**How useful is my data?**

Assessing how useful open data is can vary depending on the domain and the content. To assist this process, there are a number of best practice guidelines publishers and users can follow.

In this module we look at the 5 Stars of linked open data and discover how this can be used to measure the technical usability of data.

In this module we’ll explore the following:

What are the 5 Stars of linked open data

The first 3 stars

How to recognise the stars in data

# Assessing data usability - in 2 minutes

ODI Trainer David Tarrant introduces the 5 Stars of open data as an assessment tool and helps you understand how and where it can be used.

# The 5 Stars of open data

The 5 Stars guideline for linked open data is a way to measure how well data is integrated into the Web.

It examines the accessibility and technical usability of a dataset ranging from being available online (1 star) to being part of the web of data (5 stars). Each star must be awarded sequentially and none can be skipped.

The guidelines were developed by Sir Tim Berners-Lee in 2001 and have been adopted by publishers worldwide to help guide many open data initiatives, such as the Italian Digital Agency who attach badges to all their datasets.

[Take me to the Italian Digital Agency site](http://spcdata.digitpa.gov.it/data.html)

In this module we look at the first 3 stars, which tabular and transactional data is suited to. The final 2 stars are explored in [module 13](http://accelerate.theodi.org/en/module13/).

# The first 3 stars

The first 3 stars of linked open data allow you to establish if the available data is usable.

## 1 star - An open licence

The first star is awarded to any data that is open at a basic level. The content, in any format, must be available under an open licence.

Regardless of the quality of a dataset, it cannot qualify for the first star unless it is available under an open licence.

A PDF file on a website available under an open licence is enough to fulfil the first star.

## 2 start - Re-usable format

The second star is awarded to any data that allows for simple re-use. The guidelines state that the data must be available in a ‘highly-reusable, structured format’ that can be read by a machine and understood by a human.

The key to achieving 2 stars is to select the most re-usable format. In some cases, the most re-usable format may be a closed or proprietary one, such as an Excel or Numbers file.

Making data available in any format is better than none at all.

## 3 start - Open format

The guidelines for the third star of open data state that the data must be available in a structured, machinereadable format which is not tied to a specific software package.

An example of a dataset that would be awarded 3 stars is a CSV file with an appropriate open licence.

Enel, an Italian energy company, provides a good example of a 3-star dataset with a clearly displayed starrating badge.

[Take me to Enel 3-star open dataset](http://data.enel.com/node/4679)

# Unlocking the usability of open data

Each star gained represents an improvement in the usability of a dataset. Use the following simple checks to evaluate how many stars any dataset has.

## Star 1 - Checking the licence

Finding the licence for an open dataset can be one of the hardest steps.

An open licence is an explicit permission to use the data for both commercial and non-commercial purposes.

The metadata record about the data

Terms and conditions of the webpage or site

Click on the data file, there might be a popup

Check inside the data file itself

Websites feature licensing information in different ways, you may have to search a little on each site to find them.

## Star 2 - Checking usability

When you discover open data on the Web, there are several ways in which you can assess how usable it is. Ask yourself:

Do I think that this is the most user-friendly format in which to publish this particular data?

Does the file appear to be the same as the one the publisher uses?

Is there any other information that has been added by a human?

## Star 3 - Checking openness

Open formats are sometimes difficult to assess. The following questions will help you understand if the data is available in an open format:

Can I open the file in more than one programme on my computer and still see the full functionality of the file?

When I look up the file format (eg CSV) online, does it say I need to download a specific programme to open it?

CSV is the most popular open format and is easy to identify. CSV qualifies for 3 stars as long as the other 2 are achieved. The UK open data portal provides some good examples of 3-star CSV datasets that are clearly marked with both a licence and file format.

[Take me to the UK open data portal](https://data.gov.uk/dataset/spend-over-25-000-in-fco-services)

Many open data portals are also publishing Open Document Format (ODF) files. These files may appear perfect for open data, however it can be difficult to identify the tool that can open these files.

**Are you ready to assess openness?**

There are four simple things to remember when it comes to assessing openness. Can you recall them?

**How do you rate the openness of a dataset using the 5 Stars?**

**How would you go about assessing the openness of a dataset you found online?**

Look at each star separately and rate according to the highest achieved

Start with the first star and rate each one sequentially

Just guess

**That’s right!**

Each star must be awarded sequentially and none can be skipped.

**Are you sure?**

Each star must be awarded sequentially and none can be skipped.

**What is the second star?**

**The second star is awarded for...**

an open licence an open format a reuseable format

**That’s right!**

The second star is all about data re-usability and recommends that data should be available in its most reusable form regardless of whether or not the format is open or proprietary.

**Are you sure?**

The second star is all about data re-usability and recommends that data should be available in its most reusable form regardless of whether or not the format is open or proprietary.

# Assessing open data

Assessing how useful data is can vary widely depending on the domain and the specific content. To help with this process there are a number of guidelines, including the 5 Stars of linked open data.

The 5 Stars guideline for linked open data is a way to measure how well data is integrated into the Web.

Each star must be awarded sequentially and none can be skipped.

In this module we looked at the first 3 stars:

Open license

Reusable format

Open format

For more on the final 2 stars see module 13.

Next module

Main menu

# Module 10 – video

I’m Dave Tarrant, I’m the Trainer and Data Scientist here at the open data institute. So the five stars of linked open data is a guideline that measures how well data is integrated into the web. It examines the technical usefulness of a dataset ranging from it being available online under an open licence, one star to being part of the web of data, five stars. Achieving open data can be done with the first star. Beyond this there are different degrees of what makes open data useable. Three stars represent the lowest common denominator of usable open data. The first star simply says that the data must be available under an open licence, no matter what the format for example a .pdf file on a website. The second star says that the data must be available at a highly reusable, structured and machine readable format so for example an excel file. For the third star, this highly reusable, structured and machine readable format must not be tied to a particular software package. So the format itself, and the implementation must be open allowing people to implement new tools, new code, to read that format. The final two stars are not always appropriate and implementing these may sacrifice usability for technical excellence.