



D4.7 Guidelines for data owners and data publishers

Project Acronym Apps4EU

Grant Agreement number: 325090

Project Title: Apps for Europe

Revision: v1.0

Authors:

Maarten Brinkerink (Netherlands Institute for Sound and Vision)
Michelle Brook (OKF)
Christian Villum (OKF)
Sander van der Waal (OKF)

Project co-funded by the European Commission within the ICT Policy Support Programme

Dissemination Level

P	Public	X
C	Confidential, only for members of the consortium and the Commission Services	

Revision History

Rev.	Date	Authors	Org.	Description
V.0.1	06/12/2013	Sander van der Waal, Christian Villum	OKF	First draft
V.0.2	13/12/2013	Maarten Brinkerink, Christian Villum	OKF, NISV	Full first version with all topics covered.
V.0.9	18/12/2013	Christian Villum, Michelle Brook	OKF	Final version before comments
V.1.0	31/01/2014	Sander van der Waal	OKF	Final Version for submission

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Note:

Materials that are part of deliverables 4.2, 4.3, 4.6 and 4.7 constitute some duplicate material. This is to ensure that the guidelines constitute coherent documents and do not require the reader to go back and forth between documents to read all that is relevant for the topic.

Table of Contents

1. Introduction	4
2. Releasing the right data	4
Asking the community	4
Ease of release	5
Observe peers	5
3. Make your data discoverable.....	5
Existing tools.....	6
4. So I've Released Some Data, Now What?.....	6
Tell the world!	6
Understanding your audience	7
Finding your audience.....	7
Government data communities	7
Scientific data communities.....	7
Cultural data communities	8
Post your material on third-party sites.....	8
Making your communications (and data source) more social-media friendly	8
5. Engaging with your community	8
Establish a social media presence.....	9
Be active on discussion lists - or even setup your own discussion list.....	9
6. Getting folks in a room - meeting in person.....	9
Unconferences, Meetups and Barcamps	10
7. Keeping your community in the long term.....	10
How to maintain your community	10
Be a community member yourself - empower members to lead with you	10
Be fair and transparent.....	11
Update your data sets/keep the stream going.....	11
Highlight successful reuse cases	11

Abstract

Guidelines to help data owners and publishers to engage early and often with relevant reuse communities (for government, science and culture data). The Guidelines will also provide practical advice on how to create close feedback loops and keep the community alive after the Competition.

1. Introduction

This guide will help data owners and data publishers engage with communities to maximize impact and reuse of their data.

As a rule of thumb when releasing data it is good to start out small, simple and fast. This way you'll invite feedback from audiences of re-users eager to access more data and looking for ways to start a conversation with you to help you release not only specific data that they're looking for (thereby helping you identify what data would be mostly sought after), but also providing you with feedback on how these data most optimally should be released.

In other words, there is no requirement that all your datasets must be made available right from the beginning. Starting out by releasing just one dataset, or even one part of a large dataset, is fine. Remember this is about innovation: Moving as rapidly as possible is good because it means you can build momentum and learn from experience, as mentioned above, by collecting feedback from your end-users. Innovation is as much about failure as success and not every dataset will be useful.

Involving end-users is key: Engage early and engage often. Engage with actual and potential users and reusers of the data as early and as often as you can, be they citizens, businesses or developers. This will ensure that the next iteration of your data is as relevant as it can be.

It is essential to bear in mind that much of the data will not reach end users directly, but rather via 'info-mediaries'. These are the people who take the data and transform or remix it to be presented. Engage with infomediaries first. They will reuse and repurpose the data and help you reach a broader audience.

2. Releasing the right data

Choosing the dataset(s) you plan to make open is the first step to ensure buy-in from your users and the community – though remember that the whole process of opening up data is iterative and you can return to this step if you encounter problems later on.

If you already know exactly what dataset(s) you plan to open up then things are simpler. However, in many cases, choosing which datasets to focus on to begin with is a challenge. How should one proceed in this case? There are different ways:

Asking the community

We recommend that you ask the community right from the first instance. These are the people who will be accessing and using the data, so they are likely to have a good understanding of which data could be valuable.

1. Prepare a short list of potential datasets that you would like feedback on. It is not

essential that this list concurs with your expectations, the main intention is to get a feel for the demand.

2. Create a request for comment. Make it brief and to the point (Internet attention span is short).
3. Publicise your request on one or more web pages. Make sure that it is possible to access the request through its own URL. That way, when shared via social media, the request can be easily found.
4. Provide easy ways to submit responses. Avoid requiring registration, as it reduces the number of responses.
5. Circulate the request to relevant mailing lists, forums and individuals, pointing back to the main webpage.
6. Run a consultation event. Make sure you run it at a convenient time where the average business person, data wrangler and official can attend. This may very well be an online event.
7. Ask a politician to speak on your agency's behalf. Open data is very likely to be part of a wider policy of increasing access to government information.

Ease of release

Sometimes, rather than deciding which data would be most valuable, it could be useful to take a look at which data is easiest to get into the public's hands. Small, easy releases can act as the catalyst for larger releases and behavioural change within organisations.

Be careful with this approach however. It may be the case that these small releases are of so little value that nothing is built from them. If this occurs, faith in the entire project could be undermined.

Observe peers

Open data is a growing movement. There are likely to be many people in your topic area who understand what other areas are doing. Formulate a list of data on the basis of what those other data publishers are doing.

3. Make your data discoverable

Open data is nothing without users. You need to be able to make sure that people can find the source material. The most important thing is to provide a neutral space which is stable, reliable and technically proficient. The easier it is for users to discover and reuse data, the faster new and useful tools will be built. For to maximize appeal among potential re-users your data should be easily accessible, both legally and technically.

Make your data legally accessible

To give legal access you should apply an open license to your data which allows non-restrictive reuse - thereby removing any legal obstacles that might send your users to other sources. To

learn more about open licensing and to find a list of open licenses, visit [Open Definition](#).

Make your data technically accessible

Next, you should ensure technical availability. Either by allowing full bulk download of your data, or - even better - by offering re-users to engage with your data via an application programming interface (API) - which specifies how the software components (your data and the user's application) can interact with each other. To read more about API's visit, for instance, this [Wikipedia page](#).

A good resource for better understanding open data and it's release is the [Open Data Handbook](#).

Existing tools

There are also a number of tools which are live on the web that are specifically designed to make data more discoverable. One of the most prominent is DataHub.io, which is a free-to-use catalogue and data store for datasets from around the world. The site makes it easy for individuals and organizations to publish data and for data users to find material they need.

In addition, there are dozens of specialist catalogues for different sectors and places. Many scientific communities have created a catalogue system for their fields, as data are often required for publication. Getting your data into those, if possible, would likely make it much easier to find a broad audience.

Resist the urge to build the software to support the catalogue from scratch. There are free and open source software solutions (such as [CKAN](#)) which have been adopted by many data publishers already. As such, investing in another platform may not be needed.

4. So I've Released Some Data, Now What?

Once legal and technical issues are sorted, your focus should go into establishing a community around the data and ensuring an ongoing feedback loop to help you improve your data source. Setting up a community can be done in many ways, so our suggestions here are by no means conclusive. Many more elaborate sources and literature are available that can add to a deeper understanding, and we encourage further exploration to support the following tips.

Tell the world!

First and foremost, make sure that you promote the fact that you've embarked on a campaign to promote open data in your field. If you release a bunch of datasets, it's definitely worth spending a bit of time to make sure that people know (or at least can find out) that you've done so.

In addition to things like press releases, announcements on your website, and so on, you may consider:

- Contacting prominent organisations or individuals who work/are interested in this area
- Contacting relevant mailing lists or social networking groups (see suggestions later herein)
- Directly contacting prospective users who you know may be interested in this data

When doing so, it is often very useful to be aware of the nature of your data and seek target

audiences accordingly.

Understanding your audience

Like all public communication, engaging with the data community needs to be targeted and in this guide we'll focus on cultural data, government data and scientific data. Like all stakeholder groups, the right message can be wasted if it is directed to the wrong area.

Digital communities tend to be very willing to share new information, yet they very rapidly consume it. Write as if your messages will be skimmed over, rather than critically examined in-depth.

Finding your audience

Online communities have different ways of organizing - as well as with regards to which channels they use to communicate. For data communities, however, discussion lists seems to be the medium of choice for many groupings. In this section we highlight some of the most prominent communities within the government, scientific and cultural data spaces.

Government data communities

Government data communities are very diverse by nature and often consists not only of data users, but also members with political interests - such as activist groups, lobbyists and government officials tasked with either releasing data or learning from the experience of others in order to do so. These will therefore have different incentives to help spreading your message.

Many forums and discussion lists exists, but some of the most notable includes the [Open Government Working Group](#) of the Open Knowledge Foundation - a list of over 900 open government data enthusiasts and professionals - which also runs a [website](#) with relevant resources. Another community is the [EU Open Data discussion list](#), which as the name implies focuses on European government data.

Worth mention is also the newly launched [Global Open Data Initiative](#), which runs a [Google Group](#) with discussions among its 200+ members. The Global Open Data Initiative aims to be a point of focus and discussion for civil society around open government data.

Lastly, we encourage you to tap into twitters streams based on hashtags, for instance tags such as #opengov and #opengovernment - as well as more general ones such as #opendata.

Scientific data communities

Scientific data covers several adjacent areas including research data, educational data and even environmental and sustainability data. One of the most recognized communities is Open Knowledge Foundation's OpenScience Working Group, which convenes in a [discussion list](#) and runs the [OpenScience website](#). Another recognized space is the [Open Science Federation Google+ forum](#) - and in other social media terms tapping into twitter streams around hashtags such as #openscience and #openaccess is highly recommended.

For the research end of the spectrum, you might be either looking to engage with publicly funded research groups - often found in universities - or moving towards a more proprietary, patented space where communities are to be found in the commercial sector of companies and start-ups.

For environmental data discussions and enthusiasts, a useful resource is the Open Sustainability Working Group, also initiated by the Open Knowledge Foundation. They run a [discussion list](#) and can furthermore be interacted with via their [Twitter account](#).

Cultural data communities

Some European countries already have national initiatives that bring together cultural institutions that make open data available, and sometimes also connect them to a network of developers that are interested in cultural data and content to reuse in their projects. Examples of this are the Open Cultuur Data networks in [the Netherlands](#) and [Belgium](#) and [Avoin GLAM](#) in Finland. Another important initiative that reuses cultural data and content to specifically keep in mind is the [GLAM-WIKI project](#), that supports cultural organisations and other institutions who want to work with Wikimedia to produce open-access, freely-reusable content for the public. This initiative is often supported by the national Wikimedia Chapters and run by Wikimedia volunteers world-wide. Lastly [Europeana](#) can serve as a ‘channel’ for data owners from the cultural heritage sector who are interested to see their metadata (and content) being reused by third parties. Through their API and hackathons this initiative [promotes the reuse](#) of the CC-0 licensed metadata Europeana has collected. Lastly, the Open Knowledge Foundation’s [OpenGLAM](#) network promotes free and open access to digital cultural heritage, bringing together hundreds of cultural metadata enthusiasts.

Post your material on third-party sites

Many specialized blogs have created a large readership in specialised topic areas. It may be worthwhile approaching them to offer adding an article about your initiative on their site. These can be mutually beneficial. You receive more interest and they receive a free blog post in their topic area.

Making your communications (and data source) more social-media friendly

There are several things that you can do to make sure that your message (and data) can be more easily shared. Some tips:

- Provide unique pages for each piece of content (message or dataset): When content is shared with others, the recipient of the referral will be sent directly to the content in question.
- Make sure a short quickly-readable abstract is available on the landing page. Many visitors will give your content around 10 seconds of attention, so make sure to get the message across in a quick fashion - and then lead readers looking to learn more towards a space with elaborate background information.
- Offer pre-fabricated social media updates on your landing page, so that people can either copy/paste them, or - even better - share directly via integration the most popular social media platforms. All these have easy tools for implementation.
- Integrate social media tools directly to your data source, so that users can dive in and share links to specific parts of your data that they find particularly interesting. Some of the above mentioned services, for instance CKAN, feature such functionality out of the box.

5. Engaging with your community

Engaging with a community is in its simplest form to engage in conversation, listen and to inspire others to join and contribute. There are numerous ways of doing that, and naturally not a one-size-fits-all approach. But ultimately it requires you to set up the channels to have

conversations, be available and reachable and to include and value your community's opinions and contributions.

Establish a social media presence

While email will get you a long way and should not be discarded, it's not the most transparent or flexible type of communication and you should consider setting up others channels for engaging with people who want to join the community around your data. One of the easiest ways of doing that is via social media, which allows you to have public conversations that let's newcomers backtrack and get up to speed even if joining conversations later. Social media also let's you have a human face, and this is important especially if you're an organization or institution moving into the open data field. Furthermore it should be highlighted that social media in general are informal and should be treated as such. Consider sidestepping the organizational code of conduct if any such exist - to have more impromptu, casual and engaging conversations. In community building, people generally joins people - rather than "faceless" organizations. This goes down to even simple things like when you need to pick an avatar; use a photo of yourself or, for shared accounts, make a group photo - not your logo.

Be active on discussion lists - or even setup your own discussion list

In the open data community and adjacent communities like among developers, discussion lists and forums are often among the most used tools of conversation. To engage people there be an active contributor and be responsive to those writing you back. You might even want to consider setting up your own forum or discussion list - there are many free tools for that, among other [Mailman](#) or [Google Group](#). In the beginning of running your own forum make sure to post rather frequently to ensure that those joining feel a vibrancy. After a while they'll hopefully feel inspired to contribute with questions and ideas themselves. Therefore you should also make sure to not use it as a channel for announcements alone; rather you should make sure to send out open-ended questions which serve to engage your audience and inspire them to give you feedback.

As a bonus, running a forum or a discussion list also ensures that you do not need to answer the same questions again and again, as you can then link to the archives to bring newcomers up to speed.

Be available

If there is one thing that might potentially kill off your forum it's the unavailability of its admin. So make sure to make time for attending to questions and conversations. Or even better: Equip the most active community members with the agency to partially act on your behalf. This will be empowering to them and lighten the load on your shoulders.

6. Getting folks in a room - meeting in person

Face-to-face events can be a very effective way to encourage others to use your data. Reasons that you may consider putting on an event include:

- Finding out more about prospective reusers
- Finding out more about demand for different datasets

- Finding out more about how people want to reuse your data
- Enabling prospective reusers to find out more about what data you have
- Enabling prospective users to meet each other (e.g. so they can collaborate)
- Exposing your data to a wider audience (e.g. from blog posts or media coverage that the event may help to generate)

There are also lots of different ways of running events, and different types of events, depending on what aim you want to achieve.

Unconferences, Meetups and Barcamps

As well as more traditional conference models, which will include things like pre-prepared formal talks, presentations and demonstrations, there are also various kinds of participant driven events, where those who turn up may:

- Guide or define the agenda for the event
- Introduce themselves, talk about what they're interested in and what they're working on, on an ad hoc basis
- Give impromptu micro-short presentations on something they are working on
- Lead sessions on something they are interested in

There is plenty of documentation online about how to run these kinds of events, which you can find by searching for things like: 'unconference', 'barcamp', 'meetup', 'speedgeek', 'lightning talk', and so on. You may also find it worthwhile to contact people who have run these kinds of events before (in your own or other countries), who will most likely be keen to help you out and to advise you on your event. It may be valuable to partner with another organisation (e.g. a civic society organisation, a news organisation or an educational institution) to broaden your base participants and to increase your exposure.

7. Keeping your community in the long term

The work with releasing data shouldn't stop after the initial release of the data. Rather, the community that will hopefully build around the data source, once engaged, will be a valuable ongoing resource for you in the continuing work of releasing more data. Yet it will require ongoing nursing from your side to keep the feedback loop going and ensure the continued engagement from your community. The benefits are, however, many - as you will be able to have your community work for you in a sense: Spreading the knowledge of your data and inviting more people to engage with it - and, if given agency as mentioned earlier herein, help maintain the community with you. This can be further strengthened if you allow your users to have influence on decisions made - for instance on what data will be released next.

How to maintain your community

There are many ways to maintain a community and every community is different, so be ready to learn as you go along (or skip some steps and look up good literature on the subject). A few well-proven tips include:

Be a community member yourself - empower members to lead with you

As mentioned earlier, having a human face and being available goes a long way, and seeing yourself as a community member will often inspire other members to contribute a great lengths. You might even want to consider setting up a structure in which key members get certain responsibilities; event organizing, responding to tech support requests or act as ambassadors to

name a few.

Be fair and transparent

People are likely more eager to join a community if they know more about the organization and - in this case - the provider of the data. Make sure to run the community in a way where a certain degree of transparency allows members to know what's coming and not feel as if you are suddenly moving in new directions (ie. around technical decisions affecting the data catalog) without informing the community.

Update your data sets/keep the stream going

When the newsworthiness of your released datasets fades, make sure to have new sets ready for release. If your data source is seen as a steady source of data, your community will automatically help attract others.

Highlight successful reuse cases

A good way to inspire your community to engage further is to highlight how your data has successfully been used by other members. You might consider setting up some kind of showcase of apps, projects and social change that builds on your data.