

Authoring better metadata in spreadsheets

Last updated: 30 August 2018

Laura Huerta - lahuema@ebi.ac.uk

Spreadsheets play a significant role in researchers' workflows, especially for data entry and storage. This practical will introduce you to some recommendations for organising metadata in spreadsheets in a way that both humans and computer programs can read.

You can find attached the data describing a RNA sequencing (RNA-seq) experiment to study global transcriptional changes associated with resistance induced by *Trichoderma harzianum* T39 to downy mildew (caused by *Plasmopara vitcola*) in *Vitis vinifera* Pinot Noir leaves. Let's have a look first at the metadata:

1. What do you think about this way of recording the biological information of the experiment?
 - human read-able and/or machine read-able?
 - sortable?
 - consistent across samples?
2. Could you split the information provided under Description into different concepts?
3. Is there any biological information important for describing the samples that is missing?
4. What is the main variable in this experiment?

Let's think now on the relationship between samples and files:

5. How many samples were analysed in that experiment?
6. How many biological replicates were analysed per sample?
7. How many sequencing files were generated?
8. Do you think this experiment contains any technical replicates?