European Public Sector Information Platform
Topic Report No. 2015 / 02

# Ethical and Responsible Use of Open Government Data

Author: Karolis Granickas

Published: February 2015

# Keywords:

Responsibility, ethics, open data, re-use

# Summary

This report will map and briefly describe key concepts surrounding responsible data re-use and suggest an entry point for those organizations that are in the beginner stages of responsible and ethical data re-use considerations. The report will briefly explain rationale behind responsible data re-use. The definition of responsible data will dictate the structure for the report that will cover key concept related to it.

## Table of Contents

# Introduction

More often than not, open government data community in the world has been re-using released government data on an assumption that once data is in the public domain, it is good for re-use and further publishing. Although being familiar with key risks associated with government data re-use, organizations and individuals have usually been putting a burden of mitigating these risks on governments' shoulders and pressuring them to release data that already carries no potential privacy or security dangers. In turn, the process of release of data has often been slower as government data release structures had to be built - from the policy development to an actual finger work to release datasets.

Thanks to growing expectations from the international donor community and a number of leading organizations, such as the Engine Room[1], this attitude has been changing and there is now an increasing understanding about duties and responsibilities related to open government data re-use that organizations owe to their stakeholders.

Although many organizations are assuming increased responsibilities, there is an obvious and growing demand for skills and knowledge about ethics and responsible data re-use. The resources on the Internet remain scarce even with the outstanding work done by the Responsible Data Forum[2] and some others. Having responsible data re-use policies in place are still eagerly sought good practice examples rather than a general rule.

This report is intended to map and briefly describe key concepts surrounding responsible data re-use and to suggest an entry point for those organizations that are in the beginner stages of responsible and ethical data re-use considerations. It is in no way intended to suggest an expert advice on different components of data re-use, such as privacy and security. The key goal of this report is to suggest a sketch understanding of what conversation on ethical and responsible data re-use is about.

The report will briefly explain rationale behind addressing ethics and responsible data re-use. It will remind a definition of responsible data that will give structure to further explanation of key concepts related to it.

[1] https://www.theengineroom.org/

[2] https://responsibledata.io/

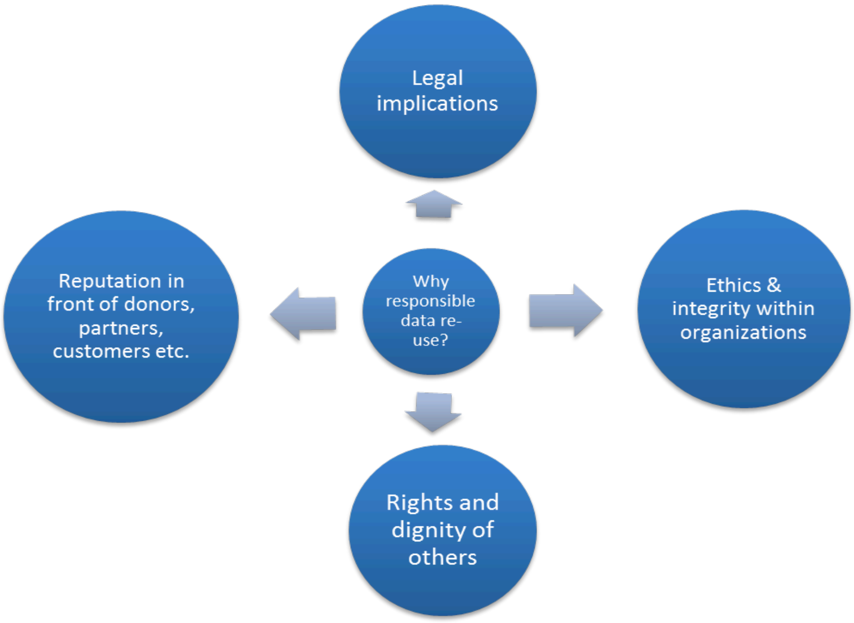
# Why should it matter to us?

Open data movement has been pressuring governments to release large amounts of information often assuming that any type of data controlled by government agencies is data that "belongs to us", and rightly so.

However, all governments have limitations as to what data can be released publicly. Most common limitations are protection of privacy, commercial or state secrecy. Governments owe a duty to their citizens to protect their privacy and secrets, as prescribed by laws.

More often than not, open data movement functioned on an assumption that anything that is released by governments is good for re-use and that the duty to respect privacy and protect secrets expires with a moment data becomes public. And that very responsibility to make sure that data released is ready for re-use often makes governments slower to publish datasets. Arguably, **sharing a moral responsibility to protect rights of others can effectively make governments trust re-users** more and make the release of data quicker.

Participants of Responsible Data Forum (it is a collaborative effort to develop useful tools and strategies for dealing with the ethical, security and privacy challenges facing data-driven advocacy lead by the Engine Room, Amnesty International and others) outlined the four underlying reasons to assume the duty to re-use data responsibly:


Legal implications
Reputation in front of donors, partners, customers etc.
Why responsible data re-use?
Ethics & integrity within organizations
Rights and dignity of others

## Legal implications

All subjects to law in the European Union have legal obligation to obey all EU, national and local data protection laws. Among a larger number of key regulations and directives that are applicable on the EU level is the **EU Data Protection Directive**[3] that finds its direct implementation in the national laws of any of the EU states. Currently, the directive is being transformed into a regulation, which will mean that we will not have to look for many additional legal acts to understand our obligations -–the regulation will be a directly applicable legal act regulating our behaviour with data.

The regulation will apply if the data controller or processor (organization) or the data subject (person) is based in the EU. Furthermore (and unlike the current Directive) the Regulation will also apply to organizations based outside the European Union if they process personal data of EU residents. According to the European Commission "personal data is any information relating to an individual, whether it relates to his or her private, professional or public life. It can be anything from a name, a photo, an email address, bank details, posts on social networking websites, medical information, or a computer's IP address.

There is much advice out there on **what laws and legal implications may occur to organization re-using open data**. One of the more practical introductions to an issue is the Open Data and Privacy Primer developed by the Open Rights Group.[4]

## Ethics and integrity within an organization

High ethical standards, respect to dignity and organizational integrity are few of key employee motivators. If an organization negligently demonstrates a lack of care towards privacy and dignity of others, it either leads to double standards or may translate into lack of care towards employees' rights. How can one expect protection of privacy within organizations from leaders who demonstrate disrespect to privacy and security outside an organization?

[3] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data

[4] Can be found here: https://www.openrightsgroup.org/ourwork/reports/open-data-and-privacy-primer

## Rights and dignity of others is data re-users' concern

Large part of open data re-users are NGOs who often declare missions that are directly linked to rights of certain social groups. Having responsible data policies send a clear signal to all stakeholders that organization does in fact care about its affected groups, especially those vulnerable.

## Reputation in front of donors, partners, customers

Having data re-use policies in place does send a clear signal to donors, partners, customers and other stakeholders that the organization treats its activities with care and high ethical standards. Increasingly so, donor community is demanding that such policies are in place if organizations are to receive funding.

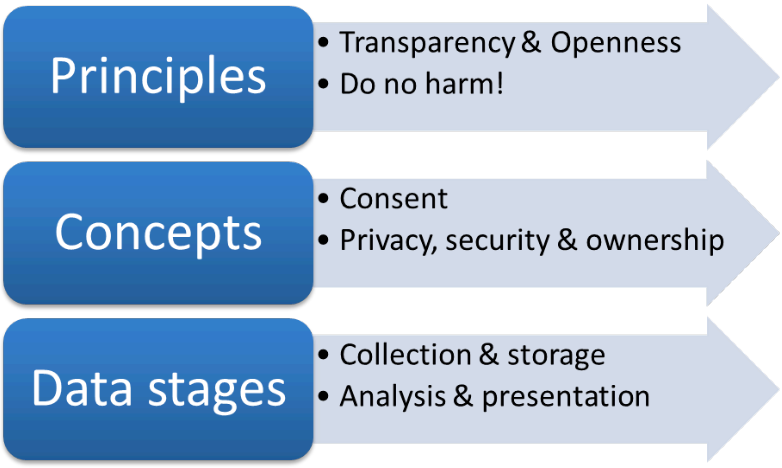
Responsible Data Forum developed a short, but structured message matrix that helps understanding possible benefits of having responsible data re-use policies to various societal groups.[5]

# What is responsible data re-use?

The Responsible Data Forum describes responsible data as a **duty** to ensure people's rights to **consent, privacy, security and ownership** around the information processes of **collection, analysis, storage, presentation and re-use**, while also respecting the values of **transparency and openness.**

This definition will dictate the structure and logical framework of this report. We will break the definition down and will briefly describe each of the components with suggestions on further read and resources.

[5] Can be found here: https://wiki.responsibledata.io/Messaging_Matrix


Principles
• Transparency & Openness
• Do no harm!
Concepts
• Consent
• Privacy, security & ownership
Data stages
• Collection & storage
• Analysis & presentation

## Transparency & Accountability vs. Do No Harm Principle

It is important to start by understanding the spirit and underlying principles of responsible and ethical data re-use. First of all, the concepts of privacy, security, commercial or state secrecy can arguably be pulled under one umbrella (in terms of data re-users' behaviour judgment) that is "do not harm" principle (although the definition is silent on "do no harm" explicitly). Although the principle originated in the context of humanitarian assistance and development, it can be applied in a broader sense and can mean, that data re-users must **do all within their powers to not cause any harm to any of the stakeholders that can rise as a direct or indirect result of open data re-use.**

Many conversations are shaped around confrontation between "do not harm" principle and concepts of transparency, accountability. Although it may appear that these are two opposites, effectively, the "do no harm" principle is merely a limitation to promotion of transparency and accountability. This means, that protection of promotion of transparency and accountability is the basic principle of modern democracies which is not questioned, but that has certain limitations linked to a "do no harm" principle encompassing concepts of privacy and security.

There are no hard rules on how to balance these two principles and the conversation around them will continue developing as our societal norms change. However, one of the criteria that may guide judgment of organizations is a suggestion by authors of the book called "Ways to practise responsible development data"[6] is that the "do no harm" is for powerless and

[6] Can be found here: https://responsibledata.io/wp-content/uploads/2014/10/responsible-development-data-book.pdf

transparency and accountability is for powerful. As the book puts it:

> When discussing specifics, it's rare to find a case where thoughtful, dedicated and informed people won't agree on what is permissible and what isn't permissible when promoting transparency and accountability. There is a key difference between personal data and data that should be made 'open'; **as a broad rule, we believe that the right to privacy is for those without power, and transparency is for those with power.**

## Right to consent

Returning once again to the definition of responsible data, the very first concept included is the **right to consent.** Largely, this is something that will function on an assumption when it comes to open government data re-use and is more explicit and central when it comes to research and direct data collection. However, it is worth understanding the concept and knowing that a right to consent (whatever stage it may be acquired in) is an essential in any of the data acquisition, analysis and re-use stages.

Informed consent is the mechanism through which **people agree to provide information for research or data collection** projects. Generally, consent has been understood as something that is given by individuals during direct interaction with researchers or surveyors, and is composed of three components: disclosure of research objectives and any risks or negative consequences of participating capacity of individuals to understand the implications of participating voluntariness of their participation. [7]

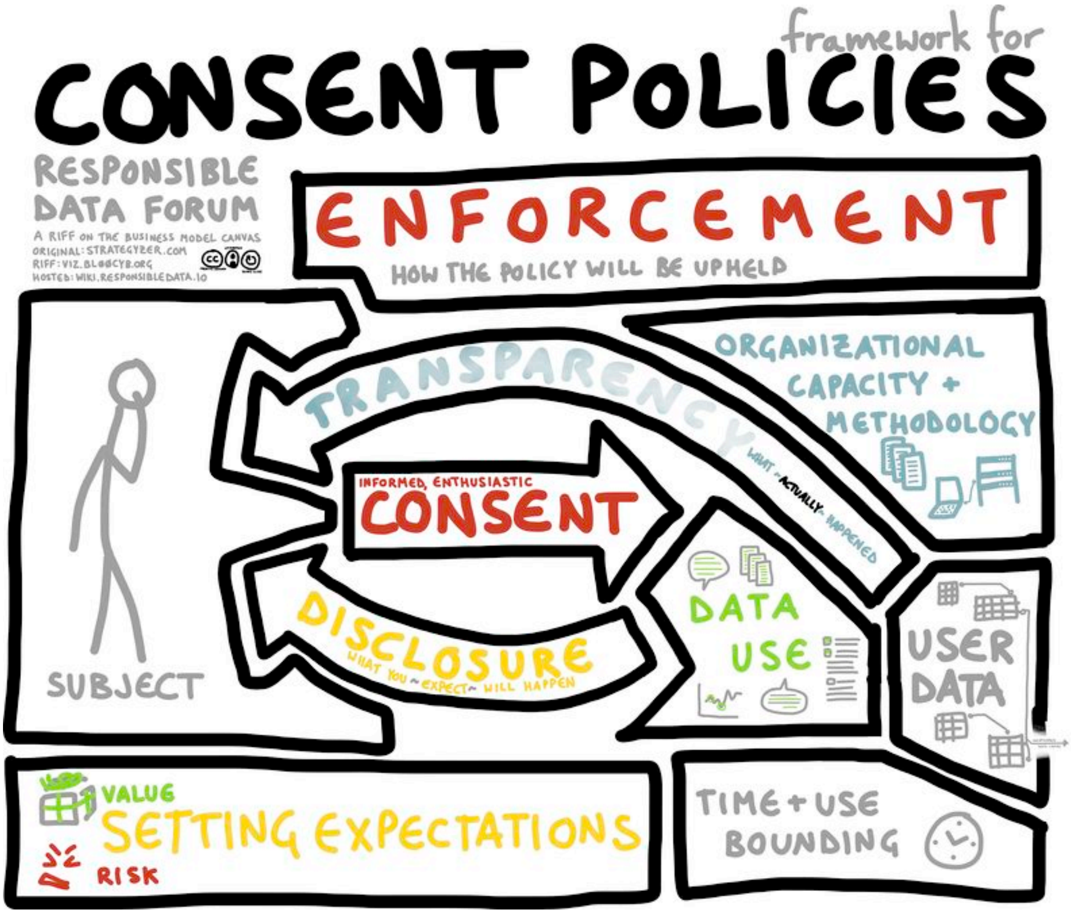
The guide developed by the Responsible Data Forum suggests that informed consent include plain language, easy-to-understand explanations of the types of data to be collected, the purposes of collecting data, the intended and potential unintended uses of that data, who has access to and control over the data, risks of data leakage to third parties, and any benefits to participation in data collection. A "responsible data" approach to informed consent would discuss specific steps taken to mitigate risks to all aspects of holistic security and participants' well-being: digital, operational-physical and psychosocial.[8]

[7] Ways to practise responsible development data: https://responsibledata.io/wp-content/uploads/2014/10/responsible-development-data-book.pdf

[8] Can be found here: https://responsibledata.io/forums/primer-responsible-data-in-development/

In case of re-use of open government data, an obligation to acquire informed consent, perhaps, is not with data re-user, but with government. However, this does not mean that data re-users, once acquired the data, have no obligations to achieve consent if there are reasonable grounds to believe that the re-use purpose is fundamentally different from the purpose to which the individual expressed consent at an earlier stage. Although this more often than not will not raise a lot of questions (when re-using government data), it is worth assuming an obligation to raise questions, whenever there are grounds to believe that there may be any sort of harm done to a particular individual or a societal group if data in hand is re-used and publicized.

Disregarding the question whether the active consent acquisition will be required in many instances during organizations work, it is highly recommended to have the consent policies in writing to both serve as mechanism for possible planning and decision-making and also as an indication of high responsibility and ethics standards at organizations. A sample frame of how to structure consent policies was helpfully suggested by the Responsible Data Forum:



# Privacy

Perhaps, largest part of the conversation about responsible and ethical data re-use is around the concept of privacy. There are countless articles, guidelines, primers and other literature on privacy online and there is no question that any organization that has anything to do with re-using data must be well familiar with the concept of privacy, legal requirements, risks and mitigations associated.

Privacy is concerned with control over information, who can access it, and how it is used. As Daniel Solove[9] notes this has many dimensions, from concerns about intrusive information collection, through to risks of exposure, increased insecurity or interference in their decisions that individuals or communities are subjected to when their 'private' information is widely known. Privacy is generally linked to individuals, families or community groups, and is a concept that is often used to demarcate a line between a 'private' and 'public' sphere.

Article 12 of the Universal Declaration on Human Rights states "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation". It has been argued that privacy is a western concept, only relevant to industrialised societies – yet work by Privacy International has found privacy concerns to be widespread across developing countries, and legal systems across the world tend to recognise privacy as a concern, even if the depth of legal rights to privacy and their enforcement varies. [10]

It is safe to say, that any organization re-using government data may at some point of activities face risks related to privacy. In these cases, it is useful to have policies in place describing possible risks and their mitigation plan. The key risk, obviously, is to publish data that can uncover particular individuals' private information.

[9] A Taxonomy of Privacy, George Washington University Law School: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=667622

[10] Discussion notes on open data & privacy by the Open Data Research Network: http://www.opendataresearch.org/content/2013/501/open-data-privacy-discussion-notes

# Privacy risks

Guidelines of Open Data and Privacy by the Government of South Australia[11] can also be useful for many other organizations. It distinguishes a number of key risks related to privacy as follows:

- disrespect to privacy can cause humiliation, embarrassment or anxiety for the individual, for example from a release of health data, it might be concluded that an individual accessed treatment for a sensitive sexual health condition;
- can have an impact on the employment or relationships of individuals;
- can affect decisions made about an individual or their ability to access services, such as their ability to obtain insurance;
- can result in financial loss or detriment;
- can pose a risk to safety, such as identifying a victim of violence or a witness to a crime.

The first step to managing privacy risks in the release of a public sector dataset is to undertake an initial assessment of the risk of making that data publicly accessible. Assessing these risks will require a detailed consideration of the data to be released. Methods used include:

- determining any specific unique identifying variables, such as name;
- cross-tabulation of other variables to determine unique combinations that may enable a person to be identified, such as a combination of age, income, postcode;
- Acquiring knowledge of other publicly available datasets and information that could be used for list matching. The level of privacy risk will be dependent on the likelihood that identification could occur from the release of the data and the consequences of such a release.

The obvious risk mitigation technique in most cases is de-identification – removing any sort of unique identifiers that may track the data to sensitive information. A further advice on how can organization do that can be found here in the Australian Government Guidelines on Open Data and Privacy. [12]

[11] http://archives.sa.gov.au/sites/default/files/20140303%20Privacy%20and%20Open%20Data%20Guideline%20Final%20V1_Copy.pdf

[12] Guidelines of Open Data and Privacy by the Government of South Australia: http://archives.sa.gov.au/sites/default/files/20140303%20Privacy%20and%20Open%20Data%20Guideline%20Final%20V1_Copy.pdf

In addition, it is worth mentioning a resource for governments, helping to shape thinking around privacy and open data policies, developed by Open Government Guide.[13]

## Security

Although security is somewhat linked to privacy, in some context it finds itself a rather different role. For instance, there is a wide ongoing conversation on security of human rights activists, working with information and its re-use. Organizations start to adopt policies that subject security of their employees in the light of possible dangers resulting from working with information.

There is an increased attention paid to organizational policies that adapt security protocols and tactics to encompass: 1) digital information security; 2) physical and operational security; and 3) psychosocial well-being required for good security implementation. These aspects comprise a new three-part approach to **"holistic security."** Digital security is not only a question of a focus on software or tools. It requires integrating emotional well-being, personal and organizational security. Good implementation of digital security tools and tactics requires attending to the practitioners' psychosocial capacities to recognize and respond dynamically to different threats to themselves and to participants related to project data collection and communications. [14]

In addition, security can also be a highly technical concept dealing with security of the data itself. While the report is not intended to cover any of the technical tips, it is worth mentioning a resource developed by the Responsible Data Forum that is a set of checklists that attempt to list all of the possible information security steps an organization can take to protect itself. It breaks these steps into the smallest atoms of action required to complete them. The overall goal of the resource is to support security trainers to produce an easy-to-use action plan that is customized and approachable for organizations.[15]

[13] Open Data and Privacy by Open Government Guide: http://www.opengovguide.com/topics/privacy-and-data-protection/

[14] Primer on Responsible data in Development by Responsible Data Forum: https://responsibledata.io/forums/primer-responsible-data-in-development/

[15] Atomized Security Plan for Organizations, Responsible Data Forum: https://responsibledata.io/forums/atomized-security-plan-for-organizations-2/

# Risk mapping

There are numerous tools out there helping to identify and mitigate risks associated with privacy and security concerns. The Responsible data Forum have developed an entry primer suggesting a snapshot of what risk mapping steps can look like, they include:

**1. Identify the Persons at Risk in the event of exposure**

Definition of Persons at Risk: Any entity at risk of being by the exposure. Therefore, not restricted to the data owner or collector.

**2. Identify Knowledge Assets that can be extracted from the data collected**

Definition of Knowledge Assets: Discrete data points, information extracted from collections of discrete data points, information extracted from meta analysis of data points, information extracted from the mashup of the collected data and external data sources.

**3. Evaluate the importance of each knowledge asset to the campaign**

The importance is used in combination with Risk assessment to determine what data to collect. Importance is rated on this scale:

- Low Importance: knowledge assets that have little or no relevance to the success of the campaign
- High Importance: knowledge assets that have significant relevance to the success of the campaign
- Must Have: knowledge assets that are crucial to the success of the campaign

**4. For each Type of Harm:**

Evaluate probability and severity of harm for each type of harm for each person at risk by each knowledge asset

Probability of Harm:

- Low - Assessed as 49% or less probability of harm
- High - Assessed as 50% or more probability of harm

Severity of Harm:

- Low - Assessed as causing little to no harm to the Person at Risk
- High - Assessed as causing moderate to severe harm to the Person at Risk

- No Go - Assessed as causing catastrophic harm to the Person at Risk

The output of this process is a high-level score for each Person at Risk, with detailed matrices for each Type of Harm as supporting documentation.[16]

[16] Risk Mapping, Responsible Data Forum: https://wiki.responsibledata.io/Responsible_Data_Risk_Mapping

# Responsible data as an organizational policy

It is understood that many of the concepts dealt with above are not new for a vast majority of organizations working with re-using open government data. However, as mentioned in the beginning, this report is not intended to suggest an in-depth analysis of any of these concepts. On the contrary, it is intended to map and sketch possible components that any responsible data policy may contain.

There is an increased pressure to deal with responsible data re-use in a well-planned manner – to have policies within organizations explaining risks and their mitigation techniques related to re-use of open government data.

Although there are not many good practice examples yet, the Oxfam in the UK went through the process of creating one and shared their useful experience in a blog post.[17]

Oxfam's Responsible Data Policy involves ensuring they obtain informed consent from those who have shared their data, not collecting unnecessary or potentially damaging data, and protecting privacy through anonymization, restricted access or encryption. Furthermore, it includes the responsibility to fairly represent the contributors of the data in analysis and perhaps involve them in the process of how data is used by adopting less extractive and more empowering methods.[18]

Although there are not many good practice examples up to now, simply put, the quicker organizations will realize that having responsible data policy is essential the quicker they will get ahead of the curve in terms of having exemplary ethical profile not only in front of the donor community, but also in front of any other stakeholder concerned.

[17] https://responsibledata.io/developing-an-organisational-policy-for-responsible-data/

[18] https://responsibledata.io/developing-an-organisational-policy-for-responsible-data/

# Concluding remarks

The conversation about responsible and ethical open government data re-use is by no means new to a vast majority of organizations working with data. However, with constantly growing mutual expectations between governments, organizations and other stakeholders it is important to no treat responsibility issue in an ad hoc and fragmented manner. Taking positive steps towards creating policies and implementing them will result in increased trust, reputation and eventually smoother processes of open data release, acquisition and re-use.

This is not to say, that every organization should now develop a hard set of universal rules and apply them strictly. Needless to say, concepts of privacy, security, consent and others are very context-sensitive and should be dealt with in a manner that takes all relevant aspects into account. Therefore, organizations will be increasingly required to deal with these concepts on a flexible, but importantly, policy level.

The remaining challenge lying ahead of all open government data community is raising awareness about importance of responsible use of government data. Organizations, such as Engine Room and some others are on the right track, but there is still a lot to be done in order to achieve that **responsible re-use of government data is an absolute must for any organization that seeks to have a reputational profile in an international and national arena**.

# About the Author

Karolis Granickas is Project Leader at Transparency International Lithuania. His focus is on open government and people engagement using ICT. He coordinates Chapter's digital initiatives such as www.manoseimas.lt (parliamentary monitoring tool) and www.parasykjiems.lt (freedom of information tool), www.jurgiokepure.lt (municipal transparency tool) among others. Karolis is also an Independent Researcher with the Open Government Partnership Independent Reporting Mechanism.

Karolis has LLB degree in International Law from Westminster University, London, and LLM degree in EU Law from Maastricht University, the Netherlands.

# Copyright information

© 2015 European PSI Platform – This document and all material therein has been compiled with great care. However, the author, editor and/or publisher and/or any party within the European PSI Platform or its predecessor projects the ePSIplus Network project or ePSINet consortium cannot be held liable in any way for the consequences of using the content of this document and/or any material referenced therein. This report has been published under the auspices of the European Public Sector information Platform.

The report may be reproduced providing acknowledgement is made to the European Public Sector Information (PSI) Platform.