

Homework 5

Question 1 (100 points)

Due at 2/20/2015 11:59pm

Please use the OpenRefine introduced in class to clean the data. To work on the homework, you need to follow the steps below.

(1) Select at least **3** fields from your homework 1 dataset that require cleaning/transformation.

Ex: splitting one column with multiple attributes into multiple columns, formatting peoples' name as first name space last name, extracting the numbers from artwork dimensions etc.

(2) Use the OpenRefine to clean the data.

References:

[1] OpenRefine <http://openrefine.org>

Submission Instructions:

You need to write a **report** containing the following contents.

1. Briefly describe the selected fields (15 points)
2. Explain the cleaning operations performed for each field (50 points)
3. Show one example of the cleaning done for each field. Show the full record before and after cleaning. (15 points)

You also need to upload a **zip file** containing the following document (20 points)

1. The dataset after the cleaning. Please put the clean data in one file.