Whitepaper

Open data, content and software at Akvo

Akvo Foundation is a non-profit foundation that builds open source internet and mobile phone software which is used to make international development cooperation and aid activity more effective and transparent. We provide the software as a service, backed by a partner-support and training team.

The content and data that our partners use our services to collect are available under an open license. We often get questions about this. Quite simply, our partners and potential partners don't understand what open source, open content and open data is in general and specifically how it is implemented at Akvo. In this whitepaper we explain the what, why and how of open source software, open content and open data, and how we implement it at Akvo. We also discuss privacy and security around data hosted at Akvo.

The short version: Default is open

In general Akvo releases everything under open source, open content and open data licenses. Exceptions are made when the data potentially violates the privacy of the individual or household. Then it is only published in aggregated or anonymised form. Data that may expose the Akvo systems and services to security breaches is closed and kept private.

Feel free to share our data, content and software. If you are non-profit, then it shouldn't be a hard decision to make. If you are a for-profit operation, look at the detail of the licenses. If in doubt, please ask, we will probably be able to help.

What does open mean?

Before delving into the why and how, it is useful to understand the concepts and terminology involved. We'll use Wikipedia's definitions.

Open source software – "Open-source software is computer software with its source code made available and licensed with an open-source license in which the copyright holder provides the rights to study, change and distribute the software for

See it happen

akvo.org

Akvo Foundation
's-Gravenhekje 1A
1011 TG Amsterdam
The Netherlands

Email: info@akvo.org
Web: www.akvo.org
Tel: +31 (0) 20 820 01 75

free to anyone and for any purpose. Open-source software is very often developed in a public, collaborative manner." [1]

Open content – Open content "is licensed in a manner that provides users with the right to make more kinds of uses than those normally permitted under the law - at no cost to the user." [2]

Open content is often considered open if it allows the use of the content in different ways, again from the Wikipedia:

> Reuse - the right to reuse the content in its unaltered / verbatim form (e.g., make a backup copy of the content)
> Revise - the right to adapt, adjust, modify, or alter the content itself (e.g., translate the content into another language)
> Remix - the right to combine the original or revised content with other content to create something new (e.g., incorporate the content into a mashup)
> Redistribute - the right to share copies of the original content, your revisions, or your remixes with others (e.g., give a copy of the content to a friend)

Open data – "Open data is the idea that certain data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control." [3]

Going open, huge benefits

> "Fixing poverty in the world and providing information technology governance tools in this context, is too important to be dominated by proprietary software companies."
>
> Thomas Bjelkeman-Pettersson, Co-founder, Akvo

At Akvo we believe that going the open route provides substantial benefits for organisations in the international development sector and more importantly, give benefits to those that need it the most. Those we are trying to help.

So what are these benefits? Here are some of them:

Collaboration saves money – A report by the Standish Group states that adoption of open-source software models has resulted in savings of about $60 billion per year

akvo.org

Akvo Foundation
's-Gravenhekje 1A
1011 TG Amsterdam
The Netherlands

Email: info@akvo.org
Web: www.akvo.org
Tel: + 31 (0) 20 820 01 75

to consumers. [4] An example how open models can save money is that a lot of international development cooperation partners perform a baseline study before a project is started and this data is rarely shared in its raw format. This means that sometimes hundreds of data collection efforts happen in the same country, in related subjects and nobody is sharing. If we share this data, then substantial sums of money and time will be saved.

No costly lock-in effect – The principle behind our work is that the organisations and governments that use our software should be able to do so without suffering crippling lock-in effects, i.e. be able to decide to move away from using Akvo as a service provider without having to pay prohibitive licensing costs. Essentially the door is open. You can leave at any time and be free to take the software and data with you. No need to pay hundreds of thousands or millions to a software vendor for this privilege.

You keep control – We believe that in the future, information technology infrastructure, like the tools we build, will be just as important for the governance and operations of your country as your other infrastructure, like roads, electricity distribution or sanitation infrastructure. And we believe that the most successfully governed countries have local ownership of the infrastructure. You don't want every road outside your door to be a toll road, owned by a private company. The same way you don't want your governance IT infrastructure to be a toll road in the future. You need to be able to have control of it, and we think that only by basing your systems on open source code do you have that option.

You can reuse the tools – If you want to use the software for something entirely different, maybe something that isn't in the focus of Akvo, you are free to do so, providing you follow the open source software license we have chosen. We think this presents excellent opportunities for others to expand on our work.

You can reuse the content and data – There are going to be many positive ways of using the content and data our partners produce, in ways we never predicted. We don't want our imagination to be the limit of the usefulness, so together we open up the content and data. A very simple example could be, if you have data about public water points and the type of wells, someone researching an outbreak of cholera could very well correlate the incidences of the disease with type of water well.

We share back – Nearly all of the software that we create is built on top of other open source software. This is our way of contributing back to the community. Together we are building a global public good [5].

See it happen

akvo.org

Akvo Foundation
's-Gravenhekje 1A
1011 TG Amsterdam
The Netherlands

Email: info@akvo.org
Web: www.akvo.org
Tel: +31 (0) 20 820 01 75

As a summary we can say that there are many ways that open source, software, content and data will save money, increase efficiency and create new insight. The potential benefits are too many to describe in this document, but a good reference is the Open Data Handbook by the Open Knowledge Foundation: http://opendatahandbook.org – See also specifically the page Why Open Data? [6]

Common questions

Naturally, a lot of people think about the implications of open licensing and the consequences for their work. We will discuss a few points that generally come up during these conversations.

Why open data should be licensed?

It might sound strange to license open data. The whole idea is to make data available, right? It might sound contradictory, but to ensure that open data is really open, licensing is sometimes necessary. This is because many jurisdictions automatically confer certain rights to information and data that is published. By attaching a license to the information you actually ensure that your open data is treated in the way you intended, as you overrule any unwanted limitations that legal jurisdictions automatically impose upon your data.

Many open data efforts actually implement some type of restrictions, such as copyright law through the use of licenses. In some countries data is automatically under copyright as soon as it is produced and you can't copy this data without express permission. So at Akvo we think that the legal framework around copyright and data should be used as a support for our efforts rather than making the legal rights under which you can use the data uncertain. So we believe that it is better to have a good open license for the data than to have no license.

Why does Akvo use so many different licenses?

Different licenses are created for different purposes. You can't meaningfully share open source software using the Creative Commons licenses, they are just not suitable for the purpose. To share software we use software licenses like AGPL. In the same way, we actually use different licenses for databases and content. We'll explain below our rational behind the choice of licenses.

akvo.org

Akvo Foundation
's-Gravenhekje 1A
1011 TG Amsterdam
The Netherlands

Email: info@akvo.org
Web: www.akvo.org
Tel: +31 (0) 20 820 01 75

Most of the software we create is released under the AGPL license, but sometimes we provide bug fixes or improvements to other open source software and then we release our software fix or improvement under the open license that this software is normally released under.

Privacy

The key information to know about privacy of the data in the Akvo systems is that data that would compromise the security of the systems or would violate the privacy of individuals is not open. Data that could violate privacy is likely to be made available in aggregate or anonymised form, where the individual or the household cannot be identified.

There are a lot of misconceptions around open data and privacy. If we say that a system it built to support open data then many incorrectly assumes that all data in the system is open. But that is not the case, as this would be an unworkable system.

Lets use Wikipedia as an example. (Even if Wikipedia is more of a content management system than a data system. More people are familiar with how it works, which makes it easier to explain.) Not all data or content in the Wikipedia is open, even if the majority is. An example of non-open data in Wikipedia is your login data. Your password is data and it is not accessible to anyone but the system itself and thereby not open.

The same goes for Akvo's systems. In Akvo RSR there are a number of settings, like how a partner organisation's RSR site looks like and behaves, which are data, which isn't public data. The results of the settings may be public, for example what URL should be shown for the page, but the data itself is not open and can't be accessed via any public means.

Akvo FLOW, our mobile phone based field data collection tool has publicly accessible data from surveys performed. Today you can see survey data on a public map, example in this blog post [7], and in the future it will be possible to retrieve this data in an automated way, via something called an Application Programming Interface or API.

See it happen

akvo.org

Akvo Foundation
's-Gravenhekje 1A
1011 TG Amsterdam
The Netherlands

Email: info@akvo.org
Web: www.akvo.org
Tel: +31 (0) 20 820 01 75

With Akvo RSR you can already access almost all of the public data via the Akvo RSR API [8]. But you can't access the private data. Private data in Akvo RSR is, for example:

> username of a user account
> password for a user account
> email address for a user account
> administrative data for maintaining the system
> project data for projects that have not been published yet

Every piece of data that can be viewed on a public Akvo system web page will be available through the Akvo APIs. Essentially this information is already open, but not easily available unless we expose it through an API, which is what we do.

Data that should be private, like household survey data in Akvo FLOW will not be made public in such a way that the individual or the individual household can be identified. Anything that would violate the privacy of individuals will be kept private. Private data is likely to be made available in aggregate or anonymised form, where the individual or the household cannot be identified. In fact one of the key objectives with Akvo FLOW is to publish data that is collected with our systems, as openly as possible, so that the benefits from data reuse can be as wide as possible.

Data security

In summary one can say that Akvo takes all reasonable precautions against computer systems security breaches. But it is unrealistic to expect our systems to withstand a skilled and dedicated security penetration attempt. Therefore we don't handle "unsafe" data.

Akvo uses well regarded systems, toolkits and frameworks to safeguard the private data in our systems, such as the web development framework Django or the Wordpress content management system. However, we should not be under any illusions. Well-documented cases of data security breaches exist. These happen all the time in systems maintained by banks, credit card companies, the US Army, the FBI and Pentagon. The list goes on. If these organisations, which spend millions on security for their data systems, cannot keep the bad guys out of their systems, we should not hold any illusions that we can either.

As a consequence of this, Akvo has as a policy to not work on "unsafe" data. We do not work on data that need very high levels of security protection or shouldn't be

akvo.org

Akvo Foundation
's-Gravenhekje 1A
1011 TG Amsterdam
The Netherlands

Email: info@akvo.org
Web: www.akvo.org
Tel: +31 (0) 20 820 01 75

published for some reason. Some data that one could collect with our systems could require high levels of security protection. One could imagine collecting data about refugee camps near to a war zone. If we have detailed data about, for example, where all the war victims are camped one could conceivably get access to this data to find the victims and commit further atrocities.

That said, of course Akvo takes all reasonable precautions against system security breaches and we work diligently with improving the security protection for our systems and services.

There is a lot more to read about open source software, open content and open data. A good starting point is the referenced articles in Wikipedia which can be found at the end of this paper.

How Akvo does it

At Akvo we apply the principles of open source software, open content and open data in a specific way. This section describes the How.

Essentially all data, content and software produced by Akvo or available on Akvo's websites and our supporting services sites is protected by copyright and belong to Akvo Foundation or third parties. However, nearly all of this data, content and software is available for your use under a number of open licenses.

Trademarks

In short, don't use the registered Akvo® trademark as if you sell a product or operate a service under this name, without our permission. For trademark restrictions, please refer to the Akvo Foundation's Terms of use [9]. Do note that Akvo® is a registered trademark of Akvo Foundation.

Software

Akvo specifically makes available any software that our collaboration partners or we create, under the open source software license called: GNU Affero General Public License 3.0, or generally called the AGPL 3.0 license [10]. In essence this license allows you to download, use, modify and redistribute the software.

See it happen

The one key difference between the AGPL license and the General Public License (GPL) license, which much of the Linux operating system is licensed under for example, is that if you use AGPL licensed software to run a web or mobile service, which is what all the Akvo tools are, then you have to share any improvements or changes you make to the software available under the same license.

Content

We share content under a few licenses. As a rule we use the Creative Commons licenses, but there could be exceptions in particular circumstances where we pick another open content license.

Akvopedia – Content within the Akvopedia is available under the Creative Commons Attribution-ShareAlike 3.0 Unported License (CC-BY-SA) [11], just like content in Wikipedia.

General Akvo website content – All content created by the Akvo Foundation and displayed on the Akvo.org website and any other websites which we operate is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 Netherlands License. (CC-BY-NC-SA Netherlands) [12]. Unless otherwise stated. Occasionally we release videos licensed CC-BY [13] for example. In other words, such a video can be used for any purpose, in a commercial news broadcast for example, and it only needs to be attributed to Akvo Foundation.

User-generated content on the Akvo websites – Content that is submitted to the Akvo services, by users and systems, i.e. photographs, updates and project descriptions and more, and displayed on the websites which we operate, is licensed under the Creative Commons Attribution-Noncommercial-Share Alike 3.0 Plus Netherlands License. (CC-BY-NC-SA Plus)[12], unless otherwise stated.

Data

Data on the Akvo RSR and Akvo FLOW services – Data that is submitted to the Akvo RSR and Akvo FLOW services by users and systems, i.e. other information than photographs, updates and project descriptions, is available from our partners and us under a dual license. Both licensed under the Creative Commons Attribution-Share Alike 3.0 License (CC-BY-SA) [11] and under the Open Database License 1.0 (ODbL) [14] (or later versions as applicable).

akvo.org

Akvo Foundation
's-Gravenhekje 1A
1011 TG Amsterdam
The Netherlands

Email: info@akvo.org
Web: www.akvo.org
Tel: +31 (0) 20 820 01 75

The nitty gritty

This section describes the detail around why we have chosen particular licenses and the particular impact we think this has.

AGPL for software

The GNU General Public License and its siblings are the most common licenses for open source software today. GPL 2.0/3.0 LGPL 2.0/3.0 in May 2013 represented more than 53% of all the open source software licenses used. [15] The sibling license we use, GNU Affero General Public License version 3.0 (AGPL 3.0), only represents 0.16% of the use. So why have we chosen this license?

A bit of background on open software licensing is needed to explain this. Many of the open source licenses, like GPL 2.0, were created when distribution of software was done via magnetic media, like tapes or floppy disks. And these licenses required that if you distributed the software you had to also share the source code of the software, any improvements and modifications, as well.

Then the internet and the web happened. And suddenly you could have widespread use of software, as web software, used through a web browser, without actually distributing the software itself. So many of the large internet services and application providers did this. Google, Yahoo, Facebook, Amazon, Apple and most other large web service operators have based their services on open source software, but are hardly sharing any of their modifications of the software. They don't do this as the software they modify and create on top of the open source software is considered their competitive advantage. They follow the letter of the law and how it applies to these licenses. Licenses that were created in a different age and a different software consumption paradigm. But they essentially break the spirit of the license and the reason they were created. This is called the "ASP loophole" or Application Service Provider loophole.

The GPL 3.0 license was initially drafted to address the ASP loophole. But many open source software advocates protested and eventually a somewhat narrower upgrade from 2.0 to 3.0 was done. The intention of closing the ASP loophole was moved into a separate license called AGPL 3.0. (The story is somewhat more complicated than that, but that is a story too long for this document. To understand it, start with reading the Wikipedia page about AGPL 3.0 [16]).

See it happen

akvo.org

Akvo Foundation
's-Gravenhekje 1A
1011 TG Amsterdam
The Netherlands

Email: info@akvo.org
Web: www.akvo.org
Tel: +31 (0) 20 820 01 75

The difference between GPL 3.0 and AGPL 3.0 is, according to the Free Software Foundation which manages the license texts:

"[AGPL 3.0] is a free software, copyleft license. Its terms effectively consist of the terms of GPLv3, with an additional paragraph in section 13 to allow users who interact with the licensed software over a network to receive the source for that program. We recommend that developers consider using the GNU AGPL for any software which will commonly be run over a network."

In other words, exactly what Akvo's tools are all about. The effect of this is that the AGPL license requires anyone who operates a service based on Akvo software, or who modifies and improves the software and then operates a service, to release the code under the same license and make the source code available.

Much of the subsidies that Akvo receives to develop software comes from budgets allocated towards eliminating poverty in the world. We think it is morally correct to release the software source code under an open license that requires others that use it, for whatever purpose, to have to contribute back improvements or changes. Most people who aren't familiar with the details of open source software licensing, in our experience, actually think that is what open source software is about. They don't know that large corporations reap huge benefits from open source software but often contribute little or nothing back to the community.

And we are not the only once that think this is the correct way of licensing our software. Water for People, who started developing Akvo FLOW before Akvo took it over, agreed with us on our choice of licensing, after we had explained it to them. They had always said they were going to open source FLOW, but didn't know what license to pick until we worked it out with them.

Additionally, in a community which is relatively small, but often not so good at collaborating around projects, having open source software were you have to contribute back any improvements or changes will lower the problems with several teams working on versions of the software and not sharing with each other.

Practically, this is also a competitive advantage for Akvo as an organisation. We are able to establish ourselves in a market where giant companies tread and not be squashed by them, as they in general don't want to work with software which uses the AGPL license. We consider this a good thing, as Akvo is working on a public good, which we need broad uptake on across the international development sector, and the companies we compete with only have one thing in mind and that is profit.

See it happen

akvo.org

Akvo Foundation
's-Gravenhekje 1A
1011 TG Amsterdam
The Netherlands

Email: info@akvo.org
Web: www.akvo.org
Tel: + 31 (0) 20 820 01 75

Akvo's main goal is to help eliminate world poverty. We think our paradigm is more in line with the times and more productive in the long run.

Creative Commons for content

We use the Creative Commons (CC) licenses for content. The use of the Creative Commons licenses have proliferated and it is today the dominating content licensing system. There are several Creative Commons licenses, which vary a bit depending on what you are trying to achieve. In Creative Commons lingo "BY" means Attribution, "SA" means Share Alike and "NC" means Non-Commercial.

There are six Creative Commons licenses. [17] We primarily use two of them and one occasionally.

The main Creative Commons licenses we use are:

CC-BY-SA – 'This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. This license is often compared to "copyleft" free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects."

CC-BY-NC-SA – "This license lets others remix, tweak, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms."

We more rarely use:

CC-BY – "This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation."

To learn more about the Creative Commons licenses we recommend you check out the website: http://creativecommons.org/licenses/

For all content that Akvo creates we generally license it under the CC-BY-NC-SA license. Content created by our partners, on the Akvo services, is also made available under the CC-BY-NC-SA license. This license stops commercial reuse of the content. An example to illustrate why we think this is necessary is: Imagine a

Akvo Foundation
's-Gravenhekje 1A
1011 TG Amsterdam
The Netherlands

Email: info@akvo.org
Web: www.akvo.org
Tel: +31 (0) 20 820 01 75

picture of a village community is uploaded to Akvo RSR to illustrate a town meeting. If this didn't have a Non-Commercial license a company could theoretically use the picture in their advertising. We don't think that would go down so well in our international development community. This is avoided by using the NC version of the license, i.e. Non-Commercial.

Some content we license under the CC-BY license, when we want commercial channels to feel more comfortable picking it up, like a TV channel.

Creative Commons and Open Database License for data

Databases and data are different animals than other content. First of all, different laws regulate the copyright of databases and data than what regulates other types of content. The differences are significant enough that new types of licenses have been created just for databases.

One of the biggest datasets in the world which is collaboratively created is the OpenStreetMap database. It has many thousands of contributors and is updated constantly. OpenStreetMap data was originally licensed under the CC-BY-SA license. But it was asserted that the Creative Commons license doesn't cover sharing data well, even by the Creative Commons team themselves. So the OpenStreetMap community moved all their data to a more suitable license, which is the Open Data base License (ODbL).

Another provider of big datasets, the World Bank, took a different approach and released a lot of their data under the CC-BY license, despite arguments that it doesn't cover datasets. Looking at what constitutes somewhat conflicting approaches, we took the advice of the founder of OpenStreetMap, who said:

"Hopefully other projects can start based on the years we have put in to this and license with the ODbL or perhaps dual-license with CC-BY-SA. With luck they will never have to know how much work it took to get here." [18]

So all data which is collected by our partners in the Akvo services, including Akvo RSR and Akvo FLOW, will be dual licensed under CC-BY-SA and ODbL.

Why not use CC-BY, like the World Bank? The World Bank has a different agenda than we have. We think that many of the organisations we work with would be uncomfortable if the data they painstakingly and expensively collected, to eliminate world poverty, was used for commercial purposes without sharing the result openly

See it happen

under the same premise. We also think that if it works for OpenStreetMap it will also work for the international development community.

References

[1] Wikipedia, Open-source software, retrieved 3 June 2013,
http://en.wikipedia.org/wiki/Open-source_software
[2] Wikipedia, Open content, retrieved 3 June 2013, http://en.wikipedia.org/wiki/
Open_content
[3] Wikipedia, Open data, retrieved 3 June 2013, http://en.wikipedia.org/wiki/Open_data
[4] Creating wealth with free software, Richard Rothwell, (2008)
http://www.freesoftwaremagazine.com/articles/creating_wealth_free_software
[5] Wikipedia, Public good, retrieved 13 June 2013,
http://en.wikipedia.org/wiki/Public_good
[6] Open Data Handbook, Open Knowledge Foundation, retrieved 3 June 2013, http://
opendatahandbook.org/en/why-open-data/index.html
[7] The FLOW must go on – doing field surveys when there's no power, 5 July 2013 by
Luuk Diphoorn, http://www.akvo.org/blog/?p=11015
[8] Akvo RSR API developer documentation,
https://github.com/akvo/akvo-rsr/wiki/Akvo-RSR-API-developer-documentation
[9] Akvo Foundation, Akvo web services general terms of use.
http://www.akvo.org/web/terms_of_use
[10] GNU Affero General Public License 3.0,
http://www.gnu.org/licenses/agpl.html
[11] Creative Commons Attribution-ShareAlike 3.0 Unported,
http://creativecommons.org/licenses/by-sa/3.0/
[12] Creative Commons Attribution-Noncommercial-Share Alike 3.0 Netherlands License,
http://creativecommons.org/licenses/by-nc-sa/3.0/nl/deed.en_US
[13] Creative Commons Attribution 3.0 unported,
http://creativecommons.org/licenses/by/3.0/
[14] Open Database License 1.0, http://opendatacommons.org/licenses/odbl/1.0/
[15] Top 20 Most Commonly Used Licenses in Open Source Projects,
http://osrc.blackducksoftware.com/data/licenses/ Accessed 30 May 2013.
[16] Wikipedia, Affero General Public License, Accessed 12 July 2013,
http://en.wikipedia.org/wiki/Affero_General_Public_License
[17] Creative Commons, About the licenses, Accessed 12 July 2013,
http://creativecommons.org/licenses/
[18] O'Reilly, Choosing the right license for open data, http://strata.oreilly.com/2011/06/
openstreetmap-creative-commons-open-database-license.html