Recipe: Spaghetti à la Hivos

Recipe: Oxfam Greens on a Bed of Worldbank Data

Does Oxfam Novib really conquer poverty in the World’s poorest countries?

Showing Hivos’ partner network using IATI data

# Begin with the end in mind

Think of what you’d like to show, what kind of data do you need to make that possible and check if you have the right data.

In this case we want to show Oxfam Novib’s projects, and get an impression if they really work in the poorest countries of the world – based on data from the World Development Index.

# Get your ingredients (the data)

We need to check if Oxfam Novib includes recipient countries for each project in its IATI data set.

1. Go to the IATI Registry and download the latest Oxfam Novib IATI file.

*It’s always a good idea to open the file in a text editor (for instance Notepad++), to see what’s in it.* [*http://notepad-plus-plus.org/*](http://notepad-plus-plus.org/)

We look at the data to see if there are elements like below.

<recipient-country code="..." xml:lang="en">...</recipient-country>

The first IATI-activities don’t contain this information, but if we scroll further down every element contains the recipient-country element.

For a further analysis of the data, we’ll transform it to a table format (CSV) first.

1. Download EditiX free edition (<http://www.freexmleditorsite.com/download.html>) and install it.
2. Download XSLT transformations here: <https://github.com/IATI/IATI-XSLT> and unzip the library to a place where you can easily find it again.
3. Transform the file to CSV using XSLT:
   * Open EditiX
   * Open the Oxfam Novib .xml file
   * Choose XSLT  Transform using XSLT
   * In the box ‘XSLT document’ browse to the library you’ve just unzipped, and on to templates  csv. Now choose iati-activities-xml-to-csv.xsl
   * Choose a location for the resulting file. Make sure to set the File type to ‘All files’ and end your filename with “.csv”
   * Click “OK”

And then we’ll open the data set in OpenRefine (*a power tool for working with messy data*)

1. Download OpenRefine from <http://openrefine.org/download.html> and install it.
2. When you run Open Refine, it will open in your browser, and let you create a project by importing data from your computer. Open the .csv file you’ve just created in step 4 and click next.
   * Make sure you set the field "Character encoding" to UTF-8
   * Choose “Columns are separated by Commas (CSV)”
   * You will see a preview of the data. In the top-right corner, click “create project” to finish importing the data.
3. Now let’s analyse the data Oxfam Novib has included in their IATI file. You can easily do this using facets and filters.
   * At the top of each column you’ll find an arrow pointing downward that offers you a range of powerful options to analyse and edit the data in that column.
   * For instance: go to the column with recipient-countries
   * Choose Facet  Text facet
   * On the left hand side, you’ll now see a box showing all the different values in this column

Notice there are 54 different values. There are 9 blank cells: Remember the first activities you encountered in Notepad++? So the remaining 1639 activities have a value in the recipient-countries column.

Also notice there are different ‘non-country’ values included in this column: Global Level, European Union, ….

1. Looking through the data set we see a couple of things:
   * All the recipient-country & recipient-country-code entries contain a single value.
   * There are only implementing partners in the data set (no funding or extending partner organizations)
   * All activities have disbursements (except the first 9)
   * Quite a few projects have multiple budgets. All budget values are given in EUR.
   * 1639 activities contain sector-code information (classified using an organization specific classification). Only codes 1, 2, 3, 4, 5 and 10 are given.
   * In fact the first 9 activities are not ‘normal’ projects, but have a different hierarchy (1). If you look closer, you’ll see that 74 activities refer to these activities: look for the related-activities-ref field!
2. In this visualization, we want to show Oxfam Novib’s projects (per country) on a map. So in fact we only need the country data provided per project.
   * The 9 programs don’t have any recipient-country data attached and the underlying projects are included in the file, so we’ll ignore the ‘odd nine’ for this case.
   * We'll save all fields related to 'recipient-country' and will also save 'sector' so that we can adjust the visualization to show a specific sector.

# Shaping & cleaning your data

Now we're going to shape the data, using our “power tool” OpenRefine.

1. First we’re going to reorder the columns in the data set. Go to the first column (All), click on the arrow and choose ‘Edit columns’  ‘Reorder / remove columns’
   * Reorder or remove a few columns and click “OK”
   * Notice how in the top-left corner the Undo / Redo tab now shows (1)
   * When you click the tab, you’ll see all the ‘transformations’ you’ve made so far. Here you can easily undo and redo your steps.
   * Also notice the buttons “Extract” and “Apply”. Here you have access to the code that OpenRefine produced to execute your transformation. You can copy this code and use it in future projects.
2. Go back to how the original data was by clicking on “0. create project”. The columns you removed are back again. Now click the “Apply...” button, and copy and paste the code below to reorder the columns in the data set, keeping the columns we want to use & save as extra information.

[

{

"op": "core/column-reorder",

"description": "Reorder columns",

"columnNames": [

"iati-identifier",

"recipient-country\_codes",

"recipient-countries",

"hierarchy",

"titles",

"default-currency",

"disbursement",

"sector\_vocabularies",

"sector\_codes",

"sector\_percentages",

"related-activity\_refs"

]

}

]

Given the fact that every activity has only 1 value for recipient country (and recipient country code) there’s no further cleaning up to do. Our data set should be in good shape now.

1. Export your data from OpenRefine. Choose ‘Export’  ‘Comma separated value’.

# On to the poverty data!

We want to see if Oxfam Novib carries out its projects in the countries with the poorest people. So we’re looking for data on the Gross National Income per capita.

1. Browse to <http://data.worldbank.org> and look for the data on Gross National Income per capita.
2. Download the dataset and open it in Excel (or any other spreadsheet program of your choice).
3. The data set contains historical data, but in this case we’re looking for recent data. So remove all columns except the name of the country and the GNI values for 2010 - 2013.
4. When you scroll down, you’ll see that the data for 2012 is more complete than the data for 2013. Poverty is not something that changes from year to year, so we better take the 2012 data. Remove the columns for 2010, 2011 and 2013. We now have a concise data set for poverty.

**On to the visualization!**

In this recipe we’ll be using CartoDB to visualize Oxfam Novib’s projects, and we’ll relate them to data about the world’s poorest countries.   
(You need a CartoDB account in order to use this tool. If you haven’t got one, set it up now)

1. Upload the Oxfam Novib .csv file from step 12 to CartoDB.
2. Looking through the resulting table, you’ll notice that CartoDB has added a column ‘geo data’ containing only *null* values. This means that it hasn’t recognized any of the columns as geo data. It will need some help.
3. Go to the column ‘recipient\_countries’ and click on the downward arrow. Choose ‘Georeference’
4. You’ll be given an number of options to georeference the data. In this case choose ‘By street addresses’, and then under ‘Country where street address is located, if known’ select the column ‘recipient\_countries’. Choosing this option will mean that we’ll miss any field that doesn’t contain a real country name. But given the fact that we’re looking at the “broader picture” (all Oxfam Novib projects worldwide) that is fine for this visualization.
5. Now upload the poverty data set you’ve just created in step 16.

You’ll notice that this dataset is not recognized as geo-data either. Luckily CartoDB contains some handy ‘common datasets’ to fix the problem.

1. Go back to the overview of all your tables (click the back arrow in the top left corner)
2. Now choose ‘Common data’ and find ‘World borders’. Add it to your tables by clicking it.
3. Once it’s imported, go to the ‘options’ menu (top right), and select ‘Merge with table…’
4. In the following wizard, choose ‘Column Join’, then select ‘name’ in the left column and the column containing country names in your Worldbank data in the right column (in my case ‘field\_1’). CartoDB will produce a new table containing the merged data.

**All the data is there. Now we’re ready to produce some visuals…**

1. Go back to the main screen, choose visualizations, and click on ‘(+) Create new visualization’.
2. First, add a layer containing the poverty data. Then add a layer containing the project data and click “Create visualization”.
3. Give your visualization a name.
4. You’ll be taken to the data view, which is not so interesting, so click on ‘Map view’.
5. Have a look at the options on the right-hand side: change the layout of layer 2 so that it nicely visualizes the richer and the poorer countries. (choose Choropleth, and select a good color-range)
6. At the bottom of the menu you’ll also find options for layer 1: the Oxfam Novib projects. Open the wizard and choose ‘Cluster’. Adjust the colors as you prefer. If you stick to the recipe choose green ;-)

Congratulations!

You have just created your first map visualization!

