

Localisation: Legal Adaptation

Many good practices for publishing open data are the same all over the world. But the legal context in which publishers make open data available depends on the jurisdiction in which they operate in three ways:

1. **Rights** - different countries confer different rights on data creators
2. **Licensing** - different countries may have created their own licences, particularly for government data
3. **Privacy** - different countries have different laws and good practices around privacy

To localise the Open Data Certificate, you need to adjust the questions in the Legal section of the questionnaire to these local conditions.

This guide outlines the questions that are asked in the Legal section and highlights areas that are likely to require changes for a particular jurisdiction.

***Note:** This summary does not include all the questions in the full questionnaire, nor does it include specific help and requirement text. These are given in the [full configuration files](#) for each jurisdiction.*

1. Rights

This **sub**-section has two goals:

- to ensure that the publisher has the right to publish the data
- to encourage them to list the sources of the data

Suggested Localisation

Only the first of these goals requires localisation. If the publisher is unsure of if they have the right to publish (probably because they are not familiar with IP law), the questions try to work out whether the data they are publishing contains third-party rights. It assumes that whoever creates data owns it, and that extracting or calculating other data from it is unlawful unless a licence has been given to do so.

In some countries, there may be laws that permit certain types of processing of data without a licence; for example, text and data-mining. If this is the case, the questions should be adjusted to reassure publishers that they can publish data that has been derived in that way.

Default Questions

`publisherRights` **Do you have the rights to publish this data as open data?**

- yes *yes, you have the rights to publish this data as open data (standard)*
- no *no, you don't have the rights to publish this data as open data*
- unsure *you're not sure if you have the rights to publish this data as open data*
- complicated *the rights in this data are complicated or unclear*

-> If you answer no you cannot get a certificate.

-> If you answer complicated to publisherRights you are asked:

rightsRiskAssessment **Where do you detail the risks people might encounter if they use this data? (pilot)**

-> If you answer yes or unsure to publisherRights you are asked:

publisherOrigin **Was all this data originally created or gathered by you?**

-> If you answer no to publisherOrigin and you answered unsure to publisherRights you are asked:

thirdPartyOrigin **Was some of this data extracted or calculated from other data?**

-> if yes you are asked:

thirdPartyOpen **Are all sources of this data already published as open data? (raw)**

crowdsourced **Was some of this data crowdsourced?**

-> if yes you are asked:

crowdsourcedContent **Did contributors to your data use their judgement?**

-> if yes you are asked:

claUrl **Where is the Contributor Licence Agreement (CLA)? (raw)**

cldsRecorded **Have all contributors agreed to the Contributor Licence Agreement (CLA)? (raw)**

-> If you answered no to publisherOrigin you are asked:

sourceDocumentationUrl **Where do you describe sources of this data?** (*pilot*)

-> If a URL is provided for sourceDocumentationUrl you are asked:

sourceDocumentationMetadata **Is documentation about the sources of this data also in machine-readable format?** (*standard*)

2. Licensing

The goals of this **sub**-section are:

- to ensure that proper permission has been given to reuse and republish the data, such that it is in fact open data according to the Open Definition
- to encourage the publication of a machine-readable rights statement that provides machine-readable information about that permission

Suggested Localisation

There are three kinds of localisation that may be applicable to this section:

1. adjust the questions to cover the rights that publishers might have over the data they are publishing; if there are database rights, for example, then those need to be licensed as well as copyright
2. add any jurisdiction-specific or popular open licences to the selection lists; for example in the UK we add the UK Open Government Licence as this is used to licence much public sector material
3. if a "not applicable" answer is likely when asking about licences (for example because there are rarely rights that apply to data, or because government data is automatically public domain), the questions could be restructured or reordered to make it easier to select those options

Default Questions

copyrightURL **Where have you published the rights statement for this data?** (*pilot*)

In Europe (and other countries that confer database rights on the creators of databases), the following questions about data licensing ask about database rights as well as copyright in the data; in other places they only ask about copyright in the data.

dataLicence **Under which licence can people reuse this data?** (*raw*)

- odc-by *Open Data Commons Attribution License*
- odc-odbl *Open Data Commons Open Database License (ODbL)*

- odc-pddl *Open Data Commons Public Domain Dedication and Licence (PDDL)*
- cc-zero *Creative Commons CCZero*
- na *Not applicable*
- other *Other...*

-> If dataLicence is na then you are asked:

dataNotApplicable **Why doesn't a licence apply to this data?**

- norights *there are no copyright [or database rights] in this data*
- expired *copyright [and database rights] have expired*
- waived *copyright [and database rights] have been waived*

dataWaiver **Which waiver do you use to waive rights in the data?**

- pddl *Open Data Commons Public Domain Dedication and Licence (PDDL)*
- cc0 *Creative Commons CCZero*
- other *Other...*

-> If dataWaiver is other then you are asked:

dataOtherWaiver **Where is the waiver for the database rights?(raw)**

-> If dataLicence is other then you are asked:

otherDataLicenceName **What's the name of the licence? (raw)**

otherDataLicenceURL **Where is the licence? (raw)**

otherDataLicenceOpen **Is the licence an open licence? (raw)**

-> If the answer is is no you cannot get a certificate.

contentRights **Is there any copyright in the content of this data?**

- norights *no, the data only contains facts and numbers*
- samerights *yes, and the rights are all held by the same person or organisation*
- mixedrights *yes, and the rights are held by different people or organisations*

-> If contentRights is norights then you are asked:

Is the content of the data marked as public domain? (standard)

-> If contentRights is samerights then you are asked:

contentLicence **Under which licence can others reuse content?**

- cc-by *Creative Commons Attribution*
- cc-by-sa *Creative Commons Attribution Share-Alike*
- cc-zero *Creative Commons CCZero*
- na *Not applicable*
- other *Other...*

-> If contentLicence is na then you are asked:

contentNotApplicable **Why doesn't a licence apply to this content?**

- norights *there is no copyright in this data*
- expired *copyright has expired*
- waived *copyright has been waived*

-> If contentNotApplicable is waived then you are asked:

contentWaiver **Which waiver do you use to waive copyright?**

- cc0 *Creative Commons CCZero*
- other *Other...*

-> If contentWaiver is other then you are asked:

contentOtherWaiver **Where is the waiver for the copyright?(raw)**

-> If contentLicence is other then you are asked:

otherContentLicenceName **What's the name of the licence? (raw)**

otherContentLicenceURL **Where is the licence? (raw)**

otherContentLicenceOpen **Is the licence an open licence? (raw)**

-> If the answer is is no you cannot get a certificate. (is this true?)

-> If contentRights is mixedrights then you are asked:

Where are the rights and licensing of the content explained? (raw)

-> Finally, if a URL is provided for copyrightURL you are asked:

copyrightStatementMetadata **Does your rights statement include machine-readable versions of**

- dataLicense *data licence (standard)*
- contentLicense *content licence (standard)*

- attribution *attribution text (standard)*
- attributionURL *attribution URL (standard)*
- copyrightNotice *copyright notice or statement (expert)*
- copyrightYear *copyright year (expert)*
- copyrightHolder *copyright holder (expert)*

In Europe and other countries with a database right, there are also the options:

- databaseRightYear *database right year (expert)*
- databaseRightHolder *database right holder (expert)*

3. Privacy

The goals of this **sub**-section are:

- to ensure that the publisher has assessed the possible risk of releasing the data to releasing personal details; this might be a legal or a reputational risk
- to ensure that they have engaged third parties in auditing privacy-related matters
- to ensure that they have provided sufficient information to reusers to enable them to comply with the law

Suggested Localisation

Different countries have very different laws around privacy. The questions here are fairly universal, but there may be more specific legal requirements in individual countries. There might also be specific guidance that can be pointed to; for example in the UK, the Information Commissioner's Office has defined a process for Privacy Impact Assessments, so the questionnaire points specifically to those.

Default Questions

dataPersonal **Can individuals be identified from this data?**

- not-personal *no, the data is not about people or their activities*
- summarised *no, the data has been anonymised by aggregating individuals into groups, so they can't be distinguished from other people in the group*
- individual *yes, there is a risk that individuals be identified, for example by third parties with access to extra information*

-> If dataPersonal is summarised then you are asked:

statisticalAnonAudited **Has your anonymisation process been independently audited?(standard)**

-> If dataPersonal is individual then you are asked:

appliedAnon Have you attempted to reduce or remove the possibility of individuals being identified?

-> If **appliedAnon** is no then you are asked:

lawfulDisclosure Are you required or permitted by law to publish this data about individuals? (pilot)

-> If no, **you cannot get a certificate. (is this true?)**

-> If yes then you are asked:

lawfulDisclosureURL Where do you document your right to publish data about individuals? (standard)

-> If **appliedAnon** is yes OR **lawfulDisclosure** is yes then you are asked:

riskAssessmentExists Have you assessed the risks of disclosing personal data? (pilot)

-> If yes then you are asked:

riskAssessmentUrl Where is your risk assessment published? (standard)

-> If a URL is provided for **riskAssessmentURL** then you are asked:

riskAssessmentAudited Has your risk assessment been independently audited?(standard)

anonymisationAudited Has your anonymisation approach been independently audited? (standard)

In Europe, which has data protection laws that mandate certain handling of personal data, you are also asked:

individualConsentURL Where is the privacy notice for individuals affected by your data?(pilot)

dpStaff Is there someone in your organisation who is responsible for data protection?

-> If yes then you are asked:

dbStaffConsulted **Have you involved them in the risk assessment process?** *(pilot)*