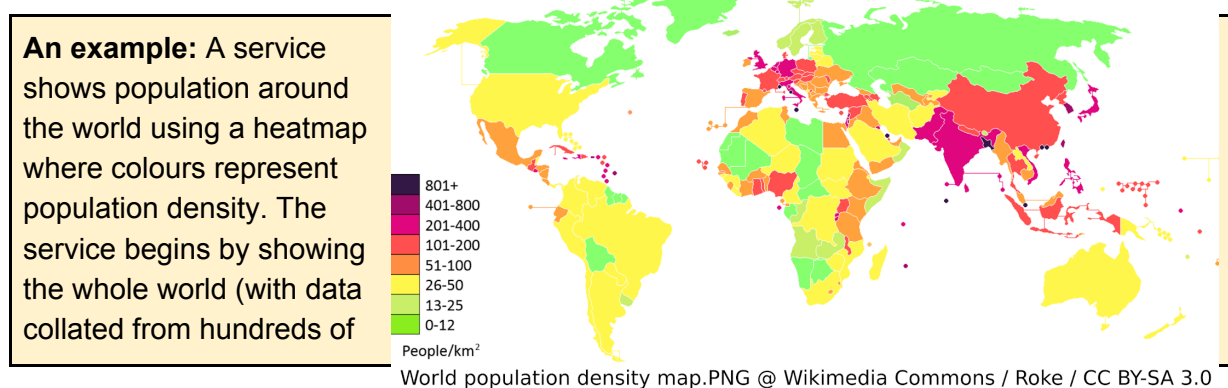# The issues with requiring attribution for reuse of data

Choosing the licensing for data may appear difficult at first. Many organisations initially explore requiring attribution for their data which is often a requirement used for texts and media files. It is sometimes thought that requiring attribution (sometimes with exemptions for specific groups) will lead to increased visibility for their organisation and would not have any direct adverse impact on reusability of the data. However, in order to facilitate data reuse it is necessary to use the most free license as possible.

Best practices for handling Public Sector Information (PSI) data recommends authorities to "make information available online and allow free re-use under standardised and generous terms".[1] This allows authorities to implement the policy objectives on increased reuse of public information. At the same time this makes it easier for authorities to obey the requirements of the PSI Directive.

An attribution requirement makes it difficult to include data sets in many kinds of services and products. Services often combine large amounts of data which are then calculated in different ways, requiring long lists of sources for each datapoint.

The attribution requirement also makes it difficult for an end user to know when they have done enough to comply with the terms of the license. Any uncertainty makes it more difficult to reuse the data as a misinterpretation now may mean having to completely rebuild the service at a later stage. This will often mean that services choose not to use data sets with unclear licenses or which require attribution or contain other limitations. This is especially common among smaller actors.

For the data providers the reduced usability of their data paradoxically leads to the organisation getting **less visibility** and the value of their own work is reduced as the information could have been included in multiple products.

**An example:** A service shows population around the world using a heatmap where colours represent population density. The service begins by showing the whole world (with data collated from hundreds of



World population density map.PNG @ Wikimedia Commons / Roke / CC BY-SA 3.0

---

[1] E-delegationen. *2. Sammanfattning och checklista för att komma igång.*
https://web.archive.org/web/20130317045612/http://feedback.edelegationen.se/vlpsi/vidareutnyttjande-a
v-offentlig-information/2-sammanfattning-och-checklista-fo%CC%88r-att-komma-iga%CC%8Ang/

different datasets and additional geodata from thousands of other sources). Collating data from many different sources allows the website to provide a complete picture, however it is very difficult to attribute thousands of sources. Should all data sources have a credit directly on the screen when the whole world is visible? Or should the source only appears when zooming down the country or region the data is for? What should the attribution look like? When would a data provider consider that their attribution requirement has been fulfilled and how do different attribution requirements interact with each other?

All of these are issues that the service creators need to deal with when releasing the product if attribution is demanded. Requiring attribution sounds a lot easier than it is to actually implement.

Wikidata is the Wikimedia project for structured data. Wikidata will evolve to become the central source of accurate data for all of the 294 language versions of Wikipedia, and its 500 million monthly readers. The Wikimedia movement practices what we preach and all of the data on Wikidata is available under the CC0 Public Domain dedication. Any datasets with an attribution requirement can therefore **not** be included in Wikidata.

# What license to choose?

**CC0 Public Domain dedication** is a legal tool that has been designed specifically for making data in databases available without restrictions on re-use. It is recommended by institutions such as the European Commission.[2] The aim of the CC0 license is to facilitate reuse of the data as much as possible and thereby stimulate the development of new products for all sizes of organisations.

CC0 is suitable for organisations that want the information they produce to be used by as many people as possible and who want to keep the public informed with new and up to date information.

CC0 means that there is no requirement to describe where the information comes from,[3] which can be viewed as negative for data owners. The question that the data owners must ask themselves is what is most important: the use and benefit to society or the visibility of the organisation? Is it important that publicly funded material is easily accessible and can be used by different actors in society?

---

[2] "[T]he CC0 public domain dedication is of particular interest. As a legal tool that allows waiving copyright and database rights on PSI, it ensures full flexibility for re-users and reduces the complications associated with handling numerous licences, with possibly conflicting provisions.", quote from Timothy Vollmer, Creative Commons. *European Commission endorses CC licenses as best practice for public sector content and data*. 17 July 2014. https://blog.creativecommons.org/2014/07/17/european-commission-endorses-cc-licenses-as-best-practice-for-public-sector-content-and-data/

[3] Read more here: https://creativecommons.org/publicdomain/zero/1.0/deed.en

It is important to note that CC0 does not prevent users from attributing the data anyway. In many cases attribution will be made even though CC0 was chosen, as many users recognize the value in emphasising that the data comes from a trusted source. E.g. on Wikipedia and Wikidata the source should always be indicated even if it is not a requirement of the data owner.



*This is how it may look on Wikidata when a source is added to a statement. The goal is that all statements should have a source, even if all the data is under CC0.*

# Organisations using the CC0 Public Domain dedication

CC0 is used by an increasing number of authorities across the world. A list of examples of organisations from around the world that use CC0 is available here: https://wiki.creativecommons.org/wiki/CC0_use_for_data

Some Swedish and European examples are:

- **European Union (EU)** – Since 2019, the European Commission uses CC0 as the standard license for metadata, raw data and other similar documents.[4] In connection with this decision, they produced a study on implementation tools for data re-use and the considerations that should be taken when licensing data.[5]

---

[4] European Commission. *Commission decision of 22.2.2019 adopting Creative Commons as an open licence under the European Commission's reuse policy*. 22 February 2019. http://ec.europa.eu/transparency/regdoc/rep/3/2019/EN/C-2019-1655-F1-EN-MAIN-PART-1.PDF
[5] European Commission. *Reuse policy: a study on available reuse implementing instruments and licensing considerations*. 2019.

- **Swedish International Development Cooperation Agency (SIDA)** – For SIDA the choice was only ever between licensing its data with the Public Domain Mark (PDM) or CC0. In the end SIDA decided to license all of the data on Openaid.se with the PDM rather than CC0 "since we as a government agency do not have copyright on our open data".[6]

- **National Library of Sweden (KB)** – Parts of the national union catalog Libris – the authority database and the Swedish National Bibliography – are licensed as CC0.[7] The National Library is actively working in the area and has concluded agreements on CC0 licensing with other parties.[8]

- **Swedish National Heritage Board (RAÄ)** – All metadata in K-samsök is under the CC0 Public Domain dedication.[9] This does not only apply to RAÄ but also all agencies and organisations who wish to share their metadata on the platform are required to release the data under CC0.[10] RAÄ argues that it is important to label the data as "although the metadata generally is not protected by copyright because it lacks originality, the database itself may be protected; the CC0 label is a way to make it clear that the free material is indeed free for people to reuse."[11]

- **Royal Armoury, Skokloster Castle, the Hallwyl Museum (LSH)** – One of Sweden's so-called central museums[12] that has chosen CC0 for their collection databases.[13]

- **Örebro Municipality** – All map data and geodata produced by the municipality is licensed as CC0.[14]

https://ec.europa.eu/jrc/en/publication/reuse-policy-study-available-reuse-implementing-instruments-and-licensing-considerations

[6] Joachim Beijmo, SIDA. *Licenser för data på Openaid.s*e. 19 March 2015. http://openaid.se/sv/blog/licensing-for-the-openaid-se-data/

[7] Öppna data från Kungliga biblioteket. https://data.kb.se/

[8] Maria Kadesjö, National Library of Sweden. *KB tecknar avtal med OCLC om CC0-licens*. 13 February 2014. https://web.archive.org/web/20160426161253/http://librisbloggen.kb.se/2014/02/13/kb-tecknar-avtal-med-oclc-om-cc0-licens/

[9] Swedish National Heritage Board. *Om K-samsök*. 22 August 2017. https://www.raa.se/hitta-information/k-samsok/om-k-samsok/

[10] Johan Carlström, RAÄ. *Samverkansavtal/överenskommelse för leverans till K-samsök*. 20 March 2013. http://www.ksamsok.se/wp-content/uploads/2013/04/K-sams%C3%B6k-avtal.pdf p. 4.

[11] Riksantikvarieämbetet. *Riktlinjer för digitalisering av kulturarvsmaterial på Riksantikvarieämbetet*. September 2013. http://www.raa.se/app/uploads/2013/10/digitaliseringsriktlinjer_riksantikvarieambetet.pdf, p. 17.

[12] Haxpett et al., Wikipedia. *Centralmuseum*. 8 January 2016. https://sv.wikipedia.org/w/index.php?title=Centralmuseum&oldid=32360788

[13] Karin Nilsson. *Öppet bildarkiv med hjälp av Vinnova*. Digitaliseringsbloggen. 17 June 2012. http://digitaliseringsbloggen.lsh.se/2012/06/projekt-oppet-bildarkiv-med-hjalp-av.html

[14] Örebro kommun. *Kartdata och geodata*. 9 November 2017. https://www.orebro.se/fordjupning/fordjupning/fakta-statistik-priser--utmarkelser/information-tillganglig-for-ateranvandning/kartdata--geodata.html