



BOMOD

Management and Development Model for Open Data

TNO innovation
for life

BOMOD

Management and Development Model for Open Data

) **BOMOD**

MANAGEMENT AND DEVELOPMENT MODEL FOR OPEN DATA

Authors

Silja Eckartz (silja.eckartz@tno.nl)

Erwin Folmer

Reviewers

Michael van Bekkum

Sandra Struijker Boudier

April 2015

Publication of TNO

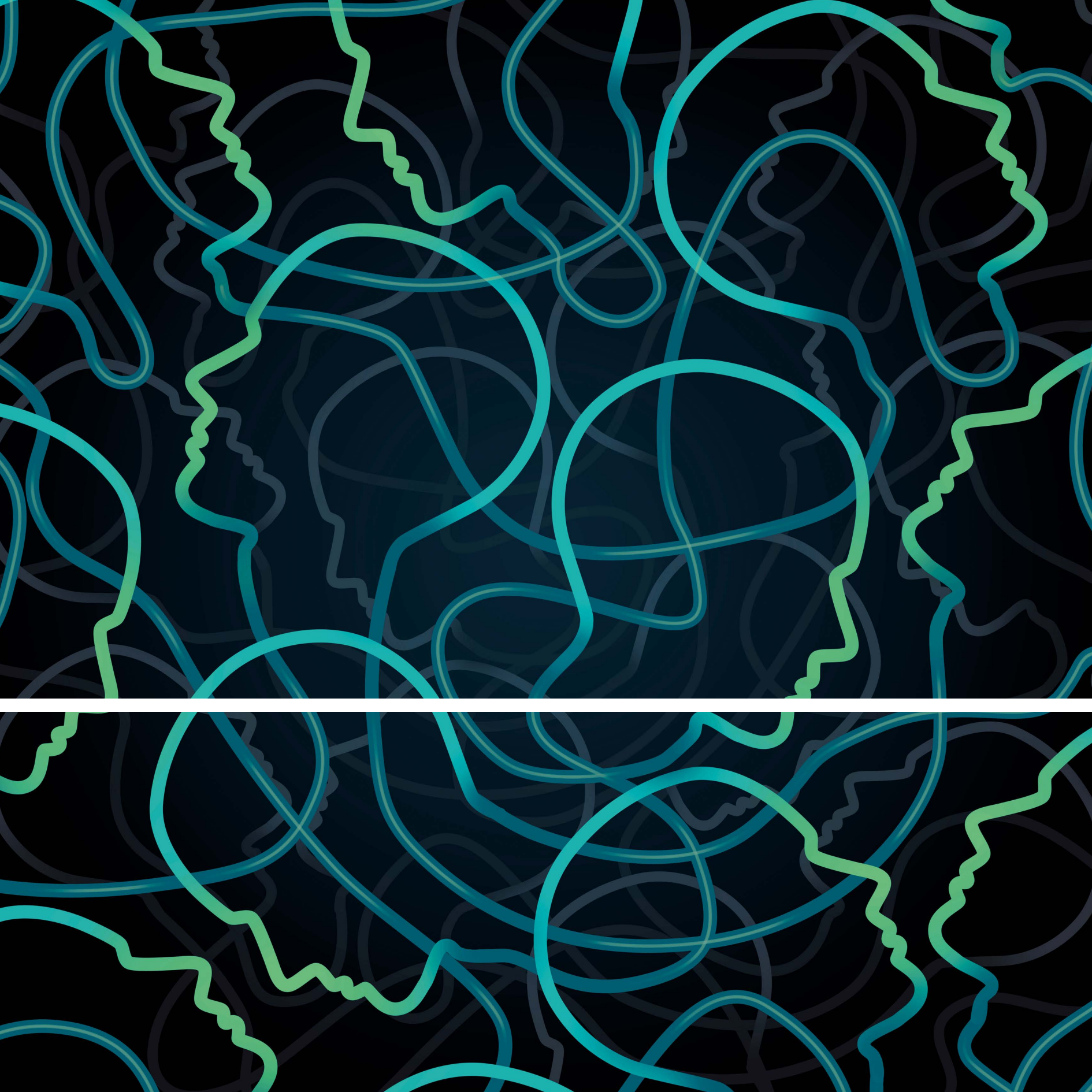
ISBN: 978-90-5986-459-7

TNO.NL

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm, website of op welke andere wijze dan ook, zonder voorafgaande toestemming van TNO.

TABLE OF CONTENT

1	INTRODUCTION	5	3	LEVEL 1 ACTIVITIES OF BOMOD	17
1.1	Barriers related to open data	5	3.1	Strategy	17
1.2	BOMOD	7	3.1.1	<i>Vision</i>	17
1.3	BOMOD and its relation to the Open Data Lifecycle model	7	3.1.2	<i>Governance</i>	20
			3.1.3	<i>Finance</i>	21
2	BOMOD	11	3.2	Tactics	22
2.1	Strategy	12	3.2.1	<i>Community</i>	22
2.1.1	<i>Vision</i>	12	3.2.2	<i>Re-Use</i>	23
2.1.2	<i>Governance</i>	12	3.2.3	<i>Data Model</i>	24
2.1.3	<i>Finance</i>	12	3.3	Operations	24
2.2	Tactics	12	3.3.1	<i>Preferences and requirements</i>	24
2.2.1	<i>Community</i>	12	3.3.2	<i>Initiation Phase</i>	25
2.2.2	<i>Rights Policy</i>	12	3.3.3	<i>Development Phase: Conceptual Level</i>	25
2.2.3	<i>Re-Use</i>	12	3.3.4	<i>Execution Phase</i>	26
2.2.4	<i>Release Management</i>	12			
2.2.5	<i>Data Model</i>	12	4	LEVEL 2 ACTIVITIES OF BOMOD	29
2.2.6	<i>Data Quality</i>	12	4.1	Tactics	30
2.3	Operations	13	4.1.1	<i>Rights Policy</i>	30
2.3.1	<i>Initiation</i>	13	4.1.2	<i>Release Management</i>	30
2.3.2	<i>Development</i>	13	4.1.3	<i>Data Quality</i>	31
2.3.3	<i>Execution</i>	13	4.2	Operations	31
2.3.4	<i>Preferences & Requirements</i>	13	4.2.1	<i>Documentation</i>	31
2.3.5	<i>Documentation</i>	13	4.3	Implementation Support	32
2.4	Implementation Support	13	4.3.1	<i>Help Desk</i>	32
2.4.1	<i>Help Desk</i>	13	4.3.2	<i>Training</i>	32
2.4.2	<i>Training</i>	13	4.4	Communication	32
2.4.3	<i>Support Tools</i>	13	4.4.1	<i>Promotion</i>	32
2.4.4	<i>Pilot</i>	13	4.4.2	<i>Publication</i>	32
2.4.5	<i>Validation & Certification</i>	13			
2.5	Communication	14	5	LEVEL 3 ACTIVITIES OF BOMOD	35
2.5.1	<i>Promotion</i>	14	5.1	Implementation Support	36
2.5.2	<i>Publication</i>	14	5.1.1	<i>Support Tools</i>	36
2.5.3	<i>Complaints Procedure</i>	14	5.1.2	<i>Pilot</i>	36
2.6	A “Maturity Model” for prioritizing BOMOD activities	15	5.1.3	<i>Validation & Certification</i>	36
			5.2	Communications	36
			5.2.1	<i>Complaints Procedure</i>	36
			6	CONCLUSION	39



1 INTRODUCTION

These days open data becomes more and more important for both private as well as public organizations. No matter if you are a private organization looking for new ways of creating value with your data, a public organization that needs to open up some of its data as open data, or if you just want to share your data for the greater good, you will very likely experience that open data is more than just publishing a data set online. Thinking about management and maintenance of the dataset is just as important as publishing the data in a format that is re-usable by others.

Research shows that many datasets published online under the label “open data” are often not re-usable at all. Links to datasets that do not work, datasets that are published in formats that cannot easily be opened, outdated data or missing data are just a few examples that limit the reusability of open data. In June 2014, only 3.3 percent of the 5700 datasets published at the Dutch open data website data.overheid.nl could actually be opened and re-used. However, during the last month they improved the quality of the datasets and removed broken links. Now they state that 92 percent of their 2500 datasets can be accessed and re-used.

This is just one example showing how important proper maintenance and management of data is to not just create random open data but data that is re-usable by others. Defining open data governance, including decisions about privacy, technical architecture, and metadata (including provenance), is crucial when publishing open data.

While public organizations are getting more and more familiar with the possibilities of open data, private organizations are often still not aware of the opportunities open data has to offer. Current literature identifies three different business models that show how private organizations can earn money with open data. These business models will be explained later on in this booklet.

1.1 BARRIERS RELATED TO OPEN DATA

When publishing data, organizations often encounter a number of barriers, including the following¹:

Barriers to publish data:

- No view on what data is more likely to be reused/ has a higher ROI potential
- Unclear business model for publishing open data
- Datasets include privacy sensitive information
- Datasets include economic sensitive information
- Poor data quality
- Data owner is not positive about open data
- Limited tool support
- Competing licenses for datasets
- Competing vocabularies for describing datasets
- Domain-specific metadata needs
- Effort required for keeping the metadata up-to-date

Another classification of open data barriers, provided by Martin et al. (2013)² as shown in Figure 1, distinguishes the following seven barrier categories: access issues, governance issues, cost issues, data issues, legal issues, metadata and skill issues.

1 <https://joinup.ec.europa.eu/sites/default/files/Open%20Data%20Support%20%40ePSIworkshop.pdf>

2 Martin, Sébastien, et al. "Risk Analysis to Overcome Barriers to Open Data." *Electronic Journal of e-Government* 11.2 (2013).

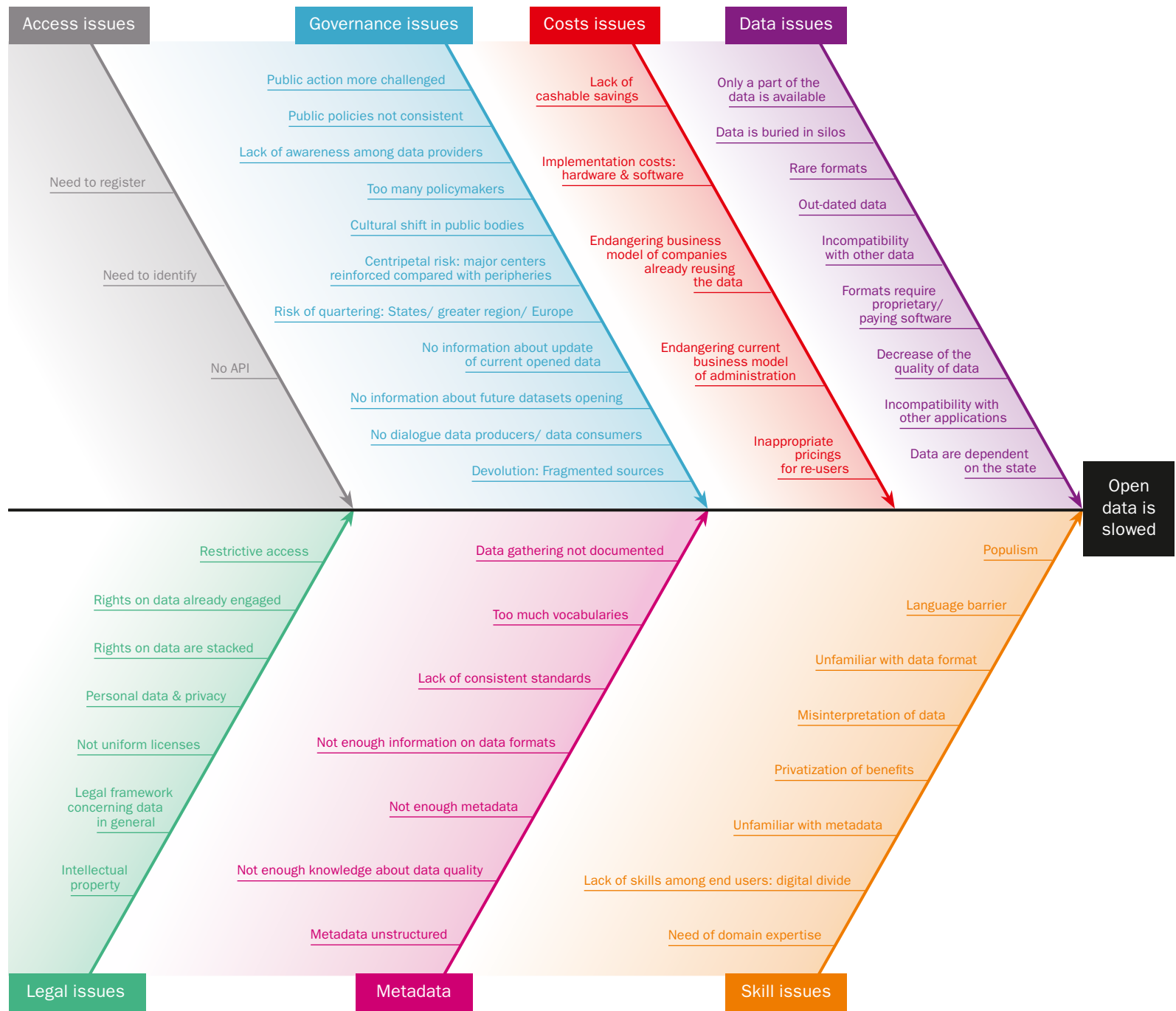


FIGURE 1 Ishikawa diagram summarising risks and barriers related to data opening (Martin et al. 2013)

Open data has a great potential to create social and economic value. However, in order to realize this value, the data has to be reused. A number of barriers specific to re-use data can be identified:

Barriers to re-use data:

- Lack of overview of existing/available datasets
- Unclear business model for reusing Open Data
- Data is often of low quality, outdated, unstructured and/or not machine-readable
- Lack of licensing information or incompatible licenses
- Different vocabularies when searching for datasets
- Lack of (good quality) metadata
- Lack of provenance information

1.2 BOMOD

This booklet will help you overcome most of the publishing barriers and some of the re-use barriers by providing you with guidelines on how to manage and govern the process of publishing data. We consider Linked Open Data as a subset of open data. We introduce BOMOD (Beheer en Ontwikkel Model Open Data) a framework with a stratified structure of activities required for the development and management of open data.

The framework is based on the structure of BOMOS as defined in BOMOS 2 and BOMOS 2i, both models to govern and develop open standards. Our goal is to make BOMOD as practical as possible, including tools and checklists where possible.

BOMOD comprises the following elements:

- Three main layers (groups of activity categories): Strategy, tactics and operations
- Two supporting layers: Implementation support and communications
- Multiple activity categories that comprise concrete activities per layer which can be carried out.

BOMOD can be applied in two scenarios: data push from the perspective of a data publisher and data pull from the perspective of a data consumer. These two different perspectives are shown in Error! Reference source not found. Figure 1.

In this booklet we focus on the instantiation of BOMOD in the data push scenario, where a data owner is looking for new ways of creating value

BOMOD: MANAGING THE (OPEN) DATA LIFECYCLE

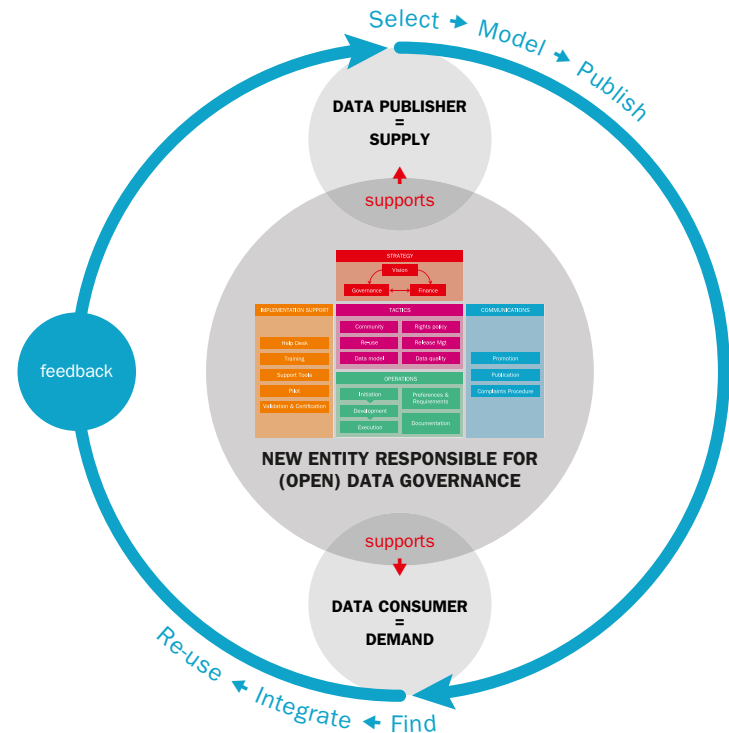
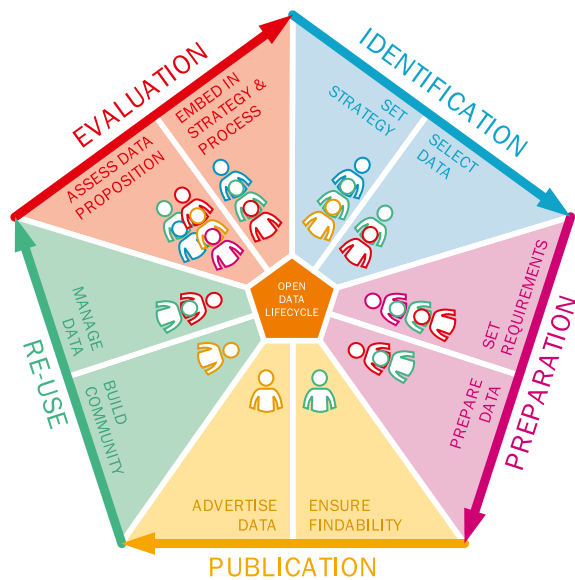


FIGURE 2 BOMOD and the Data Lifecycle



Structured approach to publish open data

Describes entire life cycle

Consists of 10 concrete steps

Describes which stakeholders should be involved in each step



FIGURE 3 Open Data Lifecycle Model

with his data through re-use by others. BOMOD is intended to support data publishers with the process of selecting, modelling and publishing data.

We introduce a maturity model showing which activities are essential and should therefore be deployed first and which activities are less crucial and can thus be deployed later. If you decide to invest in basic governance activities early on in the process of engaging in open data, you will have a solid foundation to build upon and you can avoid the experience of unexpected costs later on in the process.

The BOMOD activities described in this booklet can be performed by a newly defined data governance entity within or outside of your organization or by an existing entity within your organization responsible for data management. More information on the differences between these governance forms can be found in the description of the “Strategy” layer.

Besides BOMOD, there is also the publicized LOD Roadmap. Whereas BOMOD focuses on management aspects, the LOD roadmap provides the essential operational steps needed to publish the actual data as linked open data. BOMOD and the LOD roadmap combined are the perfect starting point for every organisation interested in publishing open data.

1.3 BOMOD AND ITS RELATION TO THE OPEN DATA LIFECYCLE MODEL

Earlier work by Van den Broek et al (2011)³ has resulted in the Open Data Lifecycle model shown in Figure 3, providing a structured approach to publish open data. BOMOD covers all phases from this life cycle model but digs deeper and provides practical guidance on how to arrange data governance in each of these phases.

3 Van den Broek, T., A. F. van Veenstra, and E. Folmer. "Walking the extra byte: A lifecycle model for linked open data." *Linked Open Data–Pilot Linked Open Data Nederland* (2011): 95-111.





2. BOMOD

After having explained how BOMOD can be used in the overall data lifecycle (Figure 2) we will now zoom in on the 22 BOMOD activities that support data owners who want to publish their data. BOMOD activities help data owners with decisions on data governance strategy, tactics, operations, implementation support and communications. Figure 4 shows the activity diagram of BOMOD. We will first provide a short overview of all activities, structured according to their higher level process. Thereafter, we will introduce a maturity model and provide detailed guidance on how to implement the activities at each of the maturity levels.

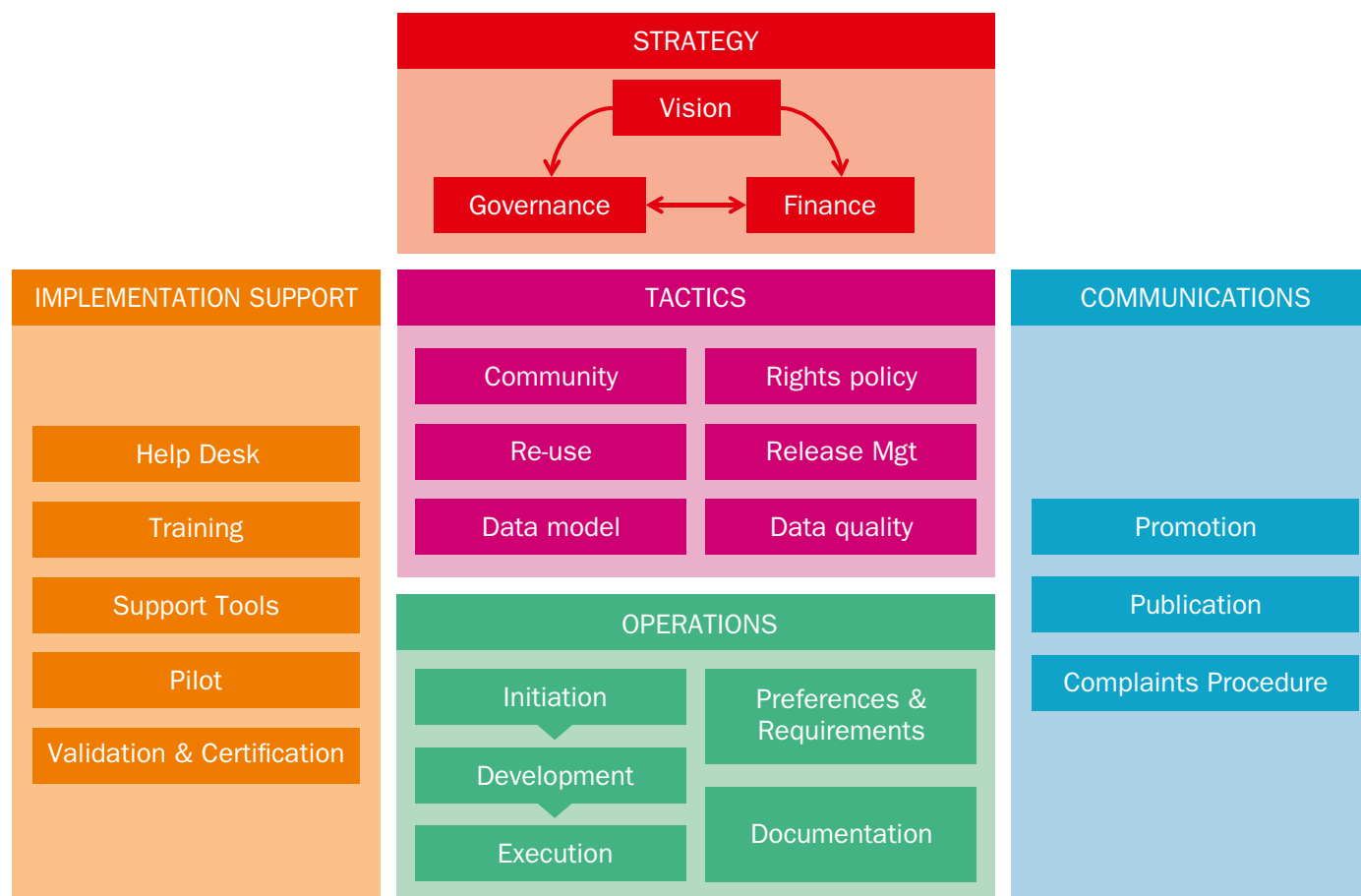
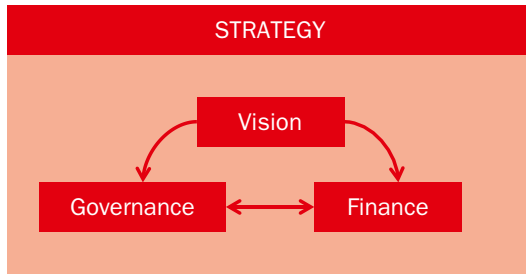


FIGURE 4 BOMOD – Activity Diagram

2.1 STRATEGY



This layer describes directing activities related to the strategic (long) term focus of the data governance entity including vision, governance structure and financial structure for publishing open data. We advise to start with defining the vision for your organization with respect to open data, as it influences the choice for a governance structure and financial structure.

2.1.1 VISION

Describes all activities related to defining what you want to do with open data now and in the future. This includes the definition of a business model for publishing open data.

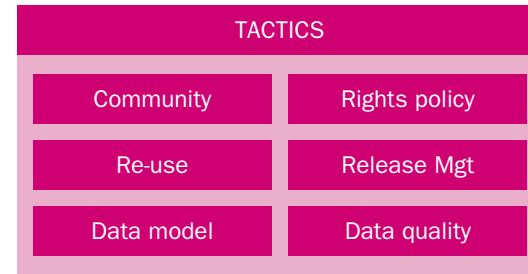
2.1.2 GOVERNANCE

Describes the governance structure of the organization that publishes open data including decision making activities.

2.1.3 FINANCE

Describes all activities related to financing open data initiatives, including setting up a sound business case.

2.2 TACTICS



The tactics layer describes six activities that steer the process of publishing data on a tactical level, namely: determine a strategy for community building, defining a rights policy, defining a re-use strategy, determine how the release of data will be managed, define a data model and define a data quality policy.

2.2.1 COMMUNITY

Describes all activities needed for community building such as ensuring that the right stakeholders participate in an open data community.

2.2.2 RIGHTS POLICY

Describes all activities related to defining the rights and obligations with respect to the data, such as defining who can be held responsible for the quality and content of the datasets.

2.2.3 RE-USE

Describes all activities, such as the definition of a re-use strategy, needed to ensure that the market re-uses the open data and that value is realized from this re-use.

2.2.4 RELEASE MANAGEMENT

Describes all activities related to managing the data lifecycle and the release of new datasets and updates.

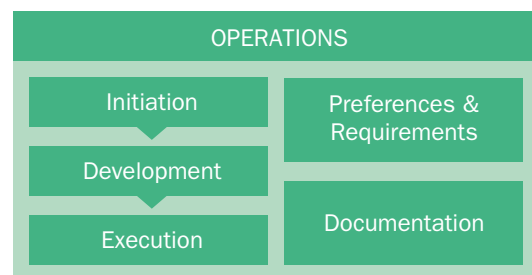
2.2.5 DATA MODEL

Describes the explicit structure of the data, including relationships and properties of objects covered by the dataset.

2.2.6 DATA QUALITY

Describes all activities related to gathering quality requirements and ensuring that the dataset meets these requirements.

2.3 OPERATIONS



The operational layer describes executive activities that lead to the actual publication of new datasets. This includes activities related to the following three operational phases: initiation, development and execution as well as activities related to defining preferences and requirements and documentation activities.

2.3.1 INITIATION

Describes all activities related to the start of a new data initiative. This can be triggered by a new idea to publish data.

2.3.2 DEVELOPMENT

Describes all activities related to the actual development of open datasets or new conceptual links between them.

2.3.3 EXECUTION

Describes all activities related to the actual publication of an open dataset.

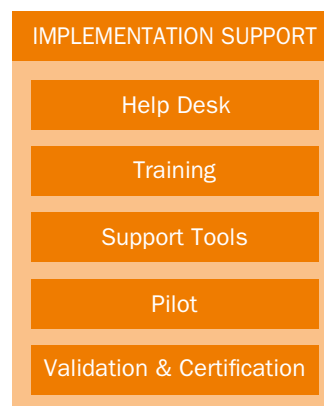
2.3.4 PREFERENCES & REQUIREMENTS

Describes all activities related to identifying preferences and requirements for open data from data users and data owners.

2.3.5 DOCUMENTATION

Describes all activities related to the documentation of the open dataset.

2.4 IMPLEMENTATION SUPPORT



This layer describes activities focusing on supporting the publication of datasets namely the support of pilot implementations and the provision of a help desk, training, support tools and validation activities.

2.4.1 HELP DESK

Describes all activities related to offering support to the various user groups.

2.4.2 TRAINING

Describes all activities related to offering training opportunities to the various stakeholders.

2.4.3 SUPPORT TOOLS

Describes all activities related to offering modules and tools that support publishing open data.

2.4.4 PILOT

Describes all activities related to pilot implementations of new applications making use of open data.

2.4.5 VALIDATION & CERTIFICATION

Describes all activities related to validating and certifying open datasets, e.g. 5 star open data.

2.5 COMMUNICATION



The communication layer describes supporting activities that focus on creating support for publishing open data, such as the definition of a promotion and publication strategy as well as the definition of a complaints procedure.

2.5.1 PROMOTION

Describes all activities related to promoting data publishing.

2.5.2 PUBLICATION

Describes all activities related to making datasets findable on the Internet.

2.5.3 COMPLAINTS PROCEDURE

Describes all activities related to ensuring that complaints are taken serious.

2.6 A “MATURITY MODEL” FOR PRIORITIZING BOMOD ACTIVITIES

The BOMOD framework provides an overview of activities that support organizations with the management of (open) data. An organization who wants to publish a single dataset should however not necessarily deploy all activities described in BOMOD. Depending on the size of the organization, the dataset and the goal of the organization with regard to its (open) data, some activities are more important than others. This difference in importance can be shown in a “maturity model” with three levels: basic/essential activities, extended activities and optimized activities.

Figure 5 shows the maturity model of activities data owners need to deploy in order to publish their data and stimulate data re-use. Most operational, tactical and strategic activities are defined as being essential during the process of publishing data. Three of the tactical activities are defined as extended activities for organizations that want to reach maturity Level 2 with respect to the process of publishing data. The activities needed to reach maturity Level 3 all fall into the supporting layer categories: communication and implementation support.

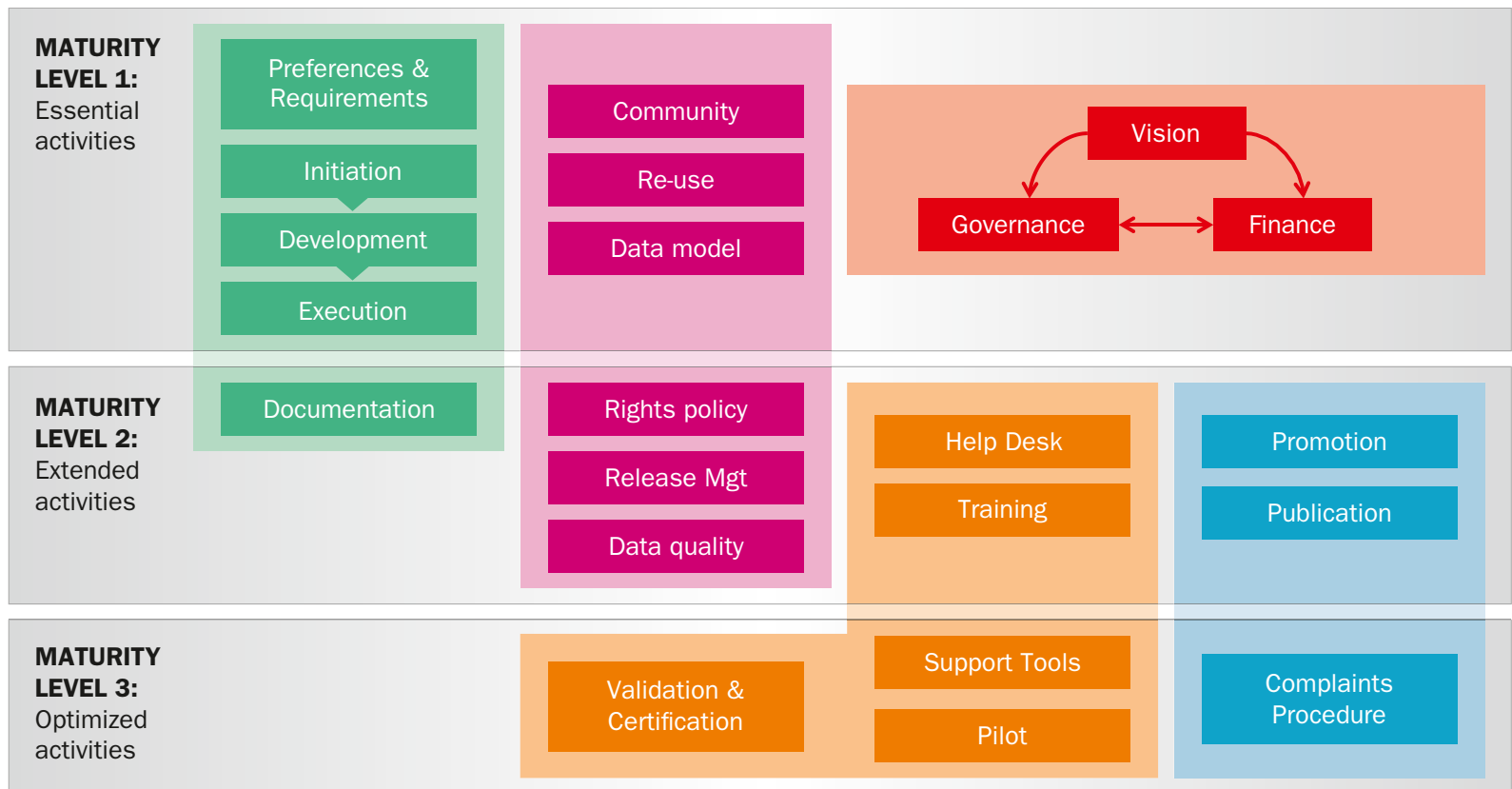


FIGURE 5 Maturity Model Data Push

(green = operations, violet = tactics, red = strategy, orange = implementation support, light blue = communication)



3. LEVEL 1 ACTIVITIES OF BOMOD

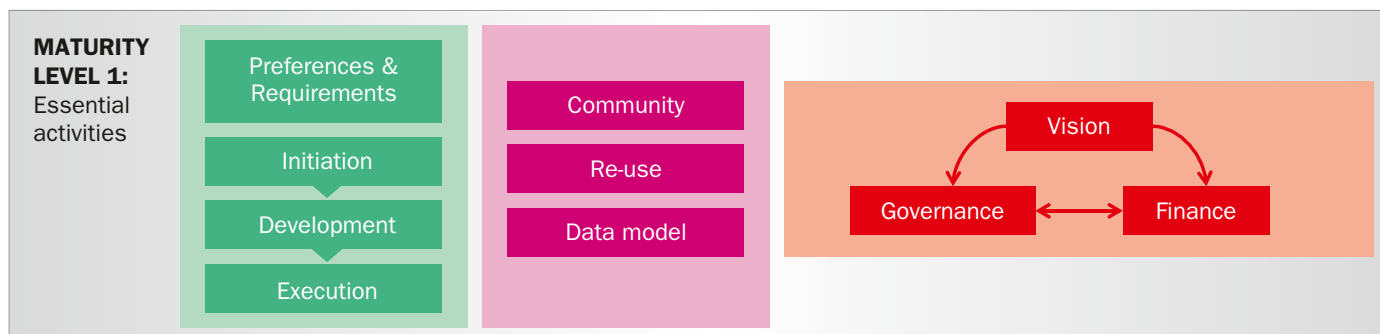


FIGURE 6: Level 1 Activities of BOMOD (green = operations, violet = tactics, red = strategy)

The activities shown in the figure above are considered to be essential activities that organizations who want to open data should start with when defining their governance. We will now discuss each of these activities in more detail, including practical tools and guidelines.

3.1 STRATEGY

3.1.1. VISION

Before you engage in any open data initiative or publish open data yourself, you need to have a clear picture of what you want to do with (open) data, now and in the future. What is the goal you want to achieve by having open data? Most data owners are driven by one of the following incentives: economic gains, social benefits, operational and technical benefits and/or legal compliance. Examples of these incentives are shown in the box on page 18. Economic gains can only be realized when others actually find and re-use the data you have opened. We will therefore pay special attention to re-use in the tactics section of this booklet. The following actions and questions can help you clarify the role data currently has in your organization and come to a clear vision on what you want to achieve by opening data.

- Recap your organizational goals. Start your open data initiative by reviewing what your organizational goals are.
- Analyse how data is used right now and how it could be used in the future.
 - Which datasets do you own?
 - Where do you collect data?
 - With whom are you sharing data at the moment?
 - Who in your value network might be benefit if you open data?
 - Are you the only organization in the value network opening data or is one of your partners also opening data that you are benefiting from?
- Clarify for yourself and important stakeholders in your network why you want to publish open data. Analyse if open data can help you to reach your organizational goals. See box on page 18 with incentives for inspiration.

Economic incentives:

- Offer new services resulting in more sales
- Efficiency gains; cost reductions
- Creating a competitive advantage or be a perceived as a leader
- Improved image resulting in more sales
- Enlarge customer base
- Collaboration to be able to make the market as a whole grow
- Fear of losing a reputation or losing out on potential gains
- Reciprocity: Open data in exchange for data from another party

Social benefits:

- Transparency, ensuring trust and accountability
- Improvement of citizen services
- Stimulation of knowledge developments

Operational and technical benefits:

- Creation of new data based on combining data
- Access to external problem solving capacity
- Easier access to data and discovery of data
- External quality checks of data (validation)

Legal incentives:

- Compliance to regulations, for example environmental, data protection, or safety regulations
- Compliance to directives regarding open data

Open Data Business Models:

- Freemium: Raw data is available for free, but you charge for extra services, e.g. API's, data being always available (SLA), no rate limits, higher quality data, data being offered in different format.
- Cross-subsidy: Data is available for free, but you get extra benefits from your data as you reach more potential customers and increase your reputation.
- Network effects: You collaborate in a rich data environment where your data is freely available for other participants. This will extend your audience and reduces your maintenance costs of data as the community jointly improves the quality of the datasets, e.g. by providing complementary data to complete incomplete datasets.

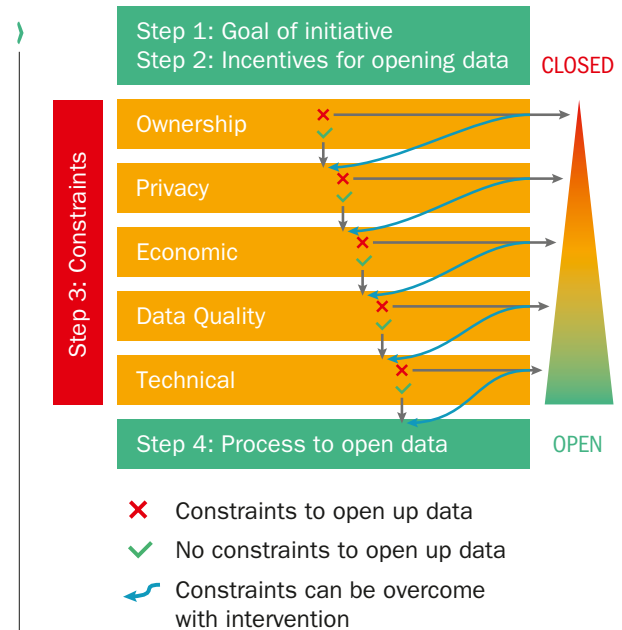


FIGURE 7: Open Data Decision Tree

- Determine which (kinds of) datasets are suitable for publication as open data. The Open data decision tree⁴ (Figure 7) can be used as a tool for identifying potential barriers and deciding if a dataset can be published.
- Determine which open data business model suits your organization and vision. See the box at the right for an overview of the three most common business models. Be honest about the reason why you publish open data.
- Develop a business case to demonstrate the benefits (incl. long term and indirect benefits) of publishing open data on an individual as well as community level.
- Determine how you will publish your datasets: raw data vs quality improved data.

The open data decision tree is structured into four steps: The identification of the goal of data sharing (Step 1), the identification of incentives for individual stakeholders (Step 2), the identification of barriers, also referred to as constraints, to data sharing (Step 3) and the definition of the process to publish data (Step 4).

Step 3 is the most extensive and the focus of the open data decision tree shown in Figure 7. Based on literature and experience we identified a hierarchy of five categories that can be used to structure barriers to open data: ownership, privacy, economic, data quality, and technical barriers. We prepared a detailed questionnaire for each of these categories to identify if a dataset can be shared or if constraints hinder data sharing. Once a constraint is identified, the next step is to determine if the barrier can be overcome with an intervention. A list of technical as well as organizational interventions is shown in the box at the right. The selected interventions are included in Step 4 when the dataset is actually published. The process to open data (Step 4) consists of step-by-step guidelines covering technical interventions and aspects, such as data conversion, metadata requirements and URI strategies as well as organizational interventions and aspects around governance. These will further be discussed in the “operations” section of this booklet.

Technical Intervention:

- Access control
- Aggregation or anonymization of data
- Improve data quality
- Publish metadata
- Support data standards

Organizational Intervention:

- Trust mechanisms
- Successful pilot projects
- Sharing of best practices
- Cultural change programs
- Collaboration mechanisms

⁴ Silja M Eckartz, Wout J Hofman, Anne Fleur Van Veenstra, A Decision Model for Data Sharing, Electronic Government, Springer (pp.253-264),2014.

3.1.2 GOVERNANCE

Your vision on open data defined earlier will influence the ideal governance structure for your organization. If (open) data is strategically important for your business processes, you might consider giving special attention to the publication of (open) data by setting up a data governance entity within your organization that is responsible for managing and supporting the value creation through data. In any case you need to define who is responsible for making decisions about open data in your organization. This responsibility includes deciding on which datasets are opened, how they are opened, with whom they are shared, what your data business models look like, how datasets are maintained and updated and what is done with user feedback and requirements. All of these activities are described in BOMOD (see the remainder of this booklet). However, you also need to decide who is actually responsible for making decisions about these activities. All these decision responsibilities are described by the governance structure.

We advise organizations to start with a working group that focuses on (open) data management and that is internal to the organization. Such a working group can have a general focus on data management, e.g. defining an (open) data strategy. Later on, once the data strategy and practices of an organization are more mature, it might be useful to set-up a working group external to the organization including potential dataset users to get feedback and control the re-use of datasets. The box at the right shows for each maturity level how the governance structure can be defined.

The following activities can help data to define their governance structure.

- Define the governance structure with respect to data management. This includes the definition of roles, such as data owner, data steward, data producer, data publisher and data user as well as the definition of working groups (first internal, then external).
- Define the decision making arrangements and processes: Who has the power/is responsible to make which decisions? For example who decides about opening a dataset. Decide who is actually responsible for making decisions about the deployment of the BOMOD activities described in this booklet.
- Define how decisions are made. Which mechanism is used? E.g. Consensus, majority decides or use of (multiple rounds of) voting?
- Define the time frame in which decisions are made.
- Define who can participate in which working group, and how internal and external stakeholder can get involved in a working group.

EXAMPLES OF HOW THE GOVERNANCE STRUCTURE CAN BE DEFINED AT THE THREE MATURITY LEVELS:

LEVEL 1: Define a contact point for data users responsible for handling complaints, feedback and comments. Define one person responsible for decision making.

LEVEL 2: Define a more structured governance process, including working groups and more diverse roles.

LEVEL 3: “Heavy weight governance scenario”

Define an extensive governance structure:

- Advisory board (incl. open data experts and potential users) – provides non-binding strategic advice to the management
- Management (Data owner organization) – decides which datasets to open, in which external data WGs to participate etc.
- Executive organization – coordinates WGs, decides on smaller changes
- Working Group – A WG can be organized intern and extern to your organization or around a certain topic or dataset. Make sure WGs have an independent chair.
 - Intern to organization: focusing on opening datasets (e.g. prepare decision to open data, prepare the actual publishing of data) or focusing on finding other datasets to use
 - Extern of the organization, including external users e.g. for BAG, to manage and support the process of data reuse by others

An example of how the governance structure of an open data initiative can look like can be found at: <http://eoverheid.nl/onderwerpen/stelselinformatiepunt/besturing>

3.1.3 FINANCE

In order to ensure continuity and realize value from your data, it is important that you think of a sustainable financial model and calculate a sound business case. Your vision on open data defined earlier and the business model chosen will influence your financial model for publishing open data. The following activities help you to define a financial structure for your open data initiative.

Sound business case:

- Show that the expected (long term) benefits of publishing or re-using open data are higher than the expected costs. While defining a sound business case you always need to try to find a balance between the features of your dataset and the costs that come along with them. You might decide to start with publishing 4 star data which is less costly and only after some time “upgrade” your data to 5 or 6 star data, which then results in higher costs.
- Benefits: Determine direct and indirect benefits of publishing datasets (both short-term as well as long-term benefits).
- Costs: Determine the costs for publishing datasets. The costs are highly dependent on the decisions you made about the format the data is published in, if it is raw data or enhanced data and if the data is real-time available.

Funding:

- Decide how the process of publishing datasets will be funded. Analyse the possibility of getting governmental open data subsidies.
- Determine the business model (if not already done in the vision activity): Freemium, cross-subsidy or network effects.
- Based on your vision, decide how the pricing structure in a business model might change over time. For example, extra services are offered for free at first, but are charged for after one year. It is important that there are as few barriers as possible for potential data users to start using the data.
- Describe a structural funding model that provides realistic guarantees that indicate that there will be an annual budget which is structurally adequate to at least keep the data opened.

EXEMPLARY BENEFITS OF OPEN DATA:

- Increase transparency
- New insights
- Knowledge sharing
- Address societal challenges
- Improved decision making
- Enable data driven innovation
- Stimulate open innovation
- Create benchmarks
- Improved services
- Empower citizens
- Increased data quality through feedback

EXEMPLARY COSTS OF OPEN DATA:

- Operations – keeping datasets open
- Maintenance of datasets
- Staff
- IT systems
- Implementation of interventions (safeguard sensitive information)

FACTORS INFLUENCING COSTS AND BENEFITS:

- Type of open data that will be published (raw vs. improved data)
- License under which dataset is opened, e.g. open license
- Size of the dataset
- Real time service or one time download

3.2 TACTICS

3.2.1 COMMUNITY

Publishing open data should not happen in isolation but within a community that includes a diverse set of stakeholders such as data owners, data producers, data stewards, data users, business users, app developers and Linked Open Data experts. These roles can be fulfilled by people within or outside of your organization. Community building and ensuring that the right stakeholders take part in an open data community is essential for publishing data. Activities that need to be carried out in order to build a community differ per stakeholder group. Examples of stakeholders and their motivation are provided in Table 1, as well as examples of methods that can be used to build a community with these stakeholders. For all stakeholder groups it holds that face-to-face events can be a very effective way to encourage others to use your data.

TABLE 1:

Examples of stakeholders and their motivation to participate in a community

STAKEHOLDERS	MOTIVATION TO PARTICIPATE IN OPEN DATA COMMUNITY	METHOD TO BUILD COMMUNITY
End users	Extra value	User events, online user groups (e.g. Linked-In)
Business user	Extra value and new applications	User events, online user groups
App developer	“quick & dirty” results	Hackatons (see box at the right)
LOD expert	More beautiful LOD according to standards	Linked-In Group

Which stakeholders need to be involved when building a community can be determined using a stakeholder analysis (see Figure 8 by Bryson⁵) and value network analysis. The following actions can be used to structure the community building process:

5 Bryson, J. (1995) Strategic Planning for Public and Nonprofit Organizations, San Francisco, CA: Jossey- Bass.

HACKATHONS

So-called “hackathons” are specialized conferences that encourage the active participation of attendees to collaborate at the event on software projects. Hackathons create a temporary boost and have successfully been used to generate interest in open data initiatives among public stakeholders. They encourage creativity and help to catalyze innovation.

However, there is no guarantee that something will automatically happen with resulting ideas once the hackathon is over,. The organizers of the hackaton should think about the continuation of the innovation and ensure that the dataset can still be used once the hackaton is done.

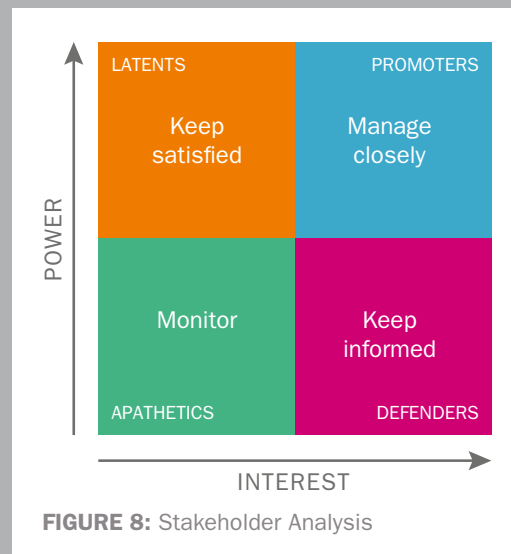


FIGURE 8: Stakeholder Analysis

VALUE NETWORK ANALYSIS:

Value network analysis can be used as a tool to unravel the many interdependencies between and evaluate the effect of open data on the network.

- Analyse your current value network using stakeholder analysis or value network analysis.
 - With whom are you exchanging data?
 - Who are potential users of your data?
 - What are the roles of important actors?
- Learn about the motivation of different stakeholder groups.
- Determine a strategy to build a community with the right stakeholders, make sure that both data owners as well as (potential) data users are involved.
- Make sure your community building strategy is in line with the motivation of the most important stakeholders.
- Determine the governance structure of your initiative: who should be involved in WGs, user groups or the advisory board.
- Build the community with the selected stakeholders. A minimum requirement is the involvement of at least two potential users of your data.
- Monitor and promote a balanced composition of the community.

3.2.2 RE-USE

In order to increase the potential value of data re-use and realize the economic gains specified in the vision activity of this booklet, it is very important that datasets can be found by potential users. Data repositories and data registries can help to increase the findability of datasets. Data repositories, such as CKAN can be used to host datasets. However, there is no such thing as a single, universal repository. There are many data storage platforms that exist, which means that just hosting the data on a random data repository does not ensure that the dataset is found by potential users. Data registries can help to increase findability by listing the dataset in a registry using tags. The situation with data registries is however similar to that of data repositories: there is also not one single registry, but again many registries, which makes the situation even more complex.

A re-use strategy might help organisations to specify how and where they want to host and register their datasets in order to ensure findability. The following steps are important when setting up a re-use strategy that makes sure that published open data is found and re-used by other parties.

- Determine where to publish datasets (e.g. data.overheid.nl or CKAN, see box at the right).
 - Determine which data registries to use in order to make datasets findable (e.g. Linked Open Data Cloud, see box at the right).
 - Define requirements for metadata about the dataset, including a description of the content of the dataset by means of tags and a tag that labels the dataset as a dataset.
- Decide if everyone can re-use the data or if the user first needs to be certified by your organization (see implementation support).
 - Decide if the dataset can be re-used for any purpose or only for previously agreed upon purposes.

REPOSITORY EXAMPLE:

Ckan.net (datahub.io) is a free tool and data management platform from the Open Knowledge Foundation that currently hosts more than 9000 datasets. Datasets can be made open to the public such as the Linked Open Data Cloud or be made available to a specified group only. CKAN requires a minimum set of information about the dataset (metadata) to be added when uploading the dataset. These are CKAN name (a unique id), title, URL, number of triples and links to other datasets. The data platform encourages data owners to add additional information to the database, such as: g. SPARQL endpoint, void description, license, and the topic of the dataset.

REGISTRY EXAMPLE:

One example of a data registry is the Linked Open Data Cloud (<http://datahub.io/group/lodcloud>) a group that catalogs datasets that are available on the Web as Linked Data and contains data links pointing at other Linked Datasets. Currently more than 30 datasets are published.

3.2.3 DATA MODEL

Modelling data is often a very time consuming task, but it helps to make the data more widely understandable and usable both within and across organizations. On a tactical level it is important to make decisions about linking the dataset to other existing vocabularies. The following activities help you make decisions about your data model.

- Make a conceptual model of the data by defining concepts and their relationships and properties.
- Investigate how others are describing similar or related data in vocabularies.
- Decide which existing, standardized and widely adopted vocabularies you want to re-use. Since others use the same vocabularies, your dataset will be linked to the dataset of others with the vocabulary as bridge. This is very important to increase the usability of the dataset. An extensive list of standard vocabularies can be found in our LOD roadmap. A few examples are given in the box at the right.
- Make an overview, in which different datasets of your organization are positioned in their context. This overview helps potential users in deciding to re-use your dataset.

3.3 OPERATIONS

The activities described under operations are both applicable to publishing new data as open data as well as changing already published datasets, e.g. updating, editing or adding data. Within the activities we will refer to the nine steps to publish open data described in the LOD roadmap developed by TNO⁶.

3.3.1 PREFERENCES AND REQUIREMENTS

In order to identify what the dataset should look like for (potential) users to be able to realize value with it, user preferences and requirements need to be carefully collected and managed. The following questions and activities can help you with identifying the preferences and requirements of your (potential) data users.

- Which datasets are interesting for potential users in- and outside of your network?
- Besides the data elements already published, which other data elements would be interesting for users?

⁶ The LOD roadmap is accessible via: www.platformlinkeddata.nl/wiki/BoekTNO/stappenplan

EXAMPLES OF STANDARD VOCABULARIES:

- Friend-of-a-Friend (foaf:) describing people:
<http://xmlns.com/foaf/spec/>
- WGS84 (geo:) describing geo positioning:
<http://www.w3.org/2003/01/geo/>
- Dublin Core Metadata Initiative (dct:) describing documentation:
<http://dublincore.org/documents/dcmi-terms/>
- Creative Commons Rights Expression Language (cc:) describing licenses:
<http://creativecommons.org/ns>
- SKOS: provides a standard way to represent knowledge organization systems using RDF

- Use Linked-In or other online discussion forums to generate creative ideas and collect feedback from your users.
- Manage maintenance requests for the datasets you already opened.
- Describe how change requests can be submitted, and what the remainder of the procedure looks like.

3.3.2 INITIATION PHASE

During the initiation phase the preferences and requirements collected earlier, are being processed and evaluated.

- Evaluate ideas for new datasets, updates of existing open datasets, new applications areas, and all activities that are required to setting them up successfully (e.g. analysis of interests, business case, planning).
- Use physical meetings, such as hackatons to generate creative ideas for new data applications.
- Select the data that you want to modify or publish.

3.3.3 DEVELOPMENT PHASE: CONCEPTUAL LEVEL

During this phase the data selected in the initiation phase is prepared for being modified and/ or published using the following activities.

- Develop new conceptual links between datasets to be published as linked open data using the conceptual model developed before (Tactics – data model) as input. Step 3 of the LOD roadmap provides detailed guidelines on linking data. We will provide a short summary of different types of data and ontology links in the box at the right.
- Re-use those existing vocabularies identified in the data model section before (Tactics). Develop new vocabularies where necessary.
- Select an URI strategy for your dataset. In Step 4 of our LOD roadmap we provide detailed guidelines on how to define a URI strategy. The most important guidelines as summarized in the box at the right.
- Develop or modify the metadata of your dataset. See Step 7 of the LOD roadmap for a detailed description of essential metadata elements.
- Decide how you want to publish the data. Do you want to publish the data as a web service that can be queried? Following, you need to make sure that your data is in the right format, e.g. by converting your dataset to RDF (read Step 5 of our LOD roadmap for technical advice on this conversion).

For a detailed step-by-step guideline on how to publish LOD please refer to our LOD roadmap, published at the platform linked data blog:
<http://www.platformlinkeddata.nl/wiki/BoekTNO/stappenplan>

DATA LINKS:

- Assign data elements to objects in order to link datasets to each other

ONTOLOGY LINKS:

- Re-use existing standardized vocabularies to make ontology links (check the data model activity above for an overview of standardized vocabularies)
- Create your own new vocabulary to link datasets

URI STRATEGY:

In our LOD roadmap we propose the following structure: `http://{domain}/{type}/{dataset|ontology}/{concept}/{reference}`

W3C describes the following as best-practice:

- Use URIs as names for things.
- Use http-URIs, so that people can look up those names.
- When someone looks up a URI, provide useful information, using standards (RDF, SPARQL)
- Include links to other URIs, so that they can discover more things.

3.3.4 EXECUTION PHASE

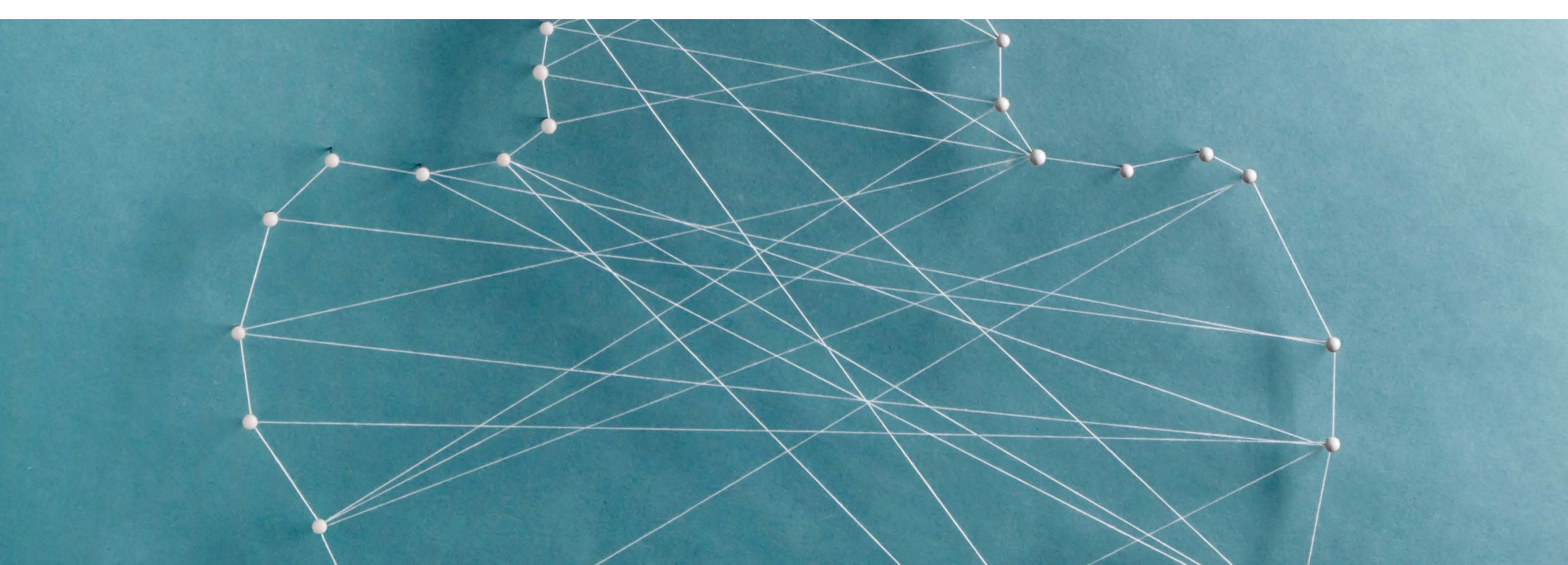
During the execution phase the dataset is actually published or modified, based on the conceptual development and according to the preferences and requirements specified earlier in the process. There are different options for publishing the dataset. A good practice is to make use of several options, so that data users have a choice and can select the method that best suits their purposes. The following points describe the activities that you need to performed to actually publish your dataset.

- If needed and decided in the strategy: Improve the data quality of datasets before publishing them (see Step 2 of LOD roadmap). A summary of important quality aspects is provided in the box at the right.
- Publish the metadata of your dataset (see Step 7 of LOD roadmap for more details on metadata).
- Publish your dataset, e.g. as a flat file. Often used syntaxes are: RDF/XML (.rdf) and Turtle (.ttl). Another, more advanced, way to make the data available is to store it in a triple store and serve it through a SPARQL-endpoint. If you provide a SPARQL Endpoint you allow others to query your linked data/ metadata. More information on this can be found in our LOD roadmap (Step 8).
- Manage version management/ updates of datasets based on tactics described in release management.

› DATA QUALITY ASPECTS:

- Validity, completeness, consistency, uniqueness, timeliness, accuracy, preciseness





4 LEVEL 2 ACTIVITIES OF BOMOD

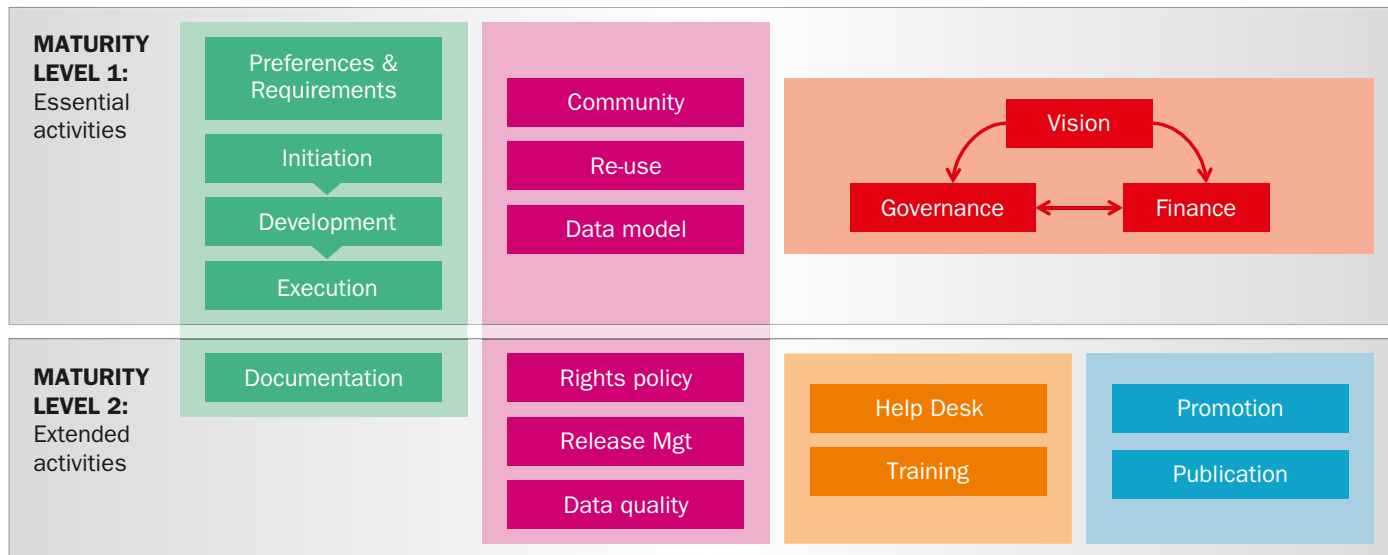


FIGURE 9: Level 2 Activities of BOMOD

(green = operations, violet = tactics, red = strategy, orange = implementation support, light blue = communication)

In addition to the activities described at Level 1, the following activities should be performed at Level 2: documentation, defining a rights policy, thinking about release management and the quality of your data, implementing a help desk, setting up training, professionalizing promotion of your dataset and its publication. Each of these activities will now be discussed in more detail.

4.1 TACTICS

4.1.1 RIGHTS POLICY

Defining and implementing a rights policy is important for data publishers who need to specify who can be held responsible for the quality and content of the datasets. The following actions help data publishers to define their rights policy:

- Define rules that specify who can be held responsible for the quality and content of the datasets. Specify these rules in “no-warranty” statements.
- Define under which license you want to publish your data. Open data should, by definition, be published under an open license (see box at the right).
- Set up an end user agreement if your situation cannot be covered by an open license in order to define specific agreements with certain user groups.
- Determine intellectual property and copyrights of your dataset.
- Based on the roles defined earlier in the strategy cluster, decide on the rights and obligations per role. Who is, based on his role, allowed to see which part of the data at which abstraction level? (closed, community, public) The rights policy needs to take ownership and privacy issues into account. The open data decision tree shown in Figure 7 can be used to identify privacy barriers and implement interventions.
- Define terms and conditions for the re-use of data: e.g. if you re-use the data you have to participate in a user working group.
- Decide if working group members and data users need to sign non-disclosure, reciprocity or other kind of agreements.
- Define and govern service level agreements about the continuation of the supply of data and the maintenance of the dataset (e.g. updates or changes).

4.1.2 RELEASE MANAGEMENT

Data owners should be deliberate about their data release policies. Each new data release often requires extensive preparation and subsequent fine-tuning. Data publishers should weigh the potential benefits and drawbacks of releasing particular datasets. Continuous maintenance and updates are important to keep the dataset usable and valuable. The amount of updates depends on the type of data, if the data is static or dynamic and how it is published (one time download vs. constant streaming). For example, a dataset with weather data that is not streamed needs to be updated more frequently than a list with cities in a specific

SPECIFY AN APPROPRIATE LICENSE:

Open licenses:

- Creative Commons
 - A flow chart for choosing a CC license: <http://creativecommons.org.au/content/licensing-flowchart.pdf>
 - Online Tool to choose a CC license: <http://creativecommons.org/choose/>
- Open Data Commons (<http://opendatacommons.org/licenses/>)
- The Open Government License (<http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/>)

area. The following activities describe how the release of data can be managed.

- Organize a launch event when publishing a dataset to get publicity for the dataset and increase value through re-use.
- Decide and specify how you will deal with updates and requests for changes. Updates and changes can range from no impact on existing re-use of the data to large impact on existing use of data. Version management might be needed to keep an overview of different versions of your dataset.

- Expectation management is very important when publishing data. Specify how often users can expect new versions of your dataset. E.g. maximum three new versions per year.
- Take requirements of users on life cycle and version management into account.

4.1.3 DATA QUALITY

Data quality policies determine the quality requirements a dataset has to fulfil before it can be published. Metadata describing the quality of the data is of utmost importance for its re-use and potential value. As this data quality activity is defined on a tactical level the following activities will focus on the perspective of an overarching entity within your organization that decides on minimum data quality requirements of datasets that are published within the name of your organization and that steers other (operational) entities within your organization to adhere to these quality requirements.

- Define and publish guidelines on what a high quality dataset published by your organization should look like with respect to the technical format used, the metadata described and other quality aspects.
- Depending on the data you want to publish, decide if you want to establish quality measurements and procedures that define and monitor data quality. This might be done by means of a compliance check or a checklist that needs to be executed before a dataset is being published (compliance check). If your data first needs to be aggregated to anonymize privacy sensitive information you might want to have a final check that the dataset does not include any sensitive information.
- When implementing a checklist/ checkpoint you might want to check the following aspects:
 - Is the dataset in a format specified as appropriate by your organization?
 - Is the dataset validated to be complete (e.g. no empty fields)?
 - Does the metadata of the dataset describe the quality of the dataset?
 - Is the dataset free of any sensitive information?
- Specify who in your organisation, e.g. a “data steward”, is responsible for continuously assessing the data quality and approving datasets before they are published.
- Monitor the usage of the data that you have published by asking feedback from the users of your data. This feedback can be collected using a LinkedIn group or another social media platform that allows for easy

interaction with your data users. You might want to ask the following questions:

- How did you use the data?
- Are you satisfied with the data quality of the dataset?
- Are you missing information in the dataset?

4.2 OPERATIONS

4.2.1 DOCUMENTATION

Providing documentation describing the dataset and its history is important to stimulate the re-use of data. The following activities are suggested for an organization that wants to make re-use of its data as easy as possible. Examples of documentation documents are shown in the box below.

- Develop manuals and frequently asked questions documentation according to fixed templates.
- Provide an overview of maintenance requests and their status.
- Make sure that all documentation (metadata, re-use agreements, FAQs etc.) is freely available and easily accessible by humans and machines (e.g. metadata can easily be queried).

EXAMPLES OF DOCUMENTATION DOCUMENTS:

- Metadata (incl. provenance)
- Frequently asked questions
- Manuals
- Re-use agreements
- Rights Policy
- List of dataset implementations/ applications

4.3 IMPLEMENTATION SUPPORT

4.3.1 HELP DESK

Offering support to the various user groups by phone or email is important to increase data re-use and user satisfaction. This role can be fulfilled by a help desk. The following points describe roles and activities to be carried out by the help desk.

- Offer support to the various user groups via different channels (e.g. LinkedIn, email, telephone) to increase data re-use and customer/user satisfaction. In order to keep it practical you might want to start by only using email to communicate with your data users.
- Work according to a service level agreement (e.g. questions are responded to within 3 working days).
- Develop and update a list with frequently asked questions.

4.3.2 TRAINING

Offering training opportunities to both people inside your organization as well as potential external users of your data, increases the value of your data. Internal data publishers might need to be trained in publishing the data according to previously defined data quality requirements. While external data users might be interested in training on how to read and interpret metadata. This can be done in various formats varying from online information to an information meeting or even a training or course.

4.4 COMMUNICATIONS

4.4.1 PROMOTION

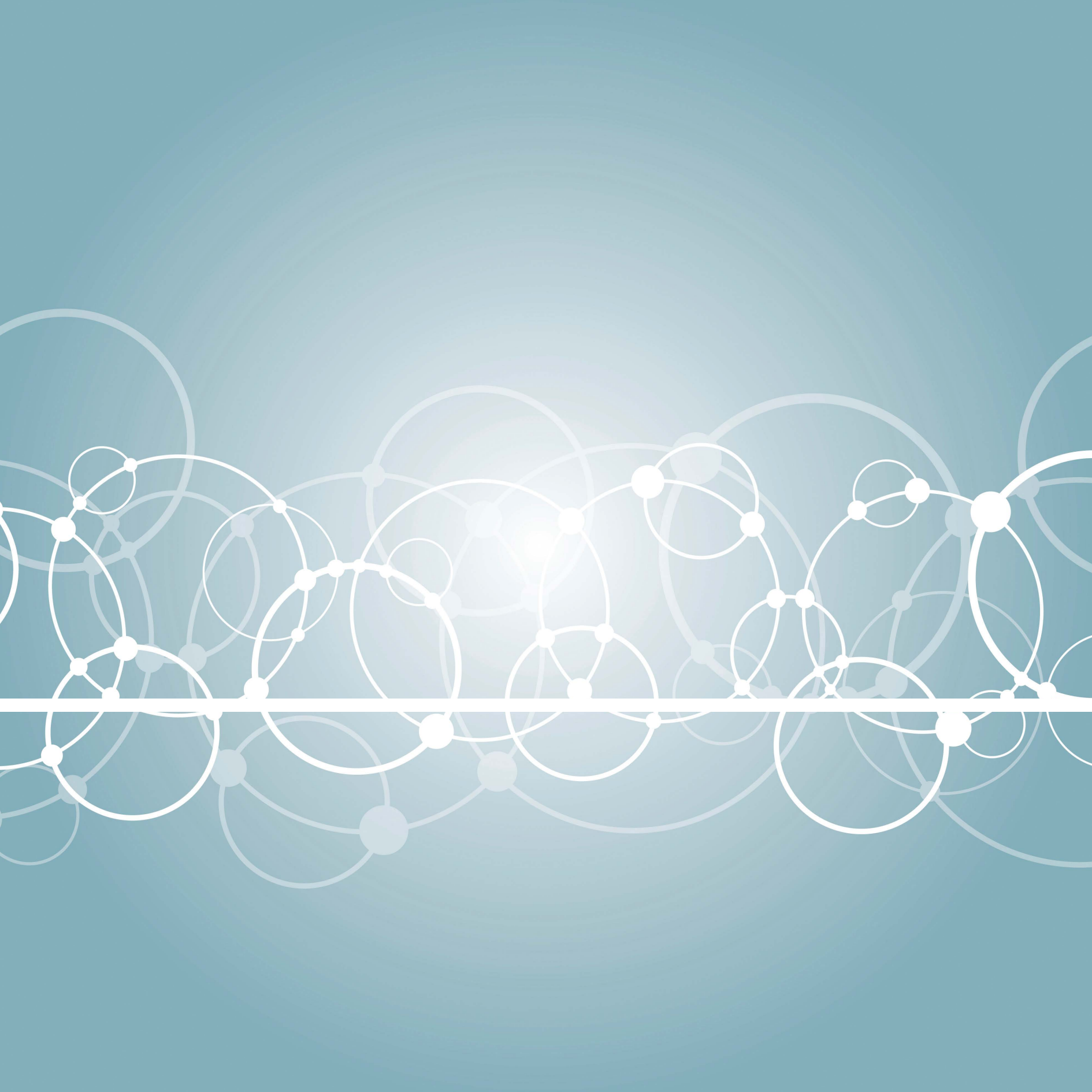
In order to increase the potential value of data re-use it is very important that datasets can be found by potential users. Besides the activities defined under the Re-use category in this booklet, it is useful to promote your dataset. The following activities help you to raise awareness for your dataset.

- Define a short “promotion strategy” for your organization. This can be as simple as one A4 page describing who your potential users are and how you want to reach them (e.g. online communities, hackatons, special events etc.).
- You might further want to define how you want to communicate about your open data in order to stimulate re-use.

4.4.2 PUBLICATION

Apart from the promotion activities, it is important to make the dataset and all information about the dataset accessible on the Internet. You can define a publication strategy to determine how often you want to inform your user community and the general public about the state of your dataset. You might want to consider publishing white papers or reports describing experiences and successful cases of re-use of your dataset.





5 LEVEL 3 ACTIVITIES OF BOMOD

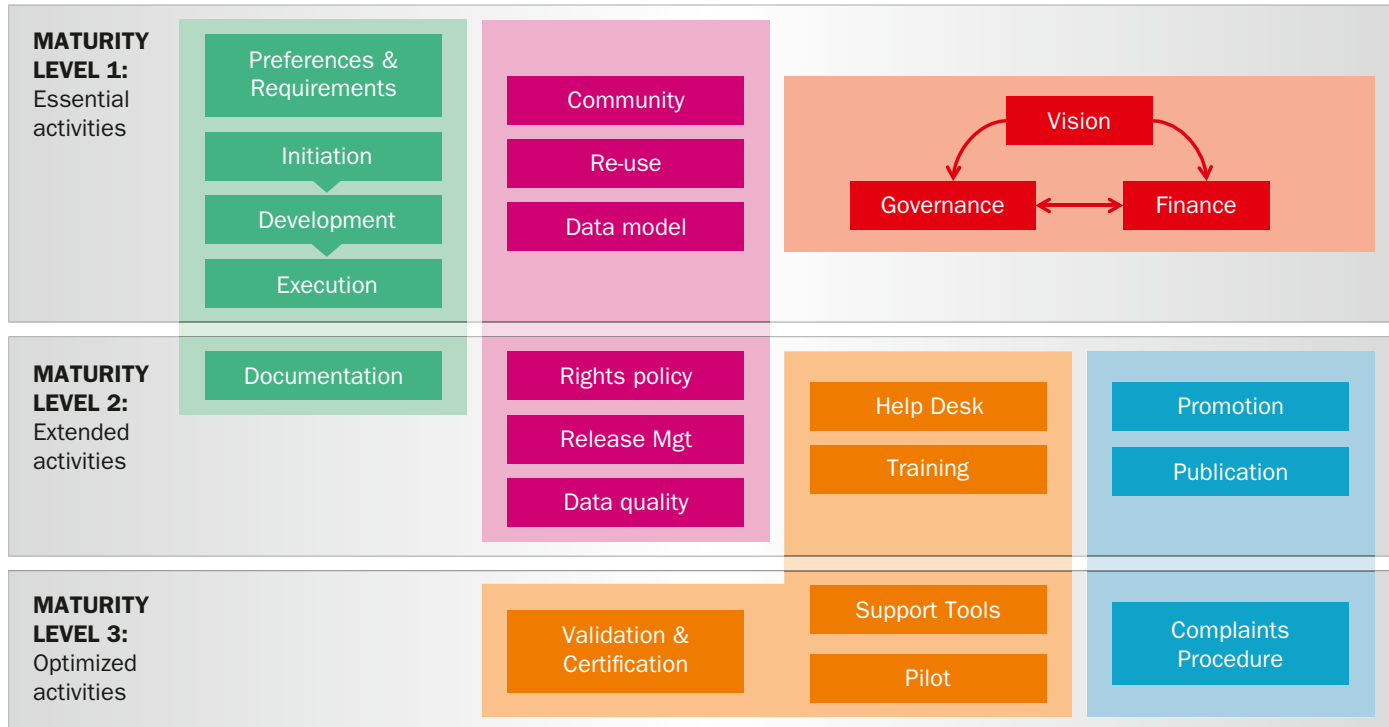


FIGURE 10: Level 3 Activities of BOMOD

(green = operations, violet = tactics, red = strategy, orange = implementation support, light blue = communication)

In addition to the activities described at Level 1 and 2, the following activities should be performed at Level 3: providing support tools, support pilots, offer validation and certification and install a complaints procedure. Each of these activities will now be discussed in more detail.

5.1 IMPLEMENTATION SUPPORT

5.1.1 SUPPORT TOOLS

In order to further support the re-use of your dataset, you might consider to offer modules and tools that support the value creation of using open data, such as data visualization tools. Going one step further you might want to offer a data platform and application programming interfaces (APIs) that allow easy access to your data.

5.1.2 PILOT

At maturity Level 3 you might consider to define actions that support users with pilot implementations of new applications making use of your data.

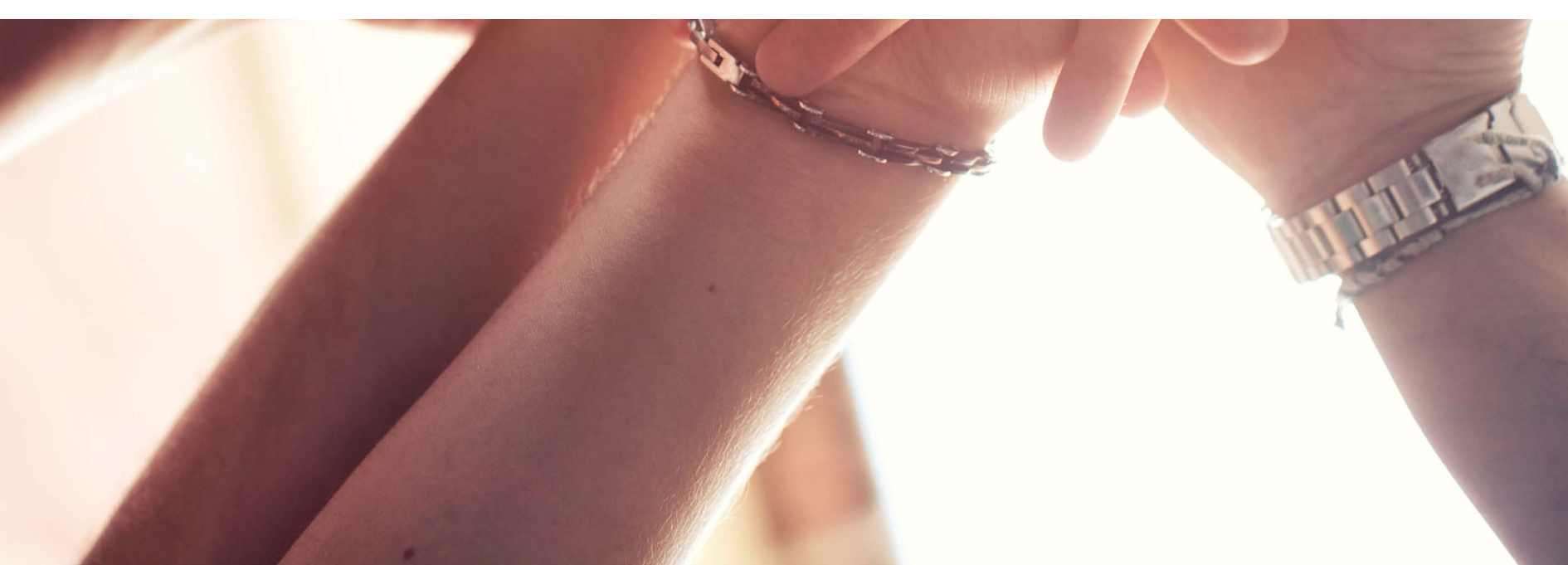
5.1.3 VALIDATION & CERTIFICATION

Once your dataset gets re-used by more users you might want to consider certifying new applications that make use of your data. However, this might be an expensive and complex process and is only recommended in cases where you want to keep a high level of control of your data.

5.2 COMMUNICATIONS

5.2.1 COMPLAINTS PROCEDURE

It is important that complaints of users with respect to your dataset are taken seriously. A complaints procedure might help to describe how complaints are handled. Complaints should always be considered as suggestions for improving the datasets.





SHARE



6 CONCLUSION

In order to have more datasets available online that fulfil the following requirements:

- being minimal 5 star,
- having sufficient metadata,
- having governance well defined,

we introduce BOMOD and the LOD roadmap. These methods support organizations to publish their data according to the requirements, ensure high technical quality, continuity and increased value from data re-use. The following guidelines are most important:

- Increase findability of datasets by describing dataset properties in metadata and using registries;
- Increase usability of datasets by describing quality aspects and intended use as well as provenance aspects in the metadata;
- Learn from your users about new ways of creating value and their feedback and requirements on your dataset;
- Think about a feasible business model for opening data.



› 2014 was a pivotal year in the evolution of open data. For the first time, most governments participated in open data initiatives, making Open Data a world-wide known and accepted concept. In practice, however, both public and private organizations are struggling to recognize and realize the potential value of open data. How will open data impact my current business model? Is my data reliable enough to build services that make a difference? How do I deal with private and sensitive information?

To help answer these questions, TNO has developed the “Beheer en Ontwikkel Model Open Data” (BOMOD). This framework can help you overcome the barriers that you might encounter when publishing data by providing a clear overview of the activities you should take into account when developing and managing open data. These activities are presented in a maturity model which allows you to easily identify the most essential activities for your specific situation. For each activity, BOMOD provides hands-on guidelines and practical tools to successfully manage and govern your process of publishing data.