

# Publishing Open Data

---

## *Devon County Council user guide*

### What is open data?

*“Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike.”*

- *Availability and Access: the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.*
- *Re-use and Redistribution: the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets.*
- *Universal Participation: everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavour or against persons or groups. For example, ‘non-commercial’ restrictions that would prevent ‘commercial’ use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.”*

<http://opendatahandbook.org/guide/en/what-is-open-data/>

Our strategy is to publish wherever possible, to help others get the information they need for whatever reason. As long as the data doesn't include personal or sensitive information we will consider it suitable for publishing by default.

There are risks involved in publishing open data, but we manage the risks in three ways:

1. Checking datasets to see if they are suitable
2. Providing guidance to our staff
3. Creating support and training from core staff who are keeping up with developments in new technologies and legislation.

This document is for data owners or project managers who need to publish open data as part of their work.

### Process

The process is only five stages long:

**IDENTIFY -> DECIDE -> PREPARE -> PUBLISH -> MAINTAIN**

... and it will help you if you remember to:

- Start with the end in mind – plan for the full life of the released data
- Think of all the people that you'll have to involve and talk to them early and often

### Identify

The first step to releasing open data is, naturally enough, identifying the data that is being considered for release. The nature of this data – and the decision to release it in some form – should trigger decisions about how it will be prepared for release, published and maintained.

Datasets can be suggested as candidates for publishing by

- The service that uses it deciding it would be good to share
- A FOIR being received for a dataset or related information
- Legislation proscribing a new dataset we must publish

- A partner organisation asking us to share information

Whoever collects and manages the dataset, that person becomes the Data Owner.

## Decide

The Data Owner will need to check three important points:

1. Is the data ours? Did we create it or collect it through our own activity, or does it come from another organisation we deal with? Note that if it's not our data, we are not the people who should be publishing it.  
Examples: postcode and address files, population statistics, birth rates
2. Does the dataset include personal or sensitive information?  
We cannot publish a dataset that includes people's names, addresses, personal contact details, ages, dates of birth, sexual orientations, medical details or use of our services in such a way that they could be identified and linked to the other information.  
There are exceptions to this – e.g. councillors' home addresses and contact details are published on our website as a standard, so they can be contacted by people living in their ward who want their help. If in doubt, always ask for help.  
We also cannot publish a dataset where some of the information is included in a summary form but the numbers in some of the summary groups are so small that they could be identified by another person e.g. through a combination of facts like gender, ethnicity and postcode. This is known as the MOSAIC effect.
3. Is the data up to date, accurate and well structured?  
Sometimes people actively want historical data, or will settle for whatever we can give them. As a rule though, we should aim to provide good quality data as current and as detailed as we can manage.

## Prepare

Preparing data for publishing means making sure it is ready to use, especially if you think people may want to use automated processing and so will need it to be good quality data that can be read and used by automatic processes. The main steps are:

1. Redact personal or sensitive data
2. Make sure, if the data groups people in categories, there are no groups smaller than 6 people
3. Check the data for typos, spelling mistakes, duplicates and empty rows. We can use a free tool called Open Refine to make this easier, especially with large datasets.
4. If the file is going to be published as comma separated variables (CSV) format, use the free CSVLint tool to make sure it's correctly structured. Just because you save something as CSV does not mean it will be in the correct structure, and bad file structures will cause problems for any programmes that call on the data.
5. Make sure than column headings are meaningful and relate to the data the column contains e.g. "Date of visit" is far more helpful to the end user than "DATVSofficeuseonlyEC302201"

## Publish

When the dataset is ready to publish, put a copy into the internal open data folder and let your open data lead know this has been done. They will need to check some details with you, so be ready to answer these questions:

- What time period does the data cover?
- How often does the data change?
- When will it be updated next?
- Who should be contacted about updates or any problems with the current file?

The open data lead can check your file for obvious errors, and refer it back to you if anything needs to be done, but it is your job to make sure it has been properly prepared and is ready to be published.

The open data lead will upload the dataset to the Devon Datastore, create an Open Data Certificate to go with it, put a link on the Devon County Council open data web pages, and add a link to the national catalogue at [data.gov.uk](http://data.gov.uk).

## Maintain

Getting your data identified, made safe, prepared and published can feel like quite an undertaking – but you can't rest there. As the data owner, you should already know how often this data needs to be updated; is it a weekly scrape of web hits, a monthly extract from an internal system, an annual collection of statistics from a survey? Whatever the answer, that's how often you should aim to publish an update.

The biggest question about maintenance is not "How often?" but rather "Whose job is it?". Make sure that you've answered this question as early in the process as possible, and that whoever is expected to update the data is prepared and equipped to do that.

Updating can mean creating a whole new file each time – one per year, say – or refreshing the existing file with a new column for the current month. If your data is very fast-moving you might want to consider providing an application programme interface (API) so developers can access it directly from the programmes they create rather than having to download a new file every time it changes.

## Summary

It can feel like there's a lot to do and a lot to remember when you're thinking about publishing open data; but most of it is common sense, and there's plenty of guidance available from the Open Data Institute, legislation such as the Transparency Code, and experts within the council. If you're in any doubt about whether you can or should share some specific data, ask for advice.

### Information Governance

<list names>

### Open Data Group

<list names>

### Digital Comms team

<list names>

## Useful links

The Open Data Institute

<http://theodi.org>

The Transparency Code

<https://www.gov.uk/government/publications/local-government-transparency-code-2015>

Open data certificates

<https://certificates.theodi.org/>

Link to Open Refine

Link to CSVLint