**Module 11 – Dave Tarrant – Turning data into diamonds**

**Turning data into diamonds**

I’m Dave Tarrant, I’m the Trainer and Data Scientist here at the Open Data institute.

**How can you prepare data to help make it most valuable?**

One dataset will usually not be enough in order to develop new knowledge. It will be necessary to combine it and enrich it with further data in order to derive the insight. One of the biggest problems is the number of errors that exist in data. In an age full of computers we still have extremely messy data, and it is critical to clean these errors before you start to develop a story, for instance.

Studies have found that actually up to 80% of the time can be used in preparing the data before you even get to telling a story and deriving this insight. Fortunately there are a number of fantastic tools out there which can ease this burden.

**How do we turn data into diamonds?**

So one of the first things to do with any dataset is to ensure it’s clean. This may involve fixing spelling errors, removing misplaced data or normalising terminology. With the dataset clean, it may be then necessary to enrich the data, to add other datasets to it, and could involve also normalising code – so finding out what that number means, or to take a list of say, company names, and find out the company number for instance that you can reference other sorts of data. And then, once you have this clean and enriched dataset you can really start to derive the insight.

**Which are the best tools and techniques to clean up your data?**

So fortunately there is a fantastic open-source and free tool to available to help you do this, called OpenRefine. OpenRefine is a power tool for dealing with messy data. Not only does it allow you to clean up the data and to export it into a number of different formats, but it also enables you to enrich the data in a number of clicks, so that you can add company data, location data, and other general knowledge data to your dataset, to really find the value and derive the insight.