**Module 7 – Ulrich Atz – Finding sustainable sources of open data**

**Finding sustainable sources of open data**

I’m Ulrich, I’m the Start-Up Programme manager at the Open Data Institute.

**What do we mean by sustainable?**

Well, by sustainable open data, we mean at its core to answer the question, “how long will the data be around”, and we can think of two main factors that affect the sustainability of an open data release. The first one is the funding, and the second is the use and re-use.

By funding we mean of course how the data release is funded, so open data may be free, but we still have cost associated with it. For example, as a publisher you have to collect data, you have to clean the data. Or as a user, you have to have the necessary skills to be able to analyse and process the data.

In terms of use, the more a dataset is used, the higher the likelihood that it will be around because you have a good story to tell why the dataset is useful. On the flipside, if there are no clear use cases of the open dataset, it undermines the sustainability of an open data release.

**Why is sustainability important when finding open data?**

Sustainability for open data is important because it gives users the trust and the confidence that the dataset will be around in coming years. For example if you are an open data business, you want to make sure your investment is worth it.

**What indicators of sustainability should you look for?**

There’s not one indicator of sustainability for open data, but you can think of several questions. For example: how is the data release funded? What are the reasons that the dataset exists? What is the licence associated with the dataset? And, how is the dataset used?

There are several indicators for sustainability for open data, for example, the first one would be how is the dataset funded, who is paying for this specific data release? Another indicator is the primary reason the dataset exists; is it collected for a specific purpose, or is it collected as an exhaust product of an existing activity? Another indicator would be the licence associated with the open dataset; does the licence specify how long the data release will be available in the future? And lastly the use and re-use of the open dataset is a very good indicator of its sustainability. The more used the dataset is, the more impact it shows, the likelier it will be available in the future.