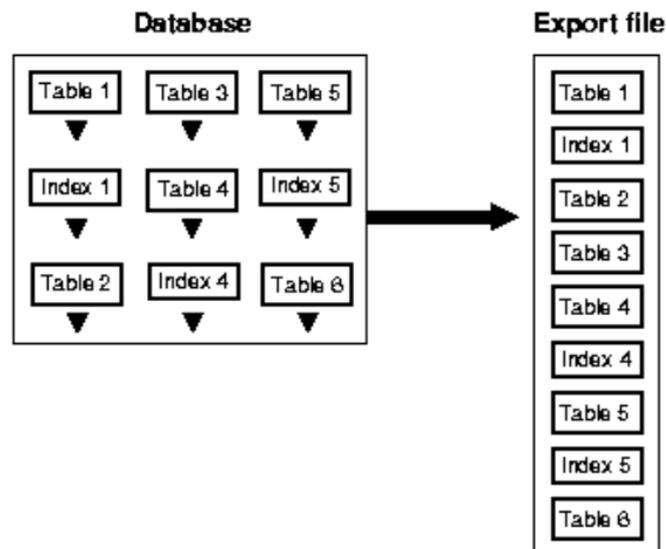
In both examples, we propose to also fully organise the relevant data as an Open Data stream and to publish them additionally on the Open Data Platform.

4.2.3 SCENARIO 3: STARTING FROM A DATABASE

What?

In many cases the core data are in a database that has been created for an application to support a business process for the public sector organisation. This is also the starting point of this scenario, namely to extract data from a database and transform them into an Open Data stream.

Most databases use a structure for storing data that is not really suitable for Open Data. This is because the main aim is to quickly transfer data to the application to be able to create, retrieve or delete data. This structure is often called OLTP (= online transaction processing) and relational in nature: data are linked to each other through a relation (key) in different tables.



In summary, the basis of this scenario is that the data must be extracted from the application database first, before they are prepared for publication as Open Dataset. An additional processing step is therefore needed to extract these data and transform them or make them consistent for Open Data publication. In this case the Transformation tools are necessary as well.

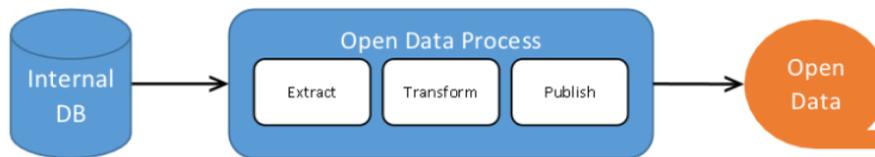
When?

The assumption in this scenario is that we have an application which is developed internally on one of the existing internal environments (for instance, built in Java, .NET or APEX) and the DB is one of the standards used by the public sector organisation (like Oracle, SQLServer, PostGress for Flemish public administration). If this is not the case, we propose proceeding with Scenario 4.

We also assume that the previous scenarios are not applicable as starting point. In other words, data will first have to be extracted from databases before they are opened up to the Open Data Platform.

How?

In this scenario the techniques 'Extract' and 'Transform' are very useful, because there is no existing process to start from or hook into. Therefore, we start directly from a database (through a query or tools).



A remark to be made in this context is that we do not equate this type of 'technical' operations with the 'content-related' operations which usually take place during the 'Transform' step. We refer in this context to complex transformations, like linking consecutive enrolments of pupils from different systems with each other (e.g. lower, secondary, higher, VDAB, etc.) to construct an entire study career. The result is the consequence of several technical operations and steps that form a logical or content-related whole. Therefore, more than 1 technical step may be required to perform the transformation.

We propose the following steps:

In general

First of all, create a logical data map in which the physical relations of the DB are ignored as much as possible. The result is that an extract from the database will very rarely or never be directly published as Open Data. However, it is possible. In most cases it will be impossible to publish a one-to-one copy and steps will have to be taken to transform the DB copy into a stand-alone format (for instance replacing keys by values, entering references, making references consistent, M = Man, etc.).

Furthermore, steps may also be needed to combine tables from the database with each other until you arrive at the logical data map. Transformation technology may possibly also have to be used to make the data consistent.

Extract

Most database systems have standard techniques for reading out tables and making them available as flat files. For example:

- Oracle: via EXPORT tool
- Microsoft: SQL Server Import and Export Wizard
- MySQL: via mysqldump tool
- Postgres: SQL Dump procedure

When data change frequently, a good alternative may be to write a programme that reads out the data from the DBMS via ODBC or JDBC drivers. ODBC (or direct SQL, which is in fact the same thing) will allow for the programming of more complex extraction logic.

Finally, use can also be made of the possibilities provided by each ETL tool set. From a practical point of view, it is also recommended for complex transformations to store the data, either temporarily or not, in a database which can then be further used to realise the definitive Open Data stream.

Two dimensions are always important in such procedures:

- either exporting the entire content of the database and publishing it as Open Data;

- or only downloading the modifications /delta with respect to the previous version and combining it with the Open Data stream.

Transform

This encompasses a thorough quality check of the data, as is the case in every data warehouse environment. For instance using uniform names for fields and content -- no cryptic abbreviations, not using 0 or 1 for gender but M = Man, storing addresses in a consistent manner, writing names in full and in the same format, etc. Because we cannot assume here that a process is already in place, all these transformation steps have to be carried out. At this stage content can also be processed, like anonymising data or combining datasets in order to achieve a uniform granularity.

It is preferred to export the extracted data as quickly and frequently as possible into Open Data, so as to allow users and citizens to have the latest reference data at their disposal. Especially when it concerns rapidly changing data. As a consequence, it is important that the transformations that these data undergo are reproducible and should preferably happen automatically.

Publish

Once the data are ready for publication as Open Data, the following steps remain to be completed:

- Collecting metadata.
- Publishing dataset (preferably automatically).
- Choosing licence model.
- Offering necessary conversions on the platform, and possibly also an API.
- Setting up a feedback loop, making sure the public sector organisation can be contacted in case of comments.
- Ensuring regular updates.

Example

At the time of writing (Q4 2014) we do not have any Open Datasets yet that have completed these steps.

4.2.4 SCENARIO 4: STARTING FROM AN EXISTING SOURCE SYSTEM

What?

The starting point in this scenario is that the public sector organisation has one or more operational systems or packages from which Open Data can be obtained, provided a number of operations are performed, either with or without intervention from the supplier or service provider that manages the package for the public sector organisation concerned.

In this scenario, the additional complexity is that a package is involved that may not allow direct access to the database.

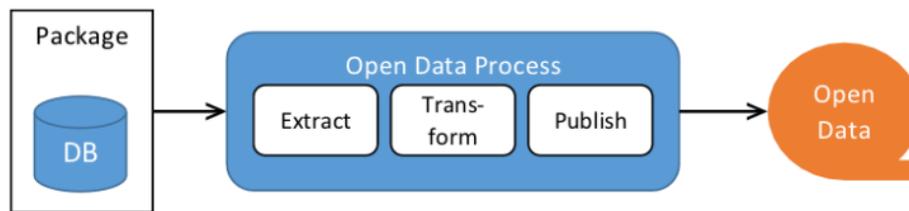
When?

This scenario applies when the data are in a (commercial) package, in which it is sometimes impossible to directly access the database.

This scenario also applies when the public sector organisation involves a service from a third party (like Software As A Service), and the service is provided externally (like a Cloud application).

How?

This scenario is an extension of the previous scenario, but differs from it in that the database cannot be directly accessed, as shown below:



The advice given in case of a package is to never copy and publish data one-to-one. Databases and field names are often described cryptically and the database structure is designed in such a way that it is only optimised for online transactions. In case of upgrades, the structure usually changes, which means you have to start all over again.

The following steps are relevant as well:

Extract

We assume that the public sector organisation can discuss with the supplier of the package how the data can be opened up:

- Through a procedure belonging to the package. Package suppliers often deliver a programme or script for reading out the data, even including specific parameters. Since this procedure is written and maintained by the supplier (for instance, in case of package upgrade), you are certain that this procedure is forward and backward compatible.
- Through APIs, if provided by the package. This means you will have to write a programme that uses the APIs to obtain data. Please note that some APIs can also carry out operations on the data before delivering them, like consolidation, aggregation, etc.
- Through a new programme or script which directly reads out the data from the package database. However, in this case one is dependent on the database design of the supplier, which tends to change with each new version or upgrade. This means that you will each time have to adjust the programme. Therefore, this approach is not recommended.

Two dimensions are always important in such procedures:

- either exporting the entire content of the database and publishing it as Open Data;
- or only downloading the modifications /delta with respect to the previous version and combining them with the Open Data stream.

Transform

This encompasses the transformation and a thorough quality check of the data, as is typically done in every data warehouse environment. For instance, using uniform names for fields and content -- no cryptic abbreviations, not 0 or 1 for gender but M = Man, storing addresses in a consistent manner, writing names in full and in the same format, etc.). Because we cannot assume here that a process is already in place, all these transformation steps have to be carried out. At this stage content can also be processed, like anonymising data or combining datasets in order to achieve a uniform granularity.

Publish

Once the data are ready for publication as Open Data, the following steps remain to be completed:

- Collecting metadata.

- Publishing dataset (preferably automatically).
- Choosing licence model.
- Offering possible conversions on the platform, and possibly also an API.
- Setting up a feedback loop, making sure the public sector organisation can be contacted in case of remarks.
- Ensuring regular updates.

Example

One example are contact details, which are often kept up-to-date in commercial packages (like CRM system) or in-house applications. These data will have to be opened up by means of a procedure within the package or by writing a programme which calls the API and reads out the DB. This is the easiest step. Making address details consistent is the next step, because different formats still often occur in publications. For this reason we advise to use existing modelling agreements for this. For contact details, please refer to the OSLO standards (see <http://purl.org/oslo/>).

At the time of writing (Q4 2014) we do not have any Open Datasets yet that have completed these steps.

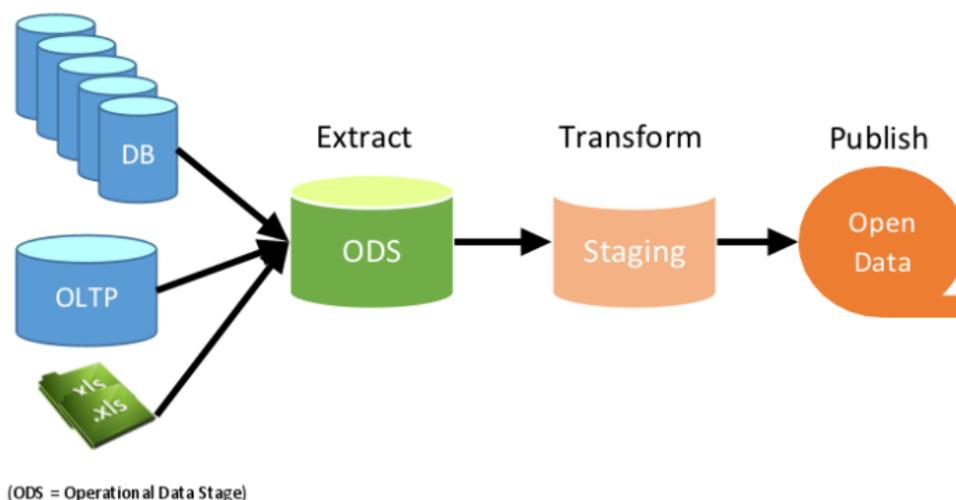
4.2.5 SCENARIO 5: STARTING FROM DIFFERENT SOURCE SYSTEMS

This is a scenario for experienced bodies involving Business Intelligence and data warehouse techniques. It expects the public sector organisation to already have some data warehouses and reporting environments and be greatly experienced in opening up data from different sources. All 'Extract' and 'Transform' techniques and tools are already in use and can be re-used here.

What?

Data published in a data warehouse are often the result of a series of operations, in terms of both content and aggregation. First, the relevant data are extracted from different sources and stored consistently in an Operational Data Store (ODS) environment. For Open Data this has the advantage that these data no longer need to be extracted separately from the source systems. An ODS is also the lowest form of granularity of the data, and therefore an ideal source for opening up all the data here.

After that, data are often processed in phases and stored in staging tables in the meantime. Several programmes then run on this to further aggregate the data and make them consistent before uploading them to a data warehouse. For Open Data we can also start from a staging table from which the Open Data stream can then be produced. The great advantage is that in this case all the corrections and operations of the data have already taken place.



When?

In this case we assume that the Open Data team will make maximum re-use of a number of existing facilities within the public sector organisation. This scenario is thus valid when an ODS environment is already in place and/or a number of staging tables are available as starting point. The Open Data team can then use the 'Extract' and 'Transform' programmes to immediately make consistent Open Data streams. However, the Open Data team must in this case also check or adjust the aggregation and granularity in order to keep the Open Data streams as fine-grained as possible. A separate or specific staging table for Open Data may therefore have to be made. In this case an extra programme should be provided within the public sector organisation's technology. When complex transformations are carried out on the source data, staging tables are therefore added.

To combine data from different bodies and, in the long run, publish one consolidated file (for instance one address file for all bodies with a similar data model).

How?

The following steps are relevant:

In general

Generally, we propose to maximally re-use existing standard ETL environments and technologies for Open Data. We do not see the need to introduce other tools. If no technology is available, please refer to Annex 2 "Technical Standards", where advice and tips are given.

Extract

In this case we assume that all Extract functionality is available within the existing BI or DWH and we do not specifically have to set it up for an Open Data stream.

Transform

Again, we start from the existing BI or DWH environments. All Open Data streams can thus be taken from the latest staging tables. A specific staging table may just have to be drawn up for Open Data, but this should be examined with respect to the granularity.

Publish

Once the data are ready for publication as Open Data, the following steps remain to be completed:

- Collecting metadata. Because we start from an existing DWH environment, sufficient metadata may already be available in this environment. Nevertheless, we advise to check the set of metadata, because usually not all Open Data metadata fields are provided for.
- Publishing dataset (preferably automatically).
- Choosing licence model.
- Offering possible conversions on the platform, and possibly also an API.
- Setting up a feedback loop, making sure the public sector organisation can be contacted in case of remarks.
- Ensuring regular updates.

Example

One example of a developed BI and DWH environment can be found within the Department of Education and Training, more specifically the Knowledge Centre project, where all the described tools and techniques are available.

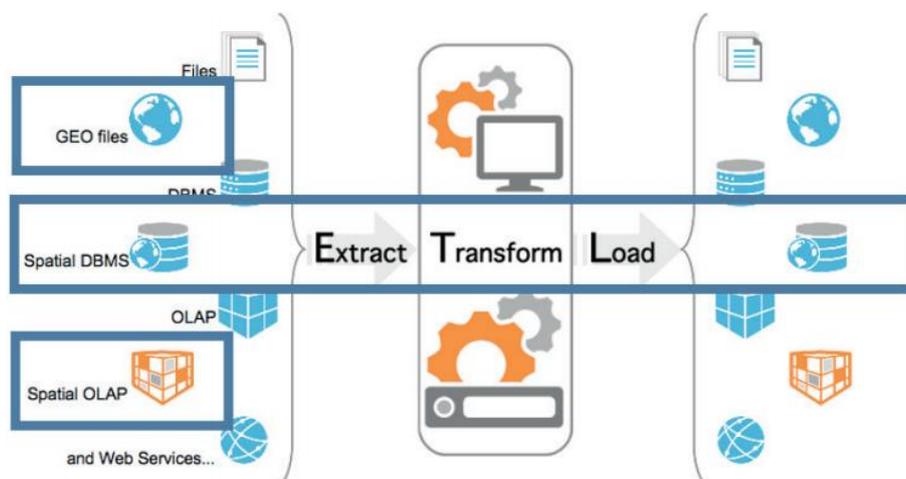
However, at the time of writing (Q4 2014) no information flow or data have already completed these steps as Open Dataset.

4.3 OPENING UP GEOGRAPHICAL AND LOCATION-BASED INFORMATION IN FLANDERS

4.3.1 POSITIONING

In this document we have so far presented a generic approach on the basis of the ETL techniques to open up data. For specific data streams, like geographical data, additional steps or other specific steps may be required for opening up these data. Therefore, we have added an extra section which zooms in on the processing of geographical data (GIS, GEO) into an Open Data stream.

On the basis of our general ETL drawing, here we specifically refer to databases, files and data warehouses that contain spatial information and serve as input source for opening up geo-information, as shown below:



We mainly consider this approach to be relevant for Scenarios 3, 4 and 5 where a complete ETL process will be executed before opening up and publishing geo-data. The general ETL techniques and scenarios described in the present document continue to be valid for the processing of geographical data. However, additional attention will be given to the specific European INSPIRE¹⁶ (i.e. Infrastructure for Spatial Information in the European Community) and Flemish GDI¹⁷ (i.e. Geographical Data Infrastructure) directives. INSPIRE and GDI provide standards for data models, protocols, formats and metadata to be used to open up geographical data.

Within the context of Open Data these directives will continue to apply. The text in the present document has been checked against these directives. However, in order to demonstrate the potential impact of these directives, we offer a guideline that administrators of geographical information may use when publishing their data.

4.3.2 GUIDELINE FOR THE PUBLICATION OF GEOGRAPHICAL DATA

In 1995, the partnership for geographical information in Flanders was started, i.e. the current GDI-Flanders. This partnership aims to optimise the production, management, exchange, use and re-use of geographical data sources and geographical services. All Flemish bodies are part of it. The steering group GDI-Flanders is the directing public sector organisation of the partnership, whereas the Flanders Geographical Information Agency (AGIV) is the implementing public sector organisation. In the course of 2015, the partnership and the Flanders Geographical Information Agency will be structurally integrated into a broader partnership and information agency.

To make data easily exchangeable between public sector bodies on the one hand and between public sector bodies and citizens, businesses and organisations on the other, these data are added to the GDI. To add these

¹⁶ <http://inspire.ec.europa.eu/>

¹⁷ <http://www.geopunt.be/voor-experts>

data to the GDI, a dossier for addition must officially be submitted to the steering group GDI-Flanders. The standard is that all environment-related datasets and services, under the management of participants in GDI-Flanders, must be added to the GDI (in conformity with Article 12 of the GDI Flemish Parliament Act¹⁸). Adding non-INSPIRE datasets, managed by participants in GDI-Flanders, to the GDI is not mandatory, unless the steering group GDI-Flanders has found that this mutual exchange is necessary to perform tasks of general interest. A re-use regulation will have to be put in place, however, for data that are not added to the GDI. For geographical data, the re-use conditions are established through a procedure via the steering group GDI-Flanders.

In order to check whether your data fall within the scope of the provisions of the INSPIRE Directive or the GDI Flemish Parliament Act and what exactly is expected of you, you can use the following guideline.

- 1) Determining the type of geographical data source:
 - i) Does INSPIRE apply to my data? Check whether your data are included in the Annexes¹⁹ of the INSPIRE Directive. If this is the case, the INSPIRE and GDI directives shall apply. The data must be added²⁰ to the partnership GDI-Flanders. The steering group GDI-Flanders establishes the terms and conditions of use and the regulation for re-use.
 - ii) Your data is not on the INSPIRE list, but is:
 - 1) environment-related
 - 2) an officially recognised authentic source of information
 - 3) an authoritative source of information
 - 4) a relevant source identified by the GDI-Flanders steering groupIf this is the case, the GDI directives shall apply. The data must be added to the GDI-Flanders. The steering group GDI-Flanders establishes the terms and conditions of use and the regulation for re-use.
 - iii) Other data: The administrator is not obliged to comply with the INSPIRE or GDI directives. If the administrator and/or users are of the opinion that the data are valuable for a broad target group, it is recommended to add the data to the GDI-Flanders.
- 2) If the data source is added to the GDI, at least the following further steps have to be taken:
 - i) Creating metadata in compliance with the “GDI-Flanders Best Practices for Metadata”²¹ and publication of these metadata in the Geopunt catalogue²².
 - ii) For INSPIRE data and authentic Flemish geographical data sources: Offering view services
 - iii) For INSPIRE data and authentic Flemish geographical data sources: Offering download services
 - iv) For INSPIRE data: Harmonisation of the data model with the corresponding INSPIRE data specification²³.

If you have geographical data at your disposal and you have any questions regarding the procedure to be followed for the publication of these data, please contact the Flanders Geographical Information Agency (*Agentschap voor Geografische Informatie/AGIV*) at contactpunt@agiv.be.

4.3.3 TECHNICAL IMPACT

The current legislation and directives for the publication of geographical information are aimed at an optimal digital exchange of this information. The European INSPIRE Directive has become effective in 2007. This legislation was transposed to the Flemish context through the GDI Flemish Parliament Act which entered into effect in 2009. For geographical information falling within the scope of this framework of agreements:

- the data model must be harmonised with the INSPIRE data specifications;

¹⁸ http://www.geopunt.be/~media/geopunt/geowijzer/inspire/documenten/vlaams%2520gdi-decreet%2520_geconsolideerd.pdf

¹⁹ <http://www.geopunt.be/voor-experts/inspire>

²⁰ <http://www.geopunt.be/geowijzer/gdi-vlaanderen/hoef%2520deelnemen>

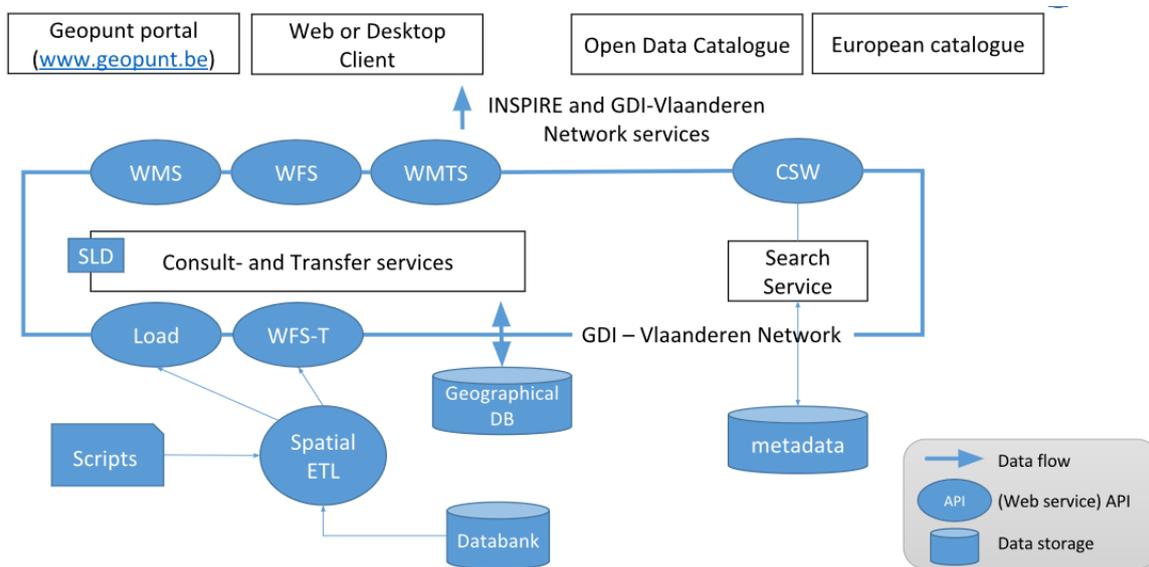
²¹ http://www.geopunt.be/~media/geopunt/geowijzer/metadata/documenten/gdi-vlaanderen_best_practices_voor_metadata-v1_0.pdf

²² <http://www.geopunt.be/catalogus>

²³ <http://inspire.ec.europa.eu/index.cfm/pageid/2>

- the metadata must be documented in compliance with the GDI-Flanders Best Practices for Metadata;
- use must be made of standardised web service interfaces for the publication.

The figure below gives an overview of the components belonging to the GDI architecture:



Impact on the ETL process

Until 2005, ETL for geographical information (Spatial ETL) was mainly the playing field of mayor GIS (Geographical Information System) software suppliers. Meanwhile, because of the growing importance of location-based information over the past years, the support for geographical information has also found its way to the most prominent commercial and open source DBMS (e.g. MS SQL Server, Oracle, IBM DB2, PostgreSQL, MySQL, etc.). These DBMS offer support in storing and managing geographical data and have a basic set of spatial functions for loading and processing these data. Since this functionality is integrated into the DBMS, it is an extension of the existing ETL functions and you can handle and process the geographical data in the same way as your other data. For data administrators that only store point locations (x,y) as object attributes, this basic support will in most cases be sufficient.

If you require more advanced Spatial ETL (e.g. model transformation, geocoding, projection, generalisation, etc.) or wish to support more complex data types (e.g. 3D, raster, etc.), there is a great chance that the available set of ETL functionality will be insufficient. In this case you can fall back on specialised commercial and open source software for GIS (e.g. Quantum GIS, Gaia, ArcGIS, MapInfo, etc.) and Spatial ETL (e.g. FME, GDAL/OGR, GeoKettle, etc.).

Impact on metadata

In order to be able to share your data and electronic services within the partnership GDI-Flanders, you must describe them using metadata. The INSPIRE Directive defines a minimum number of fields that must be provided to comply with the ISO standard for metadata (ISO 19115 for data, ISO 19119 for web services). This European regulation was transposed for Flanders into Best Practices for documenting metadata for geographical information. If you follow these Best Practices, you also meet the ISO and INSPIRE requirements.

In order to facilitate the creation and administration of metadata for geographical information, a web application is provided within Geopunt, the portal for geographical information of the Flemish public administration, to document your metadata (the Geopunt metadata centre²⁴).

Apart from a graphical web interface, the metadata environment also has a metadata service API. This API

²⁴ <https://metadata.geopunt.be/zoekdienst/apps/tabsearch/index.html?hl=dut>

applies the Open Geospatial Consortium (OGC) standard recommended by INSPIRE for the publication of metadata (Catalogue Service for the Web or CSW standard).

The metadata of all the metadata suppliers within GDI-Flanders are synchronised with the Geopunt catalogue on a daily basis. In this way a general overview is obtained of the offer of geographical information within the GDI-Flanders network.

At regular intervals, the Open Data catalogue harvests information from the Geopunt catalogue. As a result, the metadata of the geographical Open Data provision flows seamlessly into the Open Data catalogue.

Impact on the publication of data

INSPIRE and GDI-Flanders offer different ways for publishing data electronically. A distinction is made between view services and download services (INSPIRE View Services and Download Services).

View services are mainly aimed at an easy integration of geographical information into a cartographic (web) application. On the basis of a location, the user can retrieve a map cross-section in the form of an image (JPG, PNG, GIF). In order to be able to provide view services in conformity with INSPIRE and GDI-Flanders, the Technical Guidance²⁵ for the implementation of INSPIRE View Services must be followed. This guidance is based on the OGC Web Map Service (WMS) and Web Map Tile Service (WMTS) implementation standards.

The download services are provided for retrieving the geographical data itself. It is possible to offer data as a downloadable file or through direct access to the database. In order to be able to provide download services in conformity with INSPIRE and GDI-Flanders, the Technical Guidance²⁵ for the implementation of INSPIRE Download Services must be followed. This guidance is based on the OGC Web Feature Service (WFS) for direct access to the database and on the IETF ATOM standard for the publication of downloadable files. A WFS is an interface for retrieving and supplying geographical vector data. The data are supplied in an XML format that is specific for geographical data, better known as GML (Geography Markup Language). If the data is provided as a downloadable file, an ATOM feed must be set up that provides an overview of the available files. INSPIRE recommends the use of GML as file format for the exchange of data. Shapefile, DXF and KML (Keyhole Markup Language) can also be used as data exchange format.

For more information or support, please contact the Flanders Geographical Information Agency (AGIV) at contactpunt@agiv.be.

4.4 HOW TO GET STARTED WITH OPEN DATA IN YOUR PROJECTS?

In this section, we give extra tips on how you can identify an Open Data need within existing and/or new projects and immediately open up the data within the project's scope. This way, we want to show that you do not always have to start up new projects for opening up Open Data.

We distinguish the following options for this:

- Modification and/or extension of functionality of existing application/databases

What? Adding a new/additional requirement for creating an information flow to the Open Platform and publishing the data there.

How? By quickly assessing whether the information flow can meet the criteria of an Open Data stream and propose it as a scope extension within the new project. The Open Data Handbook contains sufficient indications for making this assessment.

Why? Including this as a requirement within an existing project extension will cost less than doing it afterwards or separately.

- When starting a new project

²⁵ <http://inspire.ec.europa.eu/index.cfm/pageid/5>

What? Always determine which information flows in the scope of the project can be opened up as Open Data streams

How? For every information flow that is identified within the scope of this new project it can be determined whether it can be opened up as Open Data. The collection of information flows can then together form the plan for opening up the data and formulate additional needs in terms of ETL technology and the number of flows.

Why? A new project offers an ideal opportunity to take Open Data into account from the very start rather than to think of it at a later stage.. This will reduce the total cost and provide an answer at an early stage as to which flows are required and when. The question will be asked anyway, so this is an opportunity not to be missed.

- For extensions of information flows on an existing Data warehouse environment:

What? Creating information flows to the Open Data Platform as part of flows that lead to a DWH environment.

How? A lot of information flows which now go to a DWH can be opened up to the Open Data Platform. The technical scenarios describe the additional steps and conformity requirements that are needed for this purpose.

Why? Building in an additional step into an existing process will have a lower impact and cost less than creating a completely new flow. In many cases the existing ETL scripts and technology can be applied and the existing scheduling can be used for opening up the data. This is thus a small extra step that generates great added value.

- Replacing Public Reporting / Viewers by Open Data

What? A lot of bodies themselves provide a public service to give people access to Open Data (even if the term itself did not exist at the time). Much of the functionality of these platforms (like basic reporting) can be met by the functionality of CKAN or by offering users the opportunity to import the file into their own reporting tools or technologies.

How? By offering the same information as Open Data and by removing the public component, if necessary.

Why? This allows users to decide for themselves how they want to format and save reports. The government will need to spend less time and effort on basic reporting services. This promotes the use of Open Data at a lower cost.

5 STRUCTURE YOUR DATASET(S) AND CHOOSE AN OPEN FORMAT OR API

Before publishing datasets on the Open Data Platform, these datasets or Open Data streams must meet a number of minimal requirements regarding quality and consistency. Below, recommendations are formulated about which file formats or APIs you should use to publish your data. Finally, a maturity model for Open Data is explained.

5.1 MINIMUM REQUIREMENTS

In this paragraph a number of minimum requirements are highlighted for opening up source data as Open Data. These requirements are formulated on the basis of the following dimensions:

- Safeguarding the quality of Open Data: all the technical characteristics which an Open Data stream must meet;
- Safeguarding the consistency of Open Data: ensuring the internal technical coherence of the published data.

In the next paragraphs some practical tips are given for each dimension on how these suggestions can be realised. These tips are neither exhaustive nor do they need to be read in a specific order.

5.1.1 QUALITY

In order to safeguard the technical quality of the data in an Open Data stream, it is proposed that each dataset meet the following criteria:

- Every Open Dataset must have a single and unique header row with a uniform description of the columns, either in Dutch or in English. The header row must form the first line of the dataset and should not be more than 1 line. The title should be unambiguous, short and not contain any spacing. The greatest possible use should be made of the standard glossaries to assign the title. Please refer to the section 'standards' further on in this document. For CSV files we strongly advise using an internationally accepted delimiter between the fields. A semicolon symbol (“;”) is the most generally accepted. If the settings are on Belgian standards, this is often a comma (“,”), which may, however, cause confusion with numerical values where a “.” is often used as decimal marker. For this reason it is recommended to use a point (“.”) instead of a comma (“,”) as decimal marker for numerical fields.
- Every Open Data stream must have a sufficient metadata description and at least satisfy the fields listed in the present document.
- Every Open Data stream must be associated with a process for publishing new or updated data (an update process) at frequent intervals. Of course, historical data need to remain stable. However, such data can also be extended periodically, e.g. every month or year.
- Every Open Data stream must have a version number so that it can be identified unambiguously and possible adjustments can be traced (i.e. version management).

5.1.2 CONSISTENCY

All Open Data that is published must be as consistent as possible across all data streams and bodies. If this is impossible across the bodies, then it should at least be consistent at the level of the public sector organisation itself. For instance, all addresses must always be provided in the same way to all users.. Otherwise, it will be difficult for users or businesses to form a consolidated picture of this.

We try to make this concrete using the following parameters:

1/ Open Data streams should be transparent:

- Each user of the Open Data stream should comprehend the unique origin of the data.
- Determine the status of the data (draft, validated).
- Re-use the (meta) descriptions of data that already exist within the public sector organisation.
- Describe the data model and metadata from the very start.

2/ Open Data Information should be correct and complete (within the context):

- Define (or create) validation rules on the basis of the data itself.
- Check dummy & default values (like M / F).

- Carry out a number of basic checks and correct any mistakes at the source (e.g. address information).
- Identify incorrect or missing fields.
- Aim at optimisation and avoid duplication of data.

3/ Open Data Information should be managed:

- Monitor the process from source system to Open Data stream.
 - Introduce quality tests for each step in the process.
 - Perform volume tests to prevent Open Data streams from becoming too large. Open Data is typically kept small in view of its re-use in mobile applications. If the stream grows larger (for instance for GIS data), then clearly mention this on the Open Data Platform or in the metadata.
- Separate online processing from batch processing, so as to make sure that opening up the Open Data streams will not put too great a burden on the public sector organisation's infrastructure.

4/ Open Data Information should be safe:

- Filtering out confidential data.
- Respecting privacy legislation.
- Anonymising data.
- Aggregating data, if necessary (granularity).

Recommendation 9: Check minimum criteria regarding quality and consistency before publishing a dataset on the Open Data Platform.

5.2 STANDARDS, OPEN FORMATS AND APIs

A lot of standards are available for modelling data and fields. Many agreements have also been made within the Flemish public administration. Below is a list of European or international agreements that may be applied to Open Data streams:

- General glossaries:
 - DCMI, see <http://dublincore.org>
 - OSLO, see <http://www.v-ict-or.be/kenniscentrum/OSLO>
- To describe persons:
 - vCard, see <http://en.wikipedia.org/wiki/VCard>
 - Core Person Vocabulary, see https://joinup.ec.europa.eu/asset/core_person/description
- To describe organisations:
 - Registered Organisation Vocabulary, see <http://www.w3.org/TR/vocab-regorg/> of https://joinup.ec.europa.eu/asset/core_business/news/publication-core-business-vocabulary-regorg-w3c-standards-track
- To describe locations:
 - vCard
 - Core Location Vocabulary, see https://joinup.ec.europa.eu/asset/core_location/release/100

- To describe public services:
 - Core Public Service Vocabulary of EU, see https://joinup.ec.europa.eu/asset/core_public_service/description
- To report non-urgent problems or suggestions:
 - Open311.org, see <http://open311.org/>

In addition, we refer to the initiative of the W3C organisation which goes a step further by concluding agreements between governments. See <http://www.w3.org/blog/2012/03/interoperable-governments/>.

In the Flemish public administration we want as a minimum to publish datasets that meet the 3-star description of the maturity model for Open Data (see 5.6). This means that we want to offer structured data in an open, non-proprietary file format.

Open formats for structured data include:

- CSV (http://en.wikipedia.org/wiki/Comma-separated_values), preferably in UTF-8 encoding
- TSV (http://en.wikipedia.org/wiki/Tab-separated_values), preferably in UTF-8 encoding
- XML (<http://www.w3.org/XML>)
- JSON (<http://www.json.org/>)
- ODF (<http://en.wikipedia.org/wiki/OpenDocument>)
- RDF/XML, Turtle, N-triple, JSON-LD (http://en.wikipedia.org/wiki/Resource_Description_Framework)

Open formats for geo-data include among other things:

- Shapefile (<http://en.wikipedia.org/wiki/Shapefile>)
- GeoJSON (<http://geojson.org/>)
- GML (<http://www.opengeospatial.org/standards/gml>)
- KML (<https://developers.google.com/kml/>)
- WKT (http://en.wikipedia.org/wiki/Well-known_text)

Recommendation 10: Use open formats, such as CSV, whenever possible.

Structured data that are not available in an open format (like MS Excel) can still be published by using the Datatank (software). The Open Data Platform is integrated with the Datatank for conversion services from MS Excel to open formats.

For more information about the use of the DataTank for conversion to open formats, please refer to 6.2.2. For more general information about the Datatank software, see Annex 2.

You can also consider not offering your datasets (exclusively) through data dumps, but also via an API. An API has the advantage that:

- only data are returned that are relevant for the client in response to a query, by using a filter, pre-sorted according to need, etc.
- the most recent data can be returned.

One disadvantage may be that if the API / service is not available, for example because of technical disruption,

the data can no longer be retrieved by the user.

If you want to build an API yourself, we recommend offering a RESTful Web API and granting access to the service without using an API key.

For a good description of REST and a RESTful web API, please refer to the relevant Wikipedia article (http://en.wikipedia.org/wiki/Representational_State_Transfer#RESTful_web_services).

If you already have a data dump, all you have to do is register your dataset on the Flemish Open Data Portal, more specifically in CKAN.

The integration of the Datatank software into the Open Data Portal offers an API in conformance with the aforementioned REST principles. The API documentation of the Datatank software is available on the page "Consuming Data" (http://docs.thedatatank.com/4.0/consuming_data).

Recommendation 11: If you want to build an API yourself, we recommend offering a RESTful Web API and granting access without using an API key.

5.3 LINKED OPEN DATA

For those who want to go further than the 3 stars and want to integrate into the Linked Open Data web, we offer guidelines in a separate document²⁶ on:

- how to allocate unique identifiers to entities (e.g. organisations, locations, services, etc.) that are related to the data you want to publish.
- which vocabularies (classes, attributes and relationships) to use in this description.

5.4 MATURITY MODEL FOR OPEN DATA

Given the aforementioned scenarios and minimum requirements for opening up source data as Open Data, we can always make improvements in terms of consistency and quality of the data. In order to make this transparent, we introduce a non-binding maturity model that describes criteria for achieving high quality.

The 5 star deployment scheme (<http://5stardata.info/>) for publishing Open Data already gives a number of characteristics for quality assurance. We have extended this model with additional criteria that Open Datasets should meet. These criteria apply to all the rules listed in the present document. Of course, a public sector organisation does not need to try to achieve the 5 star model immediately, since the target is currently at 3 stars.

The overview of the different levels of maturity is briefly described below. After that it is explained for each star model how it can be achieved.

Maturity level	Original explanation (English)	Additional parameters
★	Make your stuff available on the Web (whatever format) under an open licence	Publishing information without errors and with basic quality checks (for each dataset) on the Flemish Open Data Platform, supplemented with all metadata requirements.
★★	Make it available as structured data (e.g., Excel instead of image scan of a table).	As for 1 star, but now also with minimal consistency checks across the datasets (i.e. at the level of the publishing public sector organisation).
★★★	Use non-proprietary formats (e.g., CSV instead of Excel).	As for 2 stars, but information is opened up as part of a data management process including monitoring of the required quality and consistency checks (i.e. at the level of the publishing public sector organisation).

²⁶ http://www.opendataforum.info/files/URI_strategie.pdf

★ ★ ★ ★	Use URIs to identify things, so that people can point at your stuff.	As for 3 stars, but information is published on the basis of a Data Quality process, with the URI criteria being described in a separate document (at least at the level of the publishing public sector organisation).
★ ★ ★ ★ ★	Link your data to other data to provide context.	As for 4 stars, but information is published on the basis of Enterprise Data Quality with EA Governance, with the LOD criteria as described in the present document (at least at the level of the publishing public sector organisationpublic sector organisation, preferably on a general level).

6. PUBLISH YOUR DATASET(S)

The publication of Open Data does not require a great deal of technical investments. However, this does not alter the fact that the public sector organisation must provide the technical resources necessary for making the Open Data available to the public. The method chosen for making Open Data available is essential in this.

6.1 PRIOR TECHNICAL REVIEW

Each public sector organisation must provide the technical resources necessary for making the Open Data available to the public. This includes:

- Choosing the domain: the public sector organisation may opt to use its own website for making the data available, or it can do so via a separate website with its own domain name.
- Choosing the hosting: the public sector organisation must determine whether the data will be stored and made available on its own servers, or whether it will use third-party servers for this.
- Choosing the functionalities: it must be examined which database will be used, for instance whether a forum or a payment module is required. An assessment is to be made of the required server space, the total data consumption, the required speed.
- Managing the website and/or the portal: someone should be appointed to manage the website or portal. When doing so, it should also be determined which degree of availability is to be guaranteed, and which level of website monitoring and security is needed.
- Maintaining the services: if data are made available through services, someone should be appointed to monitor and guarantee the availability, functionality and performance of the services.

If the public sector organisation chooses not to make the data available itself, but to entrust this task to another public sector organisation or a third party acting as provider (for instance the dissemination by AGIV of geographical data belonging to other bodies), these technical decisions should be taken in consultation between the public sector organisation and the provider.

Recommendation 12: Provide the technical resources that are necessary for making Open Data available.

6.2 THROUGH IN-HOUSE WEBSITE (DATA DUMP OR API) OR UPLOAD TO CKAN

The public sector organisation decides whether it makes its Open Data available through its own official website, through a separate website, or through the services of a third-party provider. In order to increase the discoverability of the data for users, it is important to always include a reference to the data entered on the Flemish Open Data Platform (CKAN).

Recommendation 13: Make the data available through the in-house website or the website of a third-party provider and post a link to where the data can be found on the Flemish Open Data Platform (CKAN).

After the public sector organisation has decided how it will organise its Open Data policy, it can start with the practical implementation of the decisions taken. This implies, among other things, providing the necessary information to potential users.

7. DOCUMENT YOUR DATASET(S)

You have now published your dataset as a data dump or through an API. Take into consideration that for potential users of your data, it is not always easy to assess whether your data could be relevant for them.

For this reason, it is recommended to appoint an Open Data contact point within your organisation and to communicate a contact address for information and feedback to potential users of the data. Moreover, it is essential to inform potential users about the terms and conditions of use, (possible) charges and guarantees as to the availability. Finally, it is vital to also publish a page (or instructions for use) that provides the necessary context for the published data. It is recommended to make this information available in multiple languages.

7.1 OPEN DATA CONTACT POINT

Although opening up public sector data is not a big task and will only require limited time and effort from the public sector organisation, the public sector organisation can implement this task more efficiently by appointing one person or department to be in charge of the Open Data policy within the organisation. This responsibility is usually assumed by the person or department that is responsible for the public sector organisation's internal information policy, so as to allow a streamlined policy. If another person or department is appointed, this person or department should stay in close contact with the person in charge of the internal information policy.

Apart from the position of Open Data policy manager, it is also recommended to appoint an Open Data contact point, although this may of course be the same person or department. This person or department can play a three-fold role as Open Data contact point. Firstly, an Open Data contact point plays an important role in streamlining the internal Open Data policy. In addition, it can also offer great advantages for the organisation of data streams within the public sector organisation itself. Because the contact point knows which data are available within the public sector organisation, it can make sure that data are produced or purchased only once and are further shared within the public sector organisation. Secondly, the contact point may establish a direct relationship with the Flemish public administration. Through this relationship, the public sector organisation can ask any questions it has about Open Data to the administration and the Flemish public administration knows to whom it can provide support in the Open Data policy. Thirdly, the Open Data contact point has an important function with respect to citizens: even if the data is readily available, it is still possible that citizens have other questions regarding the format of the data, its origin, etc. That is why it is important for citizens to know exactly to whom they can pose their questions.²⁷

Recommendation 14: Appoint a person or service to be in charge of the Open Data policy within the public sector organisation. Establish an Open Data contact point for communication within the public sector organisation, with the Flemish public administration and with citizens.

²⁷ Cf below.

7.2 CONTACT ADDRESS FOR INFORMATION AND FEEDBACK

Even if an Open Dataset is clearly described in the related metadata, the re-user will sometimes still require additional information, want to ask more questions or report errors or deficiencies in the data. To be able to do so, the re-user needs an address where he can send these questions or reports. The public sector organisation can provide this by putting an e-mail address on the website that can be contacted for information and/or feedback, or create a web form that can be completed online.

It is part of the Open Data contact point's remit to answer the questions and/or respond to remarks. To encourage the use of the data, and in particular the provision of feedback, it is important that citizens' questions and/or remarks are answered with minimal delay.

Recommendation 15: Put a contact address or web form on the website that re-users of the data can use to ask for further information or give feedback.

7.3 TERMS AND CONDITIONS OF USE

In the preparatory stage, the public sector organisation has chosen a specific licence under which it wants to make its data available. In order to achieve the greatest possible harmonisation and simplification, it is preferable that the Flemish model licences are used. This not only prevents fragmentation and optimally stimulates the use of Open Data, but it also ensures that the public sector organisation itself does not have to invest in creating its own terms and conditions of use.

When the Flemish model licences are used, the public sector organisation must indicate in a clearly visible place that the data are made available under the applicable Flemish Open Data Licence, and add a link to where the licence can be found on the website of the Flemish public administration. If the public sector organisation wants re-users to apply a particular attribution statement, it should add this statement as well. Examples of attributions are given below for the different possible licences:

- “[Name of the dataset] is made available under CC0. The complete text of the English deed is available at <http://creativecommons.org/publicdomain/zero/1.0/legalcode>. A translation of the CC0 deed can be found at <http://creativecommons.org/publicdomain/zero/1.0/>”
- “[Name of the dataset] is the property of [name of public sector organisation]. [Name of the dataset] is made available under the Flemish Free Open Data Licence [link to licence]. The following attribution statement should be included for every type of use: [required attribution statement]. For further information, please contact [contact address]”.
- “[Name of the dataset] is the property of [name of public sector organisation]. [Name of the dataset] is made available under the Flemish Open Data Licence against a Reasonable Charge [link to licence]. The following attribution statement should be included for every type of use: [required attribution statement]. For further information, please contact [contact address]”.
- “[Name of the dataset] is the property of [name of public sector organisation]. [Name of the dataset] is made available for non-commercial re-use under the Flemish Free Open Data Licence [link to licence] and for commercial re-use under the Flemish Open Data Licence against a Reasonable Charge [link to licence]. The following attribution statement should be included for every type of use: [required attribution statement]. For further information, please contact [contact address]”.

By referring in a specific manner to the Flemish Open Data Licence, Google and search engines learn about the fact that a reference is made to a particular licence and that the material is available under this licence.²⁸ To that end, a reference must be made in HTML to the Open Data Licence through the "rel=licence" attribute in the link to the licence.

Apart from the legal text, the model licences will also be made available in a machine-readable version, so that Google, other search engines and web crawlers can also automatically recognise the licence (and its terms and

²⁸ Zie <http://microformats.org/rel-license>

conditions).

Recommendation 16: Use the model licences of the Flemish public administration and add a link to them in the licence provision to the data, using the "rel=licence" attribute.

7.4 CHARGES

It was already explained earlier which options bodies have to impose a reasonable charge for their Open Data. During the implementation, it is essential that re-users obtain a clear picture of how much they will have to pay for using the data and that the payment can be made quickly and transparently.

The amount of the reasonable charge must be clearly indicated with the dataset. When this amount is, for instance, a fixed amount for the whole dataset, the amount can be mentioned next to the dataset. When the amount is calculated on the basis of the volume of used data, the calculation of the charge must be explained in a transparent manner. This may be done, for instance, by adding a link to an information sheet about the charge in the metadata or in the reference to the licence. This information sheet is not part of the licence, but is added as "instructions for use". In this way it can be more easily changed. Account is of course to be taken of the user's legal certainty, and the price can only be adjusted after it has been announced clearly and in a timely manner. Information about the charge should include the method of calculation and the reasons for the imposed charge, as well as the charging method.

Asking for a reasonable charge also means that re-users will not receive access to the data until payment has been made. Consequently, a registration system is to be set up that does not give re-users the password for accessing the data or service until the public sector organisation has received payment. The related provision on the website could be as follows:

"In order to gain access to the data, please complete the application form [link to online form or Word form to be e-mailed] and pay the aforementioned amount into account number [...] with mention of [...]. As soon as the payment has been received, you will be sent a password for gaining access to the data".

The public sector organisation may also consider accepting payments by credit card. After verification of payment, immediate access can then be given to the data.

Since re-users must be able to rely on the price indicated on the website being the correct one, it must be clearly specified, each time the prices may be changed, when the new prices start to apply and until when.

Recommendation 17: If a charge is made, clearly indicate to users how and how much they must pay in order to obtain access to the data or service.

7.5 GUARANTEES AS TO AVAILABILITY

When the data are made available online through a bulk download, it is of course important that these data are actually available, that there are no dead links and that the website is online. However, continuity is even more essential when the data are made available through a service and are integrated into one of the user's services or applications, for instance via an API or plug-in. For this reason, it is vital that users are informed about the service level of the provided service. In case data are made available via a third-party provider, it is essential to harmonise the potential service level between the public sector organisation and the provider, and the provider will be the ideal party to draw up the service level engagement.

In principle, it is not mandatory to incorporate this service level engagement into the licence. Bodies are advised to rather provide the information as "instructions for use" or an "information sheet" to the service, via a link to the information concerned or via the metadata, so that they are not part of the binding licence and can be modified more easily unilaterally by the public sector organisation when circumstances require it. The following elements can be included in this information sheet;

- A best efforts obligation to have the service function permanently, but no guarantee of permanent 24/7 availability (possibly a guarantee of 90% availability, for instance).
- A warning that the service may be discontinued (preferably with a sufficiently long advance notice or

transition period).

- An indication of the service response time.
- An indication of the capacity of the service, for instance regarding the number of simultaneous requests.
- A warning that access to the service will be closed in case of overload or abuse by users (including a description of what is regarded as overload, for instance 95% of the requests originate from one single user).

It is not mandatory to include such a service level engagement. Moreover, such information is not always equally useful or important. When a link to a bulk download no longer works, the service level will only imply that the public sector organisation corrects the link when it is informed of the problem. In any case, if no service level engagement is laid down for a distribution channel of the data, the public sector organisation is expected to make every reasonable effort possible to provide the data, with due diligence.

Recommendation 18: If the data is made available via a service, add an information sheet or service level engagement to the service, explaining the performance of the service and the expectations users may have of the service's operation.

7.6 ADDON INSTRUCTIONS

Some examples of such an "instructions for use" page:

- <http://aps.vlaanderen.be/sgml/largereeksen/704.htm> with an explanation on the dataset "Participation in Pop or Rock Concerts by Flemish people by gender"
- <http://epp.eurostat.ec.europa.eu/portal/page/portal/population/introduction> for an explanation on "EU population statistics". A number of things are indeed important for end users to be able to clearly understand your data.

7.6.1 THE DATA ITSELF

Clearly describe which entities are specified in your dataset (e.g. persons, organisations, events, etc.).

Also describe the scope and granularity of your dataset, both geographically and in time. For instance, you may have data for the whole of Flanders broken down by province. You may have these data at your disposal for the years 2009 through 2012, broken down by quarter.

Also explain which attributes and relationships you use in the dataset to describe the entities. In other words, describe your logical data model.

If you use codes, make sure their meaning is known as well.

7.6.2 THE PROCESS

Furthermore, the description of why these data are collected and the processes used in collecting the data adds context. Describe therefore who collects the data and how frequently, and who manages the data. Also explain for what purposes the data are used and how they are used. List the processes using these data, if any.

Recommendation 19: Make an accompanying page for each of your datasets which explains in understandable language what the data are about, why they have been collected and what they are used for.

7.7 WEBSITE LANGUAGE

The first language in which the data, services, licences and information are to be displayed is Dutch. However, considering the growing need for data for cross-border applications, it is becoming increasingly important to make the data portals and websites on Open Data also available in other languages. Insofar as possible, the bodies are advised, as a minimum, to provide information also in English.

In order to cater for this, the Flemish public administration will also make an English version of the Open Data Licences available for bodies to refer to.

Recommendation 20: Place information on the data on the website in English as well, and refer to the English version of the Flemish Open Data Licences.

8. MAKE YOUR DATASET(S) DISCOVERABLE

You have published your dataset, but now you also want to make it more discoverable. The Flemish public administration has taken a measure to that end, namely the Open Data Portal which serves as some kind of "yellow pages".

This Open Data Portal uses CKAN software (<http://ckan.org/>).

To make your dataset known on this Open Data Portal, you must complete a metadata sheet.

We will first describe the possible metadata and then explain how to complete this metadata sheet in CKAN.

8.1 METADATA

8.1.1 WHAT IS METADATA?

According to Wikipedia, metadata is the term for describing characteristics of certain data. Metadata is thus "data about data". The metadata on a specific document (the data) can for instance be the author, the date on which the document was written, the publisher, the number of pages and the language of the text of the document. The advantage of explicitly storing metadata on the data to which it relates is that the data can be traced more easily. For instance, in a search engine that uses metadata, it is possible to immediately search for documents written by a specific author. This is more difficult with a full-text search, which means without using metadata, because each document in which the name of the author appears will be found. This can include many more documents than the ones that were actually written by that person.

Very frequent use is made of metadata to search (and find) Open Data. In all Open Data platforms a layer of metadata will be required to make the dataset in question discoverable, both for the catalogue itself and on aggregated platforms.

Apart from the search on a local platform, metadata also offers the key to look for Open Datasets on other Open Data platforms. For this reason, guidelines have been drawn up about which metadata fields are required for the Open Datasets of Flanders.

8.1.2 HOW TO PRODUCE METADATA?

The creation of metadata for Open Datasets may on the one hand be supported by (semi) automatic processes, like:

- Document properties generated in (Office) support software, like Creation Date
- Spatial and temporal information as captured by cameras, sensors, etc.
- Information from workflow publication, like the location or URL of the source

On the other hand, some properties require human intervention or production:

- What are the data about (possibly linked to an item or source of information)?
- How can this dataset be used (for instance, link to a model licence)?
- Where can you find more information about the source itself (for instance link to a website or other document)?
- Attributes describing the quality of the information (like draft, for review, not yet validated, temporary).

The method for maintaining metadata should also be adjusted to the published data. If the data do not change frequently, metadata can remain relatively stable or be adjusted in bulk at periodical intervals (like e-mail address for feedback loop). If data change often (like real-time sensor data), the metadata must be closely linked to the workflow data and changes should be made almost at once.

Furthermore, Open Datasets are part of an environment that changes. These changes need to be reflected in the metadata.

- **Organisations** tend to change, merge or transfer responsibilities to another organisation or public sector organisation.
- **New applications** can link Open Datasets with each other and may therefore require additional metadata or create the need for harmonisation of metadata across various bodies (consistency of metadata).
- **Metadata evolution:** the current requirements are minimal and will continue to evolve.

This drives the need to constantly monitor the metadata and adjust them to the evolution. However, we want to use this as pragmatically as possible, in such a way that it does not make (constant) demands on the budget.

8.1.3 METADATA GUIDELINES

Metadata is crucial for being able to correctly identify datasets and make them discoverable across the different platforms. For this reason, we propose that each public sector organisation makes an effort to provide a minimal set of metadata for every published dataset.

In 2013, agreements have been made between the Government of Flanders, the federal authorities and a number of cities like Ghent and Antwerp to implement a consistent policy in this respect whenever possible.

These agreements are based on the European DCAT recommendations and translated into advice for all publishing bodies, irrespective of the (regional) government, city, province or Region in Belgium. These agreements are included in Annex 4 ("Metadata mapping DCAT – CKAN").

One positive result of the application of these recommendations is that wherever you publish your dataset, you can in principle also easily and even automatically find it on other platforms. As a result, each public sector organisation can choose where to publish Open Datasets, but these datasets can still be harvested on another platform. For instance, all Open Datasets of AGIV which are designated as "Open Data" can also be automatically published on the Flemish Open Data Platform or even on a higher level to the European Open Data Portal.

The possibilities for providing metadata depend on the type of platform used. For now, Flanders has chosen the CKAN platform (see below). However, CKAN applies a specific approach with respect to metadata, making a

number of fields mandatory and allowing a number of free fields to be added.

Although providing metadata is a requirement to allow finding datasets quickly and easily, it still depends on the used platform whether these metadata fields are mandatory and how they must be completed. CKAN has its own specific approach, which is why we can only propose a guideline for a minimum common set of mandatory fields.

The mandatory fields within CKAN are, for the moment, limited to the following:

CKAN Veld (NL)	Field (ENG)	link with DCAT standards	Tip
Titel	Title	dct:title (dcat:Dataset)	Contains an unambiguous title of the dataset.
Omschrijving	Description	dct:description (dcat:Dataset)	Contains the description of the dataset as text. Try to keep the description as short and as relevant as possible. The idea is not to explain the operation of a public sector organisation or department here, but only to indicate what data is made available.
Licentie	Licence	dct:licence (Dataset)	The licence model under which this dataset is published. Choose one of the values from the drop-down selection. The values correspond to the standard models provided by the Flemish public administration.

In Annex 3 we have included a proposal of how we can use the free fields of CKAN to be maximally compatible with the agreements made at the Belgian level in keeping with the European DCAT recommendations. We indicate for each field how to name this in CKAN. Unfortunately, this is not yet enforceable. Therefore, special attention needs to be given to a correct spelling when defining the free fields. The Flemish public administration continues to be in contact with the CKAN development team to close this gap in the future.

Recommendation 21: In order to allow for a smooth exchange of dataset descriptions, we recommend using as many fields as possible from the DCAT profile agreed at the Belgian level, even within CKAN through free fields, if necessary. See Annex 3 for an overview.

8.2 ADDING METADATA IN CKAN

In this chapter we discuss in greater detail which metadata can be added using the "Create dataset" wizard in CKAN and how to do so.

For each field/property we give the label as shown in the interface, the name of the field in CKAN, the corresponding field from the specification of the DCAT Application Profile for data portals in Europe and the allowed/expected values for this field.

But before you can do this, you must first create an account on CKAN.

8.2.1 CREATING AN ACCOUNT

Click on "Register" in the top right-hand corner and the following screen will be displayed.

/ Inschrijving

Why Sign Up?

Dreëer datasets, groepen en andere interessante dingen

Registreren voor een nieuwe Account

GEBRUIKERSNAAM:

VOLLEDIGE NAAM:

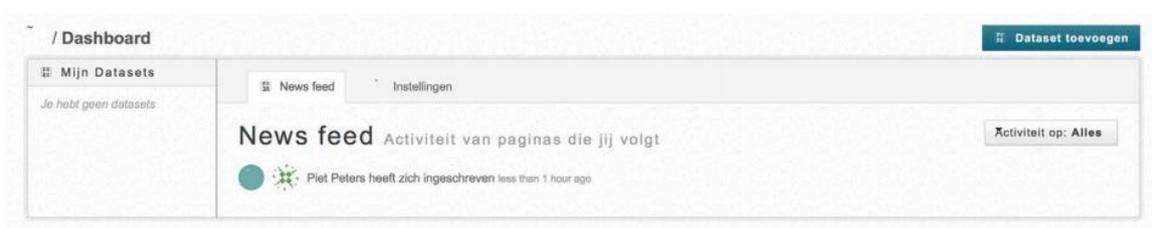
EMAIL:

WACHTWOORD:

BEVESTIG:

[ACCOUNT AANMAKEN](#)

Important: the user name should only consist of lower case alphanumeric (ASCII) characters or '-_'. If you are successful, the following screen will appear:



8.2.2 CREATING A DATASET

To create a dataset, you must log in, using your account details. When you are logged on, click on the button "Creating a dataset".



Page : Create a dataset

You will first see a page for describing your dataset.

These are the fields provided within CKAN:

Label in page	CKAN field	DCAT mapping	Values
TITLE	Title	dct:title (Dataset)	string (titlecase)
DESCRIPTION	Notes	dct:description (Dataset)	String
TAGS	Tags	dcat:keyword (Dataset)	multiple with separator ','
LICENCE	Licence	dct:licence (Dataset)	One of the drop-down list
ORGANISATION	owner_org	dct:publisher	One of the drop-down list

Within CKAN this is presented as follows:

/ Datasets / Creëer Dataset

Wat zijn datasets?

Datasets worden gebruikt om gerelateerde data te groeperen. Deze kan vervolgens gevonden worden onder een URL met een beschrijving en licentie informatie.

1 Creëer dataset

2 Data toevoegen

3 Aanvullende data

TITEL:

URL: **BIJWERK**

OMSCHRIJVING:

Hiervoor kan je gebruik maken van Markdown formatting here

TAGS:

LICENTIE:

Definities van licenties en aanvullende informatie is te vinden op gsaandefinition.org

Belangrijk: Door het indienen van inhoud gaat u ermee akkoord dat deze een bijdrage levert aan de Open Database License

Afbreken **VOLGENDE: DATA TOEVOEGEN**

On this screen you should enter the general characteristics of the dataset: the name (consisting of lower case alphanumeric (ASCII) characters or '-_'), a description, keywords and the chosen model licence.

The results can be seen in the following screen:

/ Datasets / Creëer Dataset

Wat is een hulpbron?

Een hulpbron kan een bestand of link zijn die over nuttige informatie beschikt

1 CREËER DATASET

2 Data toevoegen

3 Aanvullende data

LINK NAAR EEN BESTAND LINK NAAR EEN API UPLOAD EEN BESTAND

BRON:

NAAM:

OMSCHRIJVING:

Hiervoor kan je gebruik maken van Markdown formatting here

FORMAAT:

Dit is automatisch gegenereerd. U kunt dit bewerken als u wilt

VORIGE **OPSLAAN & EEN NIEUWE TOEVOEGEN** **VOLGENDE: AANVULLENDE INFORMATIE**

After that, go to the following screen "Adding data".

Page: Adding data

The following page allows you to add your distribution(s). These are the fields you can use for that purpose:

Label in page	CKAN field	DCAT mapping	Values
SOURCE	url	accessURL (Distribution)	url
NAME	Name	dct:title (Distribution)	string (titlecase)
DESCRIPTION	description	dct:description	String
FORMAT	Format	dct:format (Distribution)	file extension like csv,

There are three types of distribution and each has its own screen:

Screen 2.1: Data - link to an already published data dump

If you already published a data dump, then select "Link to a file" at the top of this screen. The following screen will appear:

The screenshot shows the 'Creëer Dataset' interface with three steps: 1. CREËER DATASET, 2. Data toevoegen (active), and 3. Aanvullende data. Under 'Data toevoegen', there are three options: 'LINK NAAR EEN BESTAND', 'LINK NAAR EEN API', and 'UPLOAD EEN BESTAND'. The 'BRON' field is filled with 'http://www.proxml.be/ckan/categorieën2.xlsx'. The 'NAAM' field contains 'Overzicht gebruikte thema's op open data portalen in ms-excel formaat'. The 'OMSCHRIJVING' field is empty with a placeholder 'Een nuttige informatie over de data'. Below it, there is a note: 'Hiervoor kan je gebruik maken van Markdown formatting here'. The 'FORMAAT' dropdown is set to 'xlsx'.

The most important field is "source" where you enter the (HTTP) address where the data can be found/downloaded.

Also enter the file extension in lower case (csv, xml, json, xls, xlsx, ...) under format.

Converting data into another format using Datatank software.

Datatank software supports the conversion of the following formats: xls(x), json, xml, csv, shp.

If you wish to use Datatank software to have your dump (like xls) converted to other formats, then tick "USE THEDATATANK".

FORMAAT:

 Dit is automatisch gegenwoord. U kunt dit bewerken als u wilt.

USE THE DATATANK:

XLS SHEET:

HEADER ROW:

START ROW:

To convert Excel files you are asked to add 3 additional parameters:

- The name of the sheet (page in Excel)

1	NAAM	Gemiddelde besteding korte vakantie (in euro)							
2									
3	DIMENSIES	Ruimte	Vlaanderen						
4		Tijd	1996-2010 (2-jaarlijks)						
5									
6									
7	BRON	WES, Onderzoek & Advies							
8									
9	Voor meer informatie	http://aps.vlaanderen.be/sgmi/largereksen/1729.htm							
10									
11	gemiddelde besteding (euro)								
12		2010	2008	2006	2004	2002	2000	1998	1996
13									
14	Per persoon en per korte vakantie	182	180	160	160	153	150	132	131
15	Per persoon en per nacht	92	82	75	75	72	72	65	63
16									
17									

1.1.2.1

In this case: table (see left below).

- Whether a header row is available or not (take '0' if no, '1' if yes). In the example the header row is the row containing the years.
- On which line the table begins. In this case on line 12, the line containing the years. ATTENTION: Since Datatank software index starts with 0, you must enter 11 here.

For the table above this will give the following result:

FORMAAT:

 Dit is automatisch gegenwoord. U kunt dit bewerken als u wilt.

USE THE DATATANK:

XLS SHEET:

HEADER ROW:

START ROW:

The result of entering these parameters is that the xls will be displayed as follows in CKAN thanks to the Datatank:

URL: <http://www4.vlaanderen.be/darwin/C:/ms/Exceltabellen/leersme/vakantiegereguleer/TOGRVWVAD15.xls>
 Er is geen beschrijving voor deze hulpbron

Go to the data API (powered by The DataTask)

OPENDATAFORUM/08073503-CD79-442E-8045-F7A11F2084EC

column_0	2010	2008	2006	2004	2002	2000	1998	1996	column_9	column_10	column_11	column_12
Per persoon en per korte vakantie	152	180	160	160	155	150	152	121				
Per persoon en per nacht	50	52	75	75	72	72	65	63				

Formaat
 JSON XML CSV PHP

Beschrijving
 No description provided

Source Type
 xls

Make a note of the conversion possibilities to json, xml, csv and php on the right hand side.

Screen 2.2: Data through API

If you publish your data using an API, then click on the option "link to an API" at the top. The following screen will now appear:

/ Datasets / Creëer Dataset

Wat is een hulpbron?

Een hulpbron kan een bestand of link zijn die over nuttige informatie beschikt

1 CREËER DATASET 2 Data toevoegen 3 Aanvullende data

LINK NAAR EEN BESTAND **LINK NAAR EEN API** UPLOAD EEN BESTAND

BRON:

NAAM:

OMSCHRIJVING:

Hiervoor kan je gebruik maken van Markdown formatting hier

FORMAAT:

Dit is automatisch gegenereerd. U kunt dit bewerken als u wilt

VORIGE OPSLAAN & EEN NIEUWE TOEVOEGEN VOLGENDE: AANVULLENDE INFORMATIE

Indicate in the source field at which HTTP address the API is addressable.

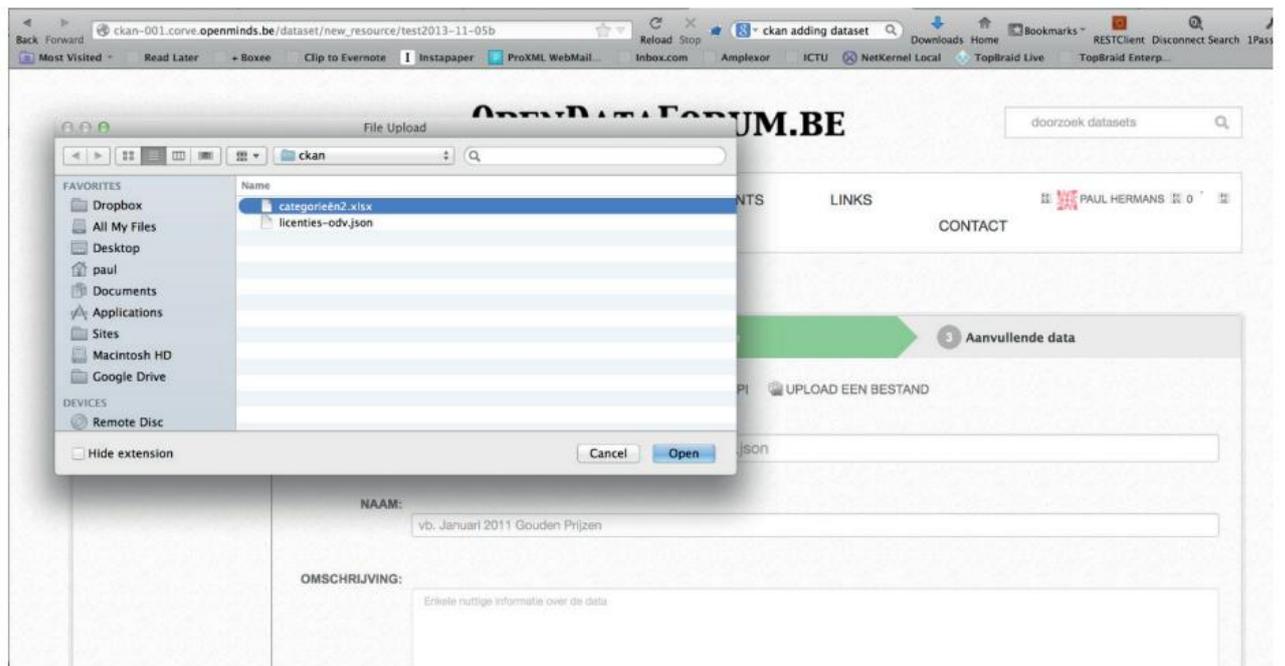
Also enter the API documentation or where it can be found in the description field.

Screen 2.3: Upload of a data dump

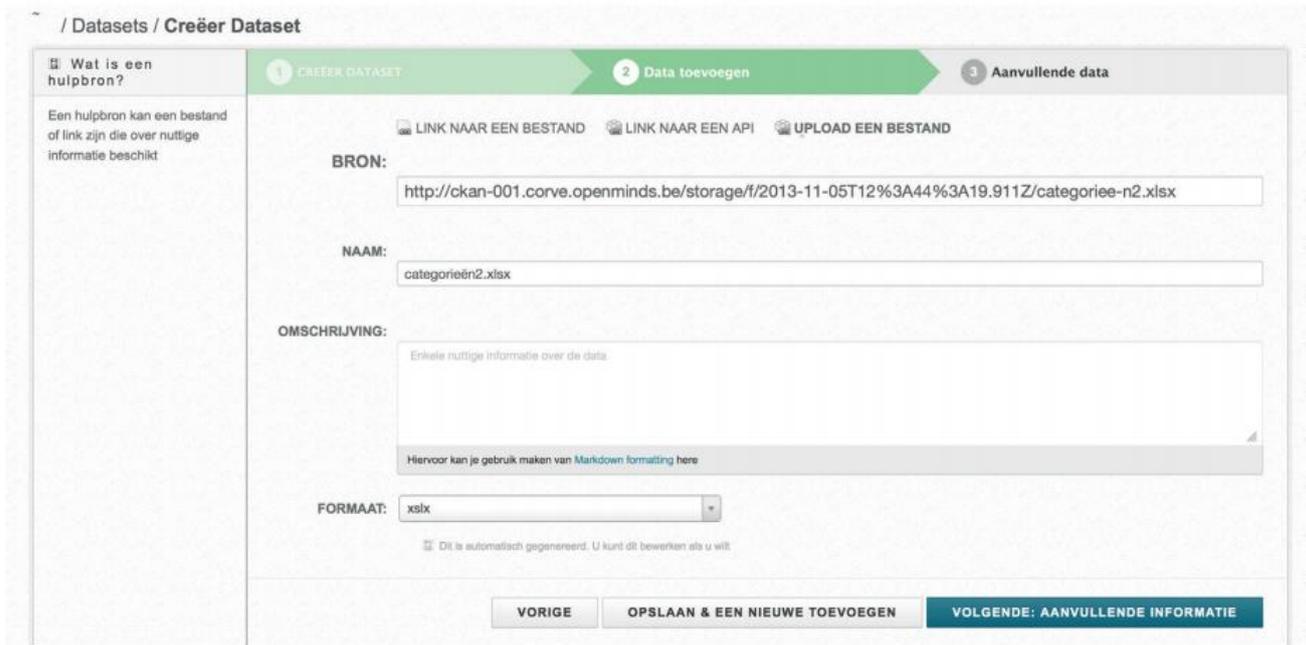
A third possibility is that you have a data dump locally, which you still want to publish. For this publication, you wish to use CKAN.

In this case choose the option "upload a file".

You now upload your local file to CKAN.



The other metadata is similar to that of the other options.



If you have entered the distribution(s), go to the next (third) screen.

Page: Additional data

This page allows you to add additional metadata.

/ Datasets / Creëer Dataset

Wat zijn datasets?
 Datasets worden gebruikt om gerelateerde data te groeperen. Deze kan vervolgens gevonden worden onder een URL met een beschrijving en licentie informatie.

1 CREËER DATASET 2 DATA TOEVOEGEN 3 Aanvullende data

AUTEUR: Vincent De Roover

AUTEUR EMAIL: Vincent.deRoover@stad.antwerpen.be

BEHEERDER: Piet Peeters

BEHEERDER EMAIL: piet@peeters.be

CUSTOM FIELD: KEY: laatst gewijzigd VALUE: 2013-11-05

CUSTOM FIELD: KEY: VALUE:

CUSTOM FIELD: KEY: VALUE:

GROEP TOEVOEGEN: Bestuurszaken

VORIGE EINDE

These are the fields provided as a standard:

Label in page	CKAN field	DCAT mapping	Values
VISIBILITY	Private	--	One of the drop-down list
AUTHOR	Author	--	String
AUTHOR EMAIL	author_email	--	Email
MAINTAINER	maintainer	--	String
MAINTAINER EMAIL	maintainer_email	--	Email
ADD GROUP	Groups	dcat:theme	One of the drop-down list

CKAN also allows you to add extra metadata using free fields.

Free fields

These are key/value pairs that can be freely added.

We propose the following options which are in line with the aforementioned DCAT profile:

Key	DCAT mapping	Values
Language	dct:language	One of (NL, FR, DE, EN)
Update frequency	dct:accruePeriodicity	One of the list with frequencies (approximately) (cf below)
Date of issue	dct:issued	date (xsd:date format)
Last modified	dct:modified	date (xsd:date format)
Version	--	MAJOR.Minor.patch (cf semantic versioning)
Version notes	adms:versionNotes	String
Temporal coverage	dct:temporal	string, for example: 1992, 1995, 1998, 2001, 2004, 2007, 2010
Spatial coverage	dct:spatial	string, for example: Flemish Region + Brussels
Conforms to standard	dct:conformsTo	cf. list of examples of standards below
Number of bytes	dcat:byteSize	The size of the distribution in bytes
Media type	dcat:mediaType	A MIME type as defined by IANA
Rights	dct:rights	A string describing the rights of the publisher
Status	adms:status	One of ("completed", "under development", "deprecated", "withdrawn"). For purposes of explanation

List of frequencies

- Annually
- Bi-annually
- Bimonthly
- Biweekly
- Continuously
- Daily
- Irregularly
- Monthly
- Three-monthly
- Half-yearly
- Half-monthly
- Half-weekly
- Three times a month
- Three times a week
- Three times a year
- Three-yearly
- Weekly

List (non-exhaustive) of standards

Name	Organisation	Link
Core business vocabulary	ISA	https://joinup.ec.europa.eu/asset/core_business/description
Core person vocabulary	ISA	https://joinup.ec.europa.eu/asset/core_person/description
Core location vocabulary	ISA	https://joinup.ec.europa.eu/asset/core_location/description
Core public service vocabulary	ISA	https://joinup.ec.europa.eu/asset/core_public_service/description
Oslo	V-ICT-OR	https://joinup.ec.europa.eu/catalogue/asset_release/oslo-open-standards-local-administrations-flanders-version-10?lang=nl

Semantic versioning

The value in the field "version" consists of three consecutive numbers with a point in-between, like for instance 5.12.0, where:

the different parts are designated (named) as MAJOR.Minor.patch, or in short "M.m.p"

the numbers will increase sequentially and be attributed in such a way that they express a compatibility of the exchanged file with other versions.

A different M 'MAJOR' version suggests incompatibility.

The same M 'MAJOR' version suggests compatibility, with the client being able to assume that his valid automatic processing of the data will still function.

Differences in m 'minor' suggest upward compatibility, which means that a client who interpreted the data correctly for an older (smaller m) version will also do so with a more recent version of the data (higher m), but may possibly lack new nuances.

8.2.3 ADDING METADATA THROUGH THE CKAN API

It is also possible to have the metadata creation of your datasets take place automatically using the CKAN API. This is explained in Annex 3.

9. EVALUATE YOUR OPEN DATA PRACTICE

Flanders is on the threshold of an unstoppable evolution towards Open Data. This also means that the Open Data policy may be adjusted, depending on the larger availability of data or the greater demand for it. In addition, possible adjustments in the legal framework around Open Data must be taken into account, such as the upcoming revision of the European Directive on the re-use of public sector information. Finally, the practical experience of bodies is also a good indicator of the success of the Flemish Open Data policy.

The Flemish public administration would like to be one of the frontrunners in the field of Open Data within the European Union. To that end, it must keep abreast of Open Data policies. For this reason, the Flemish public administration wants to evaluate on a regular basis to what degree the Open Data idea has permeated Flanders. The contribution by the bodies making data available is essential in this. This is why the Flemish public administration attaches great importance to good interaction with the bodies, giving them the opportunity to share their experiences and concerns. Concrete information, like how often certain data were downloaded, which data were retrieved and how frequently services were used, is of great value here.

Recommendation 22: Evaluate the success of your public sector organisation's Open Data practice. Share your experiences with other bodies via the Open Data Forum.

ANNEX 1 MASTER DATA MANAGEMENT

WHAT?

Databases change a lot and age quickly. Each year, 30 to 40% of the data in an address database change. Moreover, many bodies use several databases and changes and additions are often carried out by several people in various divisions. How can you keep these data up-to-date and establish mutual links between the different databases within the Flemish public administration?

Master Data Management (MDM) is the integration of data and databases into one clean data file of a high quality and consistency. Making available accurate, reliable and consistent information in all business systems can save time and money.

The biggest importance of Master Data Management (MDM) is to unambiguously establish and manage master data of an organisation, in this case the Flemish public administration as a whole. This is necessary because in almost all cases master data are stored in multiple (information) systems within an organisation and are used by several users. MDM makes sure that the master data are consistent and are preferably given a central place within the organisation. All information systems within the organisation will then use this one central source of master data. MDM also ensures that the descriptions and definitions of information correspond to the descriptions and definitions containing the same information with citizens or businesses.

Supporting processes in MDM include source identification, data collection, data transformation, normalisation, rule administration, error detection and correction, data storage, data distribution and data management.

RELEVANCE FOR OPEN DATA

Within the administration, the responsibility for the different applications is divided between several bodies, bar some exceptions. This is also the case for the relating master data. A lot of bodies are faced with the challenge of essential master data often being stored in several systems, like contacts, opening hours, etc. Bodies often do not sufficiently recognise that incorrectly entered master data in one single application can have serious consequences for another application.

Reporting systems experience a lot of problems if the master data is not consistent. In this respect it can be stated that data quality problems resulting from inconsistent master data are one of the main reasons why many Business Intelligence projects do not yield the desired result.

This problem also goes for Open Data. However, in this case it is not the internal operation, but the citizens who will be affected by poor master data. If every public sector organisation will open up data through Open Data using a different method for modelling addresses, for instance, citizens will have difficulty in sticking all the pieces together or writing a consistent application for it. For this reason, we ask to pay additional attention to the production and management of master data, at least at public sector organisation level, but preferably throughout the administration.

The good news is that the techniques for realising this for Open Data are not all that different from the techniques for applications and/or reporting systems. In other words, the process for producing and managing master data for traditional applications can also be applied to Open Datasets.

The text above is based on <http://ict-mdm.wikispaces.com/>. The licence agreement of this site (Creative Commons Attribution Share-Alike 3.0 Licence) allows us to copy this text and adjust it whenever relevant.

STEPS OF AN MDM PROCESS

The following steps generally apply for the monitoring of data quality and consistency and are therefore also applicable to finding and opening up Open Data streams. They are probably used in some or other way to manage data, but maybe not as emphatically as is indicated below. Therefore, this list should above all be regarded as a checklist for monitoring data quality and consistency and Open Data streams. The larger the number of data and the public sector organisation, the more formally a process should be regulated. Even at the level of the Flemish public administration it is interesting to keep the following steps in mind when managing all Open Data streams.

1/ Identifying Open Data master data.

On the basis of a number of criteria, it is established which data are (not) regarded as master data and included in an Open Data management process. We refer in this context to criteria for opening up data that have been listed earlier in the present document.

2/ Identifying source systems.

Where does the master data and its metadata originate from and which source systems do they produce?

3/ Collecting and analysing metadata on master data.

Collecting underlying metadata on master data. Please also refer to the previous chapter on metadata for Open Data which describes the necessary fields.

4/ Appointing data stewards.

These are individuals who have expertise in both current source systems and Open Data to make the same rules apply to all data sources.

5/ Drawing up a data governance programme and establishing a data governance council.

The programme which defines how, where and with which definitions master data are established. The actual method of normalisation which will be applied to the master data. The data governance council decides in consultation which normalisation procedure will be used. See the following chapter for a pragmatic approach adopted by this council.

6/ Developing a master data model or logical data model.

Depending on the available databases and data warehouse, if any, and the required distribution of the information, it is proposed to produce a logical and physical data model to be managed under the MDM process. We recommend extending this model to also include Open Data streams.

7/ Considering a tool.

Where many data are managed, it may be recommended to use an MDM toolset. The additional data for Open Data streams can be managed in this tool, which makes a general overview possible.

8/ Designing a supporting infrastructure.

For bodies managing large numbers of data and aiming to automatically open up data to Open Data streams, it may be considered to use supporting infrastructure for implementing ETL processes. Within the Flemish public administration this type of infrastructure is already in place. We refer in this context to the technical standards described earlier in the present document.

9/ Generating and testing master data.

The master data will have to be checked for quality and consistency during manual or automatic inspections. It will be fairly impossible to make and keep all master data accurate in one go. ETL toolsets do often contain possibilities for providing this. However, in some cases specific tests may be needed (for instance to anonymise Open Data data).

10/ Implementing maintenance processes.

Processes are never static and the management of MDM and Open Data streams most certainly is not either. Therefore, you should provide a process for maintaining metadata and ETL functionality to maintain the quality of the data.

ORGANISATIONAL IMPACT AND APPROACH

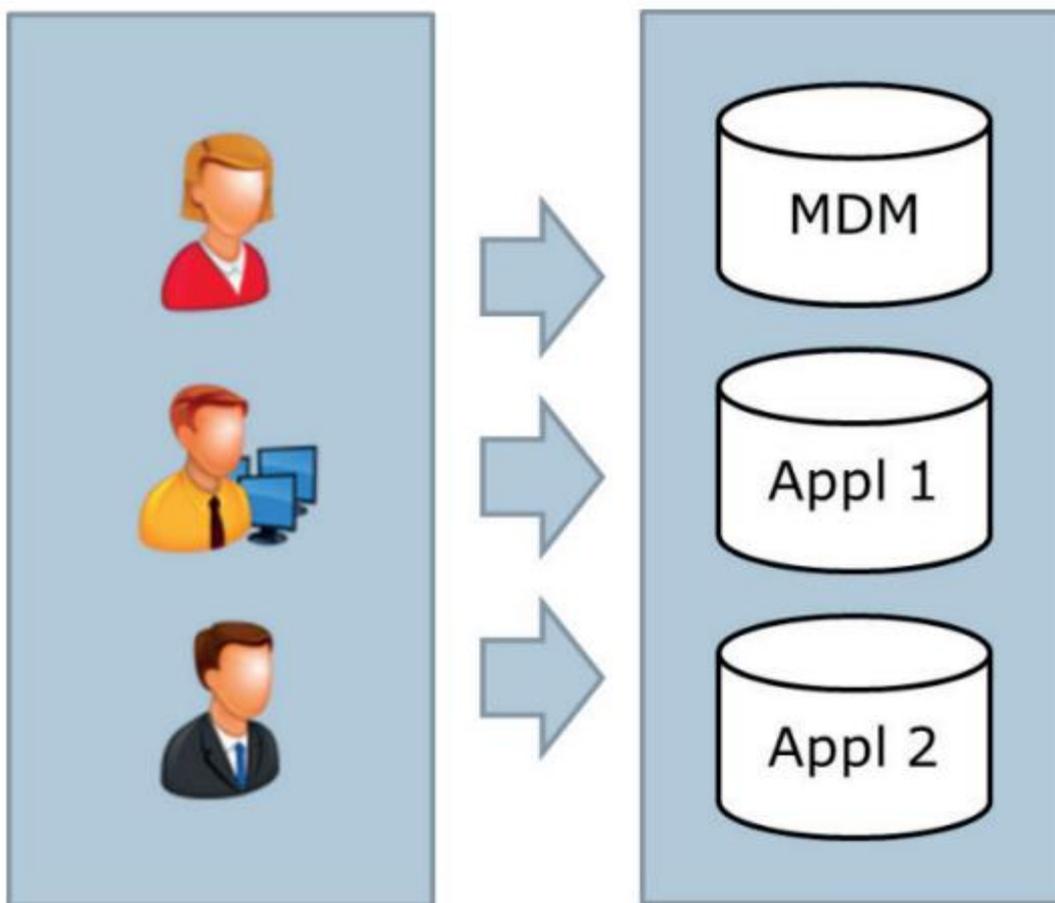
Attentive readers will understand that setting up a decent MDM process will also have an impact on the governance model of the ICT department of the public sector organisation concerned. In the first instance, we expect that the management and publication of Open Data streams will not result in additional human

resources or workload. For this reason, we propose a growth model here which has been aligned as much as possible with existing ICT contracts and can now also be used for Open Data management.

Phase 1: Short term: By public sector organisation

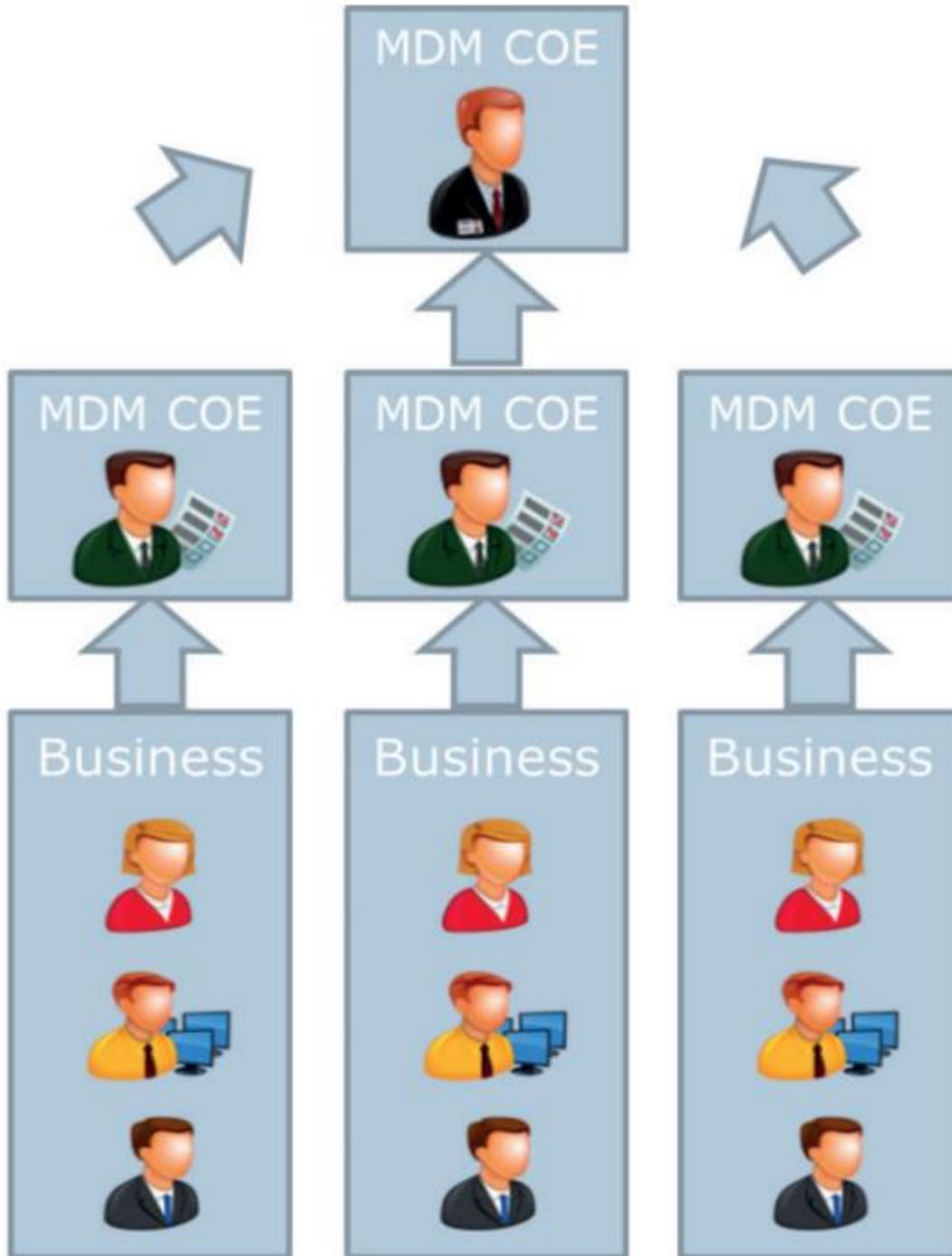
In this step we provide minimal efforts for engaging in MDM. The following characteristics apply:

- No unique or separate master data organisation or team.
- Business and/or ICT of the public sector organisation itself is in charge of drawing up standards and procedures for producing or managing master data.
- Efforts are divided between all the people within the public sector organisation and is set up as best efforts.



Phase 2: Medium-term: Federated (Open Data CoE)

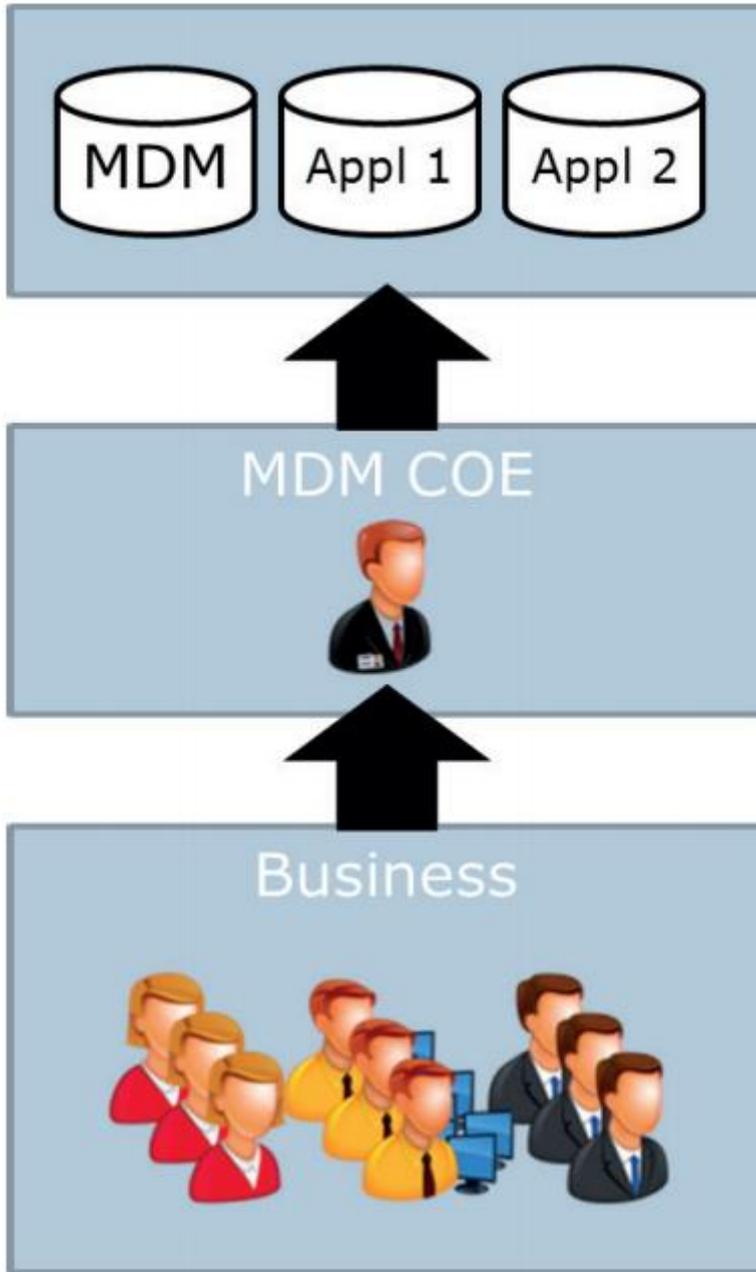
At this stage a Centre of Expertise (CoE) is established to be in charge of the public sector organisation's MDM as a whole. Standards and agreements are in place on public sector organisation level which are managed by the person who is in charge at that level. A coordinating contact person has also been appointed for the organisation's MDM. However, this person does not yet have authority to enforce these standards, but mainly has a facilitating role. The aim is to achieve the largest possible common divisor of standards and agreements.



This level of maturity thus requires people who are in charge of MDM on different levels, but also effects the standardisation.

Phase 3 (Long-term): Fully centralised

In the long term all agreements and standards regarding MDM are brought together at the highest level within the organisation. This no easy task, since it also implies the authority and responsibility for making the lower levels comply with these agreements. The advantage is that general consistency exists about all information, which facilitates the flow. Businesspeople can submit applications for changes or propose new definitions to the central team.



ANNEX 2 ADDING METADATA THROUGH THE CKAN API

CONDITION

You need an API key to make an action through the CKAN API. This key can be found on your profile page at http://ckan-001.corve.openminds.be/user/{uw_username}.



CKAN API

General information about the CKAN API is available at <http://docs.ckan.org/en/latest/api.html>.

EXAMPLE OF UPLOAD

To add a dataset through the API, you can send a JSON payload via HTTP POST.

Below, an example is given of such a JSON payload and the POST command, using CURL, a command line tool for transferring data.

Example

```
{
  "title": "Aantal aangevraagde en goedgekeurde tewerkstellingspremies 50-plus",
  "name": "aantal-aangevraagde-en-goedgekeurde-tewerkstellingspremies-50-plus",
  "url": "http://aps.vlaanderen.be/sgml/largereeksen/5867.htm",
  "notes": "Aantal aangevraagde en toegekende tewerkstellingspremies 50-plus door de Vlaamse overheid.",
  "private": false,
  "isopen": true, "license_id": "cc-zero",
  "tags": [
    {
      "vocabulary_id": null,
      "display_name": "arbeidsmarkt",
      "name": "arbeidsmarkt"
    },
    {
      "vocabulary_id": null,
      "display_name": "tewerkstelling",
      "name": "tewerkstelling"
    }
  ],
  "resources": [{
    "format": "xls",
    "name": "Aantal aangevraagde en goedgekeurde tewerkstellingspremies 50-plus in MS-Excel",
    "url": "http://www4.vlaanderen.be/dar/svr/Cijfers/Exceltabellen/arbeidsmarkt/algemeen/werkbel006.xls",
    "resource_type": "file"
  }],
  "author": "VDAB, bewerking Departement WSE",
  "maintainer": "Myriam Vanweddingen",
  "maintainer_email": "myriam.vanweddingen@dar.vlaanderen.be",
  "groups": [{"name": "arbeidsmarkt"}],
  "owner_org": "svr",
  "extras": [
    {
      "value": "arbeidsmarkt - algemeen", "key": "beleidsdomein"
    },
    {
      "value": "2006-2012",
      "key": "dimensie tijd"
    },
    {
      "value": "Vlaams gewest",
      "key": "dimensie ruimte"
    },
    {
      "value": "2012-09-07",
      "key": "laatst gewijzigd"
    }
  ]
}
```

template

```
{
  "title": "",
  "name": "",
  "url": "",
  "notes": "",
  "private": false,
  "isopen": true,
  "license_id": "",
  "tags": [
    {
      "vocabulary_id": null,
      "display_name": "",
      "name": ""
    },
    {
      "vocabulary_id": null,
      "display_name": "",
      "name": ""
    }
  ],
  "resources": [
    {
      "format": "",
      "name": "",
      "url": "",
      "resource_type": ""
    }
  ],
  "author": "",
  "maintainer": "",
  "maintainer_email": "",
  "groups": [{"name": ""}], "owner_org": "",
  "extras": [
    {
      "value": "",
      "key": "dekking in tijd"
    },
    {
      "value": "",
      "key": "geografische dekking"
    },
    {
      "value": "",
      "key": "laatst gewijzigd"
    }
  ]
}
```

Explanation

Field	Explanation
url	This is the HTTP address of the context page (the instructions for use).
license_id	One of the following values: notspecified, cc-zero, free-open-data-licence, open-data-licence-against-reasonable-charge, free-open-data-licence-for-non-commercial-re-use, free-open-data-licence-for-non-commercial-re-use + open-data-licence-against-reasonable-charge-for-commercial-re-use
Format	Lowercase file extension (csv, tsv, xml, json, rdf, xls, ...)
resource_type	file, file.upload, api
groups/name	One of the following values: labour market, public governance, foreign policy, foreign trade, culture, demography, economy, energy, finance and budget, equal opportunities, health, ICT, agriculture, media, environment and nature, mobility, monuments and landscapes, education, development cooperation, spatial planning, sport, tourism, welfare and disadvantage. science and innovation
owner_org	The list of organisations is available at http://ckan-001.corve.openminds.be/organization where you can add your own organisation.

CURL COMMAND

CURL is an open source command line tool for transferring data with URL syntax. More information is available on the CURL website <http://curl.haxx.se/>

Example of a "create" action:

```
curl -d "@payload.json" http://ckan-001.corve.openminds.be/api/3/action/package_create -H "Authorization:api-key"
```

Example of a "delete" action:

```
curl -d '{"id": "aantal-aangevraagde-en-goedgekeurde-tewerkstellingspremies-50-plus"}' http://ckan-001.corve.openminds.be/api/3/action/package_delete -H "Authorization:api-key"
```

-

ANNEX 3 METADATA DCAT-CKAN

Below you can find the DCAT profile which has been agreed between the different governments in Belgium.

Metadata is provided on the level of the dataset itself as well as on the level of distribution.

For each level, a distinction is also made between mandatory, recommended and optional fields.

For each field, the following properties are indicated: name, URI or uniform resource identifier, the value(s) which the field may contain, an explanation on how to use the field and the number of times the field may/can occur (cardinality).

As already indicated in the chapter on metadata, a number of constraints apply within the CKAN platform for making all these fields mandatory during manual entry or through upload script. Some fields can only be entered through 'free fields'. That is why we have added an additional column which indicates how this can be solved in CKAN.

Dataset

Mandatory fields for a dataset are:

Property	URI	Range	Usage note	Card	CKAN
Description	dct:description	rdfs:Literal	This property contains a free-text account of the Dataset. This property can be repeated for parallel language versions of the description.	1..n	Description (1)
Title	dct:title	rdfs:Literal	This property contains a name given to the Dataset. This property can be repeated for parallel language versions of the name. Use of Title Case.	1..n	Title (1)

(1) These fields are provided anyway - and in this case also mandatory - within CKAN.

Recommended fields for a dataset are:

Property	URI	Range	Usage note	Card	CKAN
contact point	dcat:contactPoint	v:VCard (will be deprecated in new vCard standard.)	This property contains contact information that can be used for flagging errors in the Dataset or sending comments. Initially only email address used.	0..n	Maintainer (1)
dataset distribution	dcat:distribution	dcat:Distribution	This property links the Dataset to an available Distribution.	0..n	Page « Data and page « Data and resources »
keyword/ tag	dcat:keyword	rdfs:Literal	This property contains a keyword or tag describing the Dataset.	0..n	Tags (1)
Publisher	dct:publisher	foaf:Agent	This property refers to an entity (organisation) responsible for making the Dataset available. String can be used; will be transformed into URI.	0..1	Organisation
theme/ category	dcat:theme, subproperty of dct:subject	skos:Concept	This property refers to a category of the Dataset. A Dataset may be associated with multiple themes. Existing domains used on portals can be supplied as string. Will be transformed into URI's.	0..n	Add group (1)
landing page	dcat:landingPage	foaf:Document	This property refers to a web page that provides access to the Dataset, its Distributions and/or additional information. Is the link to the leaflet 'bjsluite' page.	0..1	Page « Data and resources »

(1) These fields are provided anyway within CKAN and cannot be chosen.

Language	dct:language	dct:LinguisticSystem	This property refers to a language of the Dataset. This property can be repeated if there are multiple languages in the Dataset. <i>One of ('en', 'nl', 'fr', 'de', 'zxx'). Will be translated into URIs.</i>	0..n	Language (2)
release date	dct:issued	rdfs:Literal typed as xsd:dateTime	This property contains the date of formal issuance (e.g., publication) of the Dataset.	0..1	Date of issue (2)
update/ modification date	dct:modified	rdfs:Literal typed as xsd:date or xsd:dateTime	This property contains the most recent date on which the Dataset was changed or modified.	0..1	Last modified (2)

(2): proposal for entering name as free field in CKAN

Optional metadata for a dataset are:

Property	URI	Range	Usage note	Card	CKAN
conforms to	dct:conformsTo	dct:Standard	This property refers to an implementing rule or other specification. <i>Example: URI of OSLO.</i>	0..n	Conforms to standard (2)
Frequency	dct:accrualPeriodicity	dct:Frequency	This property refers to the frequency at which Dataset is updated. <i>Will be using the value vocabulary supplied at http://purl.org/cld/freq/ Dutch translations of the labels: * Jaarlijks * Tweejaarlijks * Tweemaandelijks * Tweewekelijks * Doorlopend * Dagelijks * Onregelmatig * Maandelijks * Driemaandelijks * Halfjaarlijks * Halfmaandelijks * Halfweeklijks * Drie keer per maand * Drie keer per week * Drie keer per jaar * Driejaarlijks * Wekelijks</i>	0..1	Update frequency (2)
Identifier	dct:identifier	rdfs:Literal	This property contains the main identifier for the Dataset, e.g. the URI or other unique identifier in the context of the Catalogue.	0..n	See URI
other identifier	adms:identifier	adms:Identifier	This property refers to a secondary identifier of the Dataset, such as MAST/ADS, DataCite, DOI, EZID or W3ID.	0..n	extra field: identifier (see VRIND identifier)
spatial/ geographic coverage	dct:spatial	dct:Location	This property refers to a geographic region that is covered by the Dataset. <i>needs further elaboration (NGI, AGIV, ...)</i>	0..n	Geographical coverage (2)
temporal coverage	dct:temporal	dct:PeriodOfTime	This property refers to a temporal period that the Dataset covers. <i>Use: schema:startDate schema:endDate.</i>	0..n	Temporal coverage (2)
Version	schema:version	rdfs:Literal	This property contains a version number or other version designation of the Dataset. <i>Use semantic versioning: MAJOR.Minor.patch</i>	0..1	Version (2)
version notes	adms:versionNotes	rdfs:Literal	This property contains a description of the differences between this version and a previous version of the Dataset.	0..1	Version Notes (2)

(2): proposal for entering name as free field in CKAN

Distribution

This is about allocating metadata to the distribution (dump or API).

Mandatory fields for a distribution are:

Property	URI	Range	Usage note	Card	CKAN
access URL	dc:accessURL	rdfs:Resource	This property contains a URL that gives access to a Distribution of the Dataset. The resource at the access URL may contain information about how to get the Dataset.	1..n	URL of distribution page

Recommended fields for a distribution are:

Property	URI	Range	Usage note	Card	CKAN
Description	dc:description	rdfs:Literal	This property contains a free-text account of the Distribution. This property can be repeated for parallel language versions of the description.	0..n	Description (1)
Format	dc:format	dc:MediaTypeOrExtent	This property refers to the file format of the Distribution. <i>Use file extension in lower case.</i>	0..1	Media type (2)
Licence	dc:licence	dc:LicenceDocument	This property refers to the licence under which the Distribution is made available. <i>For Flemish government: use one of the 4 licences.</i>	0..1	Licence (1)
byte size	dc:byteSize	rdfs:Literal typed as xsd:decimal	This property contains the size of a Distribution in bytes. <i>In bytes.</i>	0..1	Number of Bytes (2)
Title	dc:title	rdfs:Literal	This property contains a name given to the Distribution. This property can be repeated for parallel language versions of the description.	0..n	Title (1)
release date	dc:issued	rdfs:Literal typed as xsd:date or xsd:dateTime	This property contains the date of formal issuance (e.g., publication) of the Distribution.	0..1	Release date (2)
update/ modification date	dc:modified	rdfs:Literal typed as xsd:date or xsd:dateTime	This property contains the most recent date on which the Distribution was changed or modified.	0..1	Last modified (2)

(1): these fields are provided anyway within CKAN and cannot be chosen.

(2): proposal for entering name as free field in CKAN

Optional properties for distribution are:

Property	URI	Range	Usage note	Card	CKAN
Check sum	????	rdfs:Literal	This property contains the check sum	0..1	NVT
download URL	dc:downloadURL	rdfs:Resource	This property contains a URL that is direct link to a downloadable file in a given format.	0..n	URL of download
media type	dc:mediaType, subproperty of dc:format	dc:MediaTypeOrExtent	This property refers to the media type of the Distribution if this is defined in IANA. <i>Use mime type. Will be transformed into URI.</i>	0..1	Media type (2)
Rights	dc:rights	dc:RightsStatement	This property refers to a statement that specifies rights associated with the Distribution. <i>Use to make the rights of the publisher explicit.</i>	0..1	Rights (2)
Status	adms:status	skos:Concept	This property refers to the maturity of the Distribution <i>One of: 'completed', 'under development', 'deprecated', 'withdrawn'. Will be transformed into URI.</i>	0..1	Status (2)

(1): these fields are provided anyway within CKAN and cannot be chosen.

(2): proposal for entering name as free field in CKAN

ANNEX 4: OVERVIEW OF RECOMMENDATIONS

- 1/** Verify whether intellectual property rights exist in the data you wish to make available. If the public sector organisation is not the holder of these intellectual property rights, it concludes an agreement with the current right holder. Include in any future agreements or public contracts with third parties for the creation of datasets or documents a provision giving the public sector organisation the necessary rights to make the results available as Open Data.
- 2/** Check whether making the data available is not in violation of any interests protected by the Flemish Parliament Act of 26 March 2004 on Open Government.
- 3/** Check whether the data contain personal data before they are made available.
- 4/** First publish the data which are already available and may generate added value for citizens, businesses or organisations.
- 5/** Determine whether a charge will be made for re-using the data. Establish criteria for the calculation of this charge and organise the procedure for payment of the charge.
- 6/** If the distinction is to be made between charges for commercial and non-commercial re-use, give a clear definition of the concept 'commercial', which regards the use by traders for purposes of profit as 'commercial'.
- 7/** Use the model licences of the Flemish public administration for making Open Data available. Preferably choose one licence for all types of re-use, without distinguishing between commercial and non-commercial purposes.
- 8/** Choose and apply a scenario for opening up source data in view of the publication as Open Data.
- 9/** Check minimum criteria regarding quality and consistency before publishing a dataset on the Open Data Platform.
- 10/** Use open formats like CSV whenever possible.
- 11/** If you want to build an API yourself, we recommend offering a RESTful Web API and granting access without using an API key.
- 12/** Provide the technical resources that are necessary for making Open Data available.
- 13/** Make the data available through the in-house website or the website of a third-party provider and post a link to where the data can be found on the Flemish Open Data Platform (CKAN).
- 14/** Appoint a person or service to be in charge of the Open Data policy within the public sector organisation. Establish an Open Data contact point for communication within the public sector organisation, with the Flemish public administration and with citizens.
- 15/** Put a contact address or web form on the website which re-users of the data can use to ask further information or give feedback.
- 16/** Use the model licences of the Flemish public administration and add a link to them in the licence provision

to the data, using the rel="licence" attribute.

17/ If a charge is made, clearly indicate to users how (much) they must pay in order to obtain access to the data or service.

18/ If the data is made available via a service, add an information sheet or service level engagement to the service, explaining the performance of the service and the expectations users may have of the service's operation.

19/ Make an accompanying page for each of your datasets which explains in understandable language what the data are about, why they have been collected and what they are used for.

20/ Place information on the data on the website in English as well, and refer to the English version of the Flemish Open Data Licences.

21/ In order to allow for a smooth exchange of dataset descriptions, we recommend using as many fields as possible from the DCAT profile agreed at the Belgian level, even within CKAN through free fields, if necessary. See Annex 3 for an overview.

22/ Evaluate the success of your public sector organisation's Open Data practice. Share your experiences with other bodies via the Open Data Forum.

GLOSSARY

Term	Explanation
API	Application Programming Interface, specification to communicate with software components
DWH	Data Ware House
ETL	Extract, Transform, Load: process and technique from the data warehouse world to consistently collect and publish data in a data warehouse
ETP	Extract, Transform, Publish: process derived from ETL, in which the same techniques are used for collecting information and making it consistent. Only the last step (Publish) contains a number of different tools and techniques to (also) publish the dataset as Open Data.
Structured Data	Data which are labelled within a file and can therefore be easily retrieved and processed (machine-processable)
MDM	Master Data Management
Open Data Platform	"Yellow Pages" platform listing all Open Datasets, with reference to the datasets themselves. For the Flemish public administration, this platform is built on CKAN.
Open Dataset	Dataset which meets all the criteria of Open Data (format, quality, machine-readable, etc.) and which can be used by third parties
Open format	A published specification for the storage of digital data, mostly maintained by a standardisation organisation, which can therefore be used and implemented by everyone.
REST	Representational state transfer (REST), software architecture for distributed systems like the WWW
URI	Uniform Resource Identifier
VRIND	Flemish Regional Indicators
Web API	API using web technology, in many cases through the HTTP protocol.