

# JOB DESCRIPTION

## Overview

Job Title: Scraping balance sheets of all 8,088 Italian municipalities, 1998-2014

Software: I prefer that you use Python, but I'm open to alternatives

Start date: Immediately

End date: Setting up the scraping job should be elementary. Most of the time will be spent checking whether the script captured all the data. I expect the job to be completed within a week–10 days max.

Preferred skills (not required): Knowledge of the Italian language

Deliverables: excel files or text files (if excel files would be impractical). More info on the deliverables throughout this document and at this end in the section “Deliverables”

## Introduction

The balance sheets are publicly available on the website of the Italian government. They are located in the same place and they have the same or very similar structure. I will give you a list of 8,088 codes, one for each municipality. The codes should be plugged in the URL to access the balance sheet of the municipality.

The URL has the following recurring structure:

[http://finanzalocale.interno.it/apps/floc.php/certificati/index/codice\\_ente/CODE/anno/YEAR/cod/4/md/0/tipo\\_modello/U](http://finanzalocale.interno.it/apps/floc.php/certificati/index/codice_ente/CODE/anno/YEAR/cod/4/md/0/tipo_modello/U)

You will have to replace two components in the URL:

- **CODE** with the identifying code that I will provide
- **YEAR** with the year of balance sheets – 1998 to 2014 (all years in between)

When you land on the first page of the balance sheet of one municipality in one year, you will see the following page:



Anno: **2000**
Ente selezionato: **ABANO TERME (PD)**

Tipo Ente: **COMUNE**
Codice Ente: **2050540010**

Estrazione dati al 13/05/2016 18:37:41

Quadri presenti nel modello	
Quadro 1	DATI GENERALI AL 31 DICEMBRE
Quadro 2	ENTRATE
Quadro 3	RIEPILOGO GENERALE DELLE SPESE
Quadro 4	SPESE CORRENTI
Quadro 5	SPESE IN CONTO CAPITALE
Quadro 6	ANALISI DELLA SPESA PER TRASFERIMENTI
Quadro 7	DETTAGLIO DELLE SPESE IN CONTO CAPITALE PER L'ACQUISIZIONE DI BENI IMMOBILI
Quadro 8	ACCENSIONE E RIMBORSO PRESTITI PER ENTE EROGATORE
Quadro 9	QUADRO RIASSUNTIVO DELLA GESTIONE FINANZIARIA
Quadro 10	DEBITI FUORI BILANCIO AL 31.12 CHE RISULTANO AI SOTTOSCRITTI FIRMATARI DEL CERTIFICATO/PIGNORAMENTI
Quadro 11	GESTIONE DEI RESIDUI ATTIVI
Quadro 12	GESTIONE DEI RESIDUI PASSIVI
Quadro 13	SERVIZI INDISPENSABILI per Comuni ed unioni di comuni
Quadro 14	SERVIZI A DOMANDA INDIVIDUALE per Comuni ed unioni di comuni
Quadro 15	SERVIZI DIVERSI per Comuni ed unioni di comuni
Quadro 16	INFORMAZIONI INTEGRATIVE DEI SERVIZI INDISPENSABILI per Comuni ed unioni di comuni
Quadro 17	INFORMAZIONI INTEGRATIVE DEI SERVIZI A DOMANDA INDIVIDUALE per Comuni ed unioni di comuni
Quadro 18	INFORMAZIONI INTEGRATIVE DEI SERVIZI DIVERSI per Comuni ed unioni di comuni
Quadro 19	INFORMAZIONI INTEGRATIVE DEI SERVIZI DIVERSI per Comuni ed unioni di comuni
Quadro 50	CERTIFICAZIONE DEI PARAMETRI OBIETTIVI PER I COMUNI AI FINI DELL'ACCERTAMENTO DELLA CONDIZIONE DI ENTE STRUTTURALMENTE DEFICITARIO PER IL TRIENNIO 1998-2000

The blue box lists the different panels (quadri in Italian) that are available for this municipality and this year. The available panels are 1 to 19 and 50. Not all panels, however, are available for each municipality in every year.

The URL to access each panel has a recurring structure

[http://finanzalocale.interno.it/apps/floc.php/certificati/index/codice\\_ente/\*\*CODE\*\*/cod/4/anno/\*\*YEAR\*\*/md/0/cod\\_modello/CCOU/tipo\\_modello/U/cod\\_quadro/\*\*PANEL\*\*](http://finanzalocale.interno.it/apps/floc.php/certificati/index/codice_ente/CODE/cod/4/anno/YEAR/md/0/cod_modello/CCOU/tipo_modello/U/cod_quadro/PANEL)

where **PANEL** can be 01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 50

The scraping script will have to download the information contained in each panel and save it into various excel files. The next sections contain additional information on each panel.

## Info on each panel

### Panel 1



Anno: **2000**    Ente selezionato: **ABANO TERME (PD)**  
Tipo Ente: **COMUNE**    Codice Ente: **2050540010**  
Estrazione dati al 13/05/2016 18:42:10

### QUADRO 1 - DATI GENERALI AL 31 DICEMBRE (gli importi sono espressi in euro)

Notizie varie	
VOCI	Dati
Popolazione residente (ab.)	18.536
Nuclei familiari (n)	6.999
Circoscrizioni (n)	N.C.
Frazioni geografiche	1
Superficie totale del Comune (ha)	2.175
Superficie urbana (ha)	800
Lunghezza delle strade esterne (Km)	56
Di cui: in territorio montano (Km)	N.C.
Lunghezza delle strade interne centro abitato (Km)	49
Di cui: in territorio montano (Km)	N.C.

Assetto del territorio e problemi dell'ambiente	
VOCI	Dati
Piano regolatore approvato	S
Piano regolatore adottato	N
Programma di fabbricazione	N
Programma pluriennale di attuazione	N
Piano edilizia economica e popolare	S
Piano per gli insediamenti produttivi	N
_ industriali	N
_ artigianali	N
_ commerciali	N
Piano delle attività commerciali	S
Piano urbano del traffico	S
Piano energetico ambientale comunale	N

Quadri  
:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 50

This panel is usually composed by two tables. Do you see how the tables have the same column names (VOCI and Dati)? They can be reported in the same excel or text file. I would like to create for each municipality and year, an excel file named panel\_1\_ **CODE\_YEAR**.xlsx (replacing code and year with the relevant values) with the following columns:

code	year	table_name	voci	dati
2050540010	2000	Notizie varie	Popolazione residente (ab.)	18.536
2050540010	2000	Notizie varie	Nuclei familiari (n)	6.999
2050540010	2000	Notizie varie	Circoscrizioni (n)	N.C.
2050540010	2000	Notizie varie	Frazioni geografiche	1
2050540010	2000	Notizie varie	Superficie totale del Comune (ha)	2.175
2050540010	2000	Notizie varie	Superficie urbana (ha)	800
all the other rows in this table ....				
2050540010	2000	Assetto del territorio e problemi dell'ambiente	Piano regolatore approvato	S
2050540010	2000	Assetto del territorio e problemi dell'ambiente	Piano regolatore adottato	N
2050540010	2000	Assetto del territorio e problemi dell'ambiente	Programma di fabbricazione	N
All the other rows in this table ....				

As an example, I populated the table with the values from the picture:

- code and year: I would always like you to insert these variables to identify the source of the data
- table\_name: here I would like to put the name of the tables to identify what is the nature of the data
- voci and dati: after the first three columns, just copy the info that you find in the webpage (in this case, the two columns are names voci and dati)

## Panel 2

This panel is composed by different tables with the same three columns. Repeat the procedure that you followed for panel 1:

- scrape the data in one big table for each municipality and year
- put the code and year at the beginning
- put the table name in the third column+
- put the actual data in the remaining three columns and name these variables after the column name in the tables (which should be the same across all different tables)

## Panel 3

Usually, it is composed by only one table. It should be straightforward.

## Panel 4

Anno: 2000	Ente selezionato: ABANO TERME (PD)
Tipo Ente: COMUNE	Codice Ente: 2050540010
Estrazione dati al 13/05/2016 19:05:44	

**QUADRO 4 - (A) - IMPEGNI**  
(gli importi sono espressi in euro)

Funzioni e servizi / Interventi correnti	Personale	Acquisto di beni di consumo e/o di materie prime	Prestazioni di servizi	Utilizzo di beni di terzi	Trasferimenti	Interessi passivi e oneri finanziari diversi	Imposte e tasse	Oneri straordinari della gestione corrente	Ammortamenti di esercizio	Totale
Funzioni generali di amministrazione di gestione e di controllo. Di cui:	1.094.408,32	174.166,31	1.344.097,16	72.675,81	71.257,62	8.078,42	247.024,95	66.160,71	0,00	3.977.867,76
- Segreteria generale, personale e organizzazione	1.129.421,52	16.860,25	611.809,82	50.984,63	65.439,74	0,00	74.172,51	0,00	0,00	1.948.688,97
- Gestione economica, finanziaria, programmazione, provveditorato e controllo di gestione	272.288,99	64.157,89	142.700,66	0,00	0,00	8.078,42	98.500,21	19.292,25	0,00	605.020,48
- Gestione delle entrate tributarie e servizi fiscali	111.006,21	0,00	77.161,24	0,00	5.769,86	0,00	30.238,55	41.316,55	0,00	265.492,93
- Gestione dei beni demaniali e patrimoniali	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
- Ufficio tecnico	303.999,96	88.220,65	295.186,62	0,00	0,00	0,00	20.467,70	0,00	0,00	707.874,93
- Anagrafe, stato civile, elettorale, eva e servizio statistico	131.225,50	4.160,58	1.890,23	21.691,19	48,03	0,00	8.874,80	0,00	0,00	167.890,33
Funzioni relative alla giustizia	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Funzioni di polizia locale. Di cui:	565.594,67	14.999,97	18.207,17	0,00	0,00	0,00	42.076,26	0,00	0,00	640.878,60
- Polizia municipale	565.594,67	14.999,97	18.207,17	0,00	0,00	0,00	42.076,26	0,00	0,00	640.878,60
Funzioni di istruzione pubblica. Di cui:	84.467,56	67.382,65	638.478,10	0,00	194.622,65	67.414,67	5.416,60	0,00	0,00	1.057.783,78
- Scuola materna	0,00	0,00	32.640,59	0,00	124.955,20	0,00	0,00	0,00	0,00	157.596,31
- Istruzione elementare	0,00	20.081,91	295.410,76	0,00	24.273,47	0,00	15,49	0,00	0,00	339.782,16
- Istruzione media	0,00	0,00	87.747,06	0,00	12.730,66	0,00	0,00	0,00	0,00	100.477,72
- Refezione e altri servizi	84.467,56	47.300,22	222.679,17	0,00	32.662,80	67.414,67	5.401,11	0,00	0,00	499.927,08
Funzioni relative alla cultura ed ai beni culturali. Di cui:	276.574,03	42.294,72	527.652,14	0,00	50.610,19	77.379,19	23.137,27	0,00	0,00	997.648,57
- Biblioteche, musei e pinacoteche	192.454,56	41.530,88	174.928,60	0,00	9.812,68	77.379,19	15.075,38	0,00	0,00	511.182,32
Funzioni nel settore sportivo e ricreativo. Di cui:	21.505,78	7.221,10	183.847,81	0,00	119.601,09	30.015,96	1.667,64	0,00	0,00	363.860,41
- Piscine comunali	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
- Stadio comunale, palazzo dello sport ed altri impianti	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Funzioni nel campo turistico	28.832,24	3.597,12	238.162,03	1.797,27	116.309,19	0,00	5.061,28	0,00	0,00	393.759,65
Funzioni nel campo della viabilità e dei trasporti. Di cui:	90.165,11	18.540,80	394.754,35	0,00	0,00	291.267,75	5.909,82	0,00	0,00	1.000.638,34
- Viabilità, circolazione stradale e servizi connessi	90.165,11	9.244,58	225.511,42	0,00	0,00	291.267,75	5.909,82	0,00	0,00	622.099,19
- Illuminazione pubblica e servizi connessi	0,00	9.296,22	369.242,93	0,00	0,00	0,00	0,00	0,00	0,00	378.539,15
Funzioni riguardanti la gestione del territorio e dell'ambiente. Di cui:	269.877,13	51.998,66	3.657.874,68	6.271,34	7.410,84	349.769,40	249.958,43	0,00	0,00	4.593.118,21
- Urbanistica e gestione del territorio	0,00	0,00	21.040,45	0,00	0,00	0,00	0,00	0,00	0,00	21.040,45
- Locale e piani di edilizia economico-popolare	97.231,79	0,00	0,00	6.271,34	7.410,84	105.262,70	6.476,89	0,00	0,00	222.654,38
- Servizio idrico integrato	149.833,44	34.602,61	671.899,06	0,00	0,00	244.504,64	9.570,46	0,00	0,00	1.110.411,25
- Servizio smaltimento rifiuti	22.811,38	0,00	2.304.430,68	0,00	0,00	0,00	233.909,53	0,00	0,00	2.561.151,60
Funzioni nel settore sociale. Di cui:	482.178,62	48.482,91	539.096,30	0,00	471.415,66	13.915,93	38.858,73	0,00	0,00	1.593.949,19
- Asili nido, servizi per l'infanzia e per i minori	273.846,11	31.069,53	95.018,77	0,00	0,00	0,00	18.086,84	0,00	0,00	418.022,28
- Strutture residenziali e di ricovero per anziani	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
- Assistenza, beneficenza pubblica e servizi diversi alla persona	168.568,95	7.317,68	418.728,28	0,00	471.414,63	11.967,86	18.127,64	0,00	0,00	1.096.126,06
- Servizio necroscopico e cimiteriale	39.763,05	10.095,18	25.348,74	0,00	0,00	1.948,08	2.644,26	0,00	0,00	79.800,34
Funzioni nel campo dello sviluppo economico	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Funzioni relative a servizi produttivi	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
<b>TOTALE</b>	<b>3.013.603,47</b>	<b>428.644,25</b>	<b>7.742.169,74</b>	<b>80.744,42</b>	<b>1.031.227,05</b>	<b>837.841,31</b>	<b>619.110,97</b>	<b>66.160,71</b>	<b>0,00</b>	<b>14.619.504,51</b>

**QUADRO 4 - (B) - PAGAMENTI IN CONTO COMPETENZA**  
(gli importi sono espressi in euro)

Funzioni e servizi / Interventi correnti	Personale	Acquisto di beni di consumo e/o di materie prime	Prestazioni di servizi	Utilizzo di beni di terzi	Trasferimenti	Interessi passivi e oneri finanziari diversi	Imposte e tasse	Oneri straordinari della gestione corrente	Ammortamenti di esercizio	Totale
Funzioni generali di amministrazione di gestione e di controllo. Di cui:	1.437.568,62	111.883,67	929.165,87	72.675,81	63.907,93	8.078,42	186.018,48	33.542,84	0,00	2.842.338,55
- Segreteria generale, personale e organizzazione	607.622,90	12.000,39	431.271,47	50.984,63	62.019,76	0,00	43.711,88	0,00	0,00	1.207.611,54
- Gestione economica, finanziaria, programmazione, provveditorato e controllo di gestione	269.877,13	51.998,66	3.657.874,68	6.271,34	7.410,84	349.769,40	249.958,43	0,00	0,00	4.593.118,21
- Urbanistica e gestione del territorio	0,00	0,00	21.040,45	0,00	0,00	0,00	0,00	0,00	0,00	21.040,45
- Locale e piani di edilizia economico-popolare	97.231,79	0,00	0,00	6.271,34	7.410,84	105.262,70	6.476,89	0,00	0,00	222.654,38
- Servizio idrico integrato	149.833,44	34.602,61	671.899,06	0,00	0,00	244.504,64	9.570,46	0,00	0,00	1.110.411,25
- Servizio smaltimento rifiuti	22.811,38	0,00	2.304.430,68	0,00	0,00	0,00	233.909,53	0,00	0,00	2.561.151,60
Funzioni nel settore sociale. Di cui:	482.178,62	48.482,91	539.096,30	0,00	471.415,66	13.915,93	38.858,73	0,00	0,00	1.593.949,19
- Asili nido, servizi per l'infanzia e per i minori	273.846,11	31.069,53	95.018,77	0,00	0,00	0,00	18.086,84	0,00	0,00	418.022,28
- Strutture residenziali e di ricovero per anziani	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
- Assistenza, beneficenza pubblica e servizi diversi alla persona	168.568,95	7.317,68	418.728,28	0,00	471.414,63	11.967,86	18.127,64	0,00	0,00	1.096.126,06
- Servizio necroscopico e cimiteriale	39.763,05	10.095,18	25.348,74	0,00	0,00	1.948,08	2.644,26	0,00	0,00	79.800,34
Funzioni nel campo dello sviluppo economico	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Funzioni relative a servizi produttivi	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Panel 4 is composed by three big tables (only 2 shown in the picture) with the same columns. You can put the data in one excel file for each municipality and year. Make sure to put the table name in the third column, though! Notice how the table name in the panel is in bold and just above each table, instead of being embedded in the first row of the table as before? The table names are:

- name of table 1: Impegni
- name of table 2: Pagamenti in conto competenza
- name of table 3: Pagamenti in conto residui

## Panel 5

Exactly the same structure of panel 4. Repeat the same procedure.

## Panel 6

Panel 6 might have separate tables or only one. Usually, when there are extra tables, they have the same columns. In these case, repeat the same procedure followed in panel 4 and 5.

## Panel 7

Usually, it is composed by only one table. It should be straightforward.

## Panel 8

	Anno: 2000	Ente selezionato: ABANO TERME (PD)
	Tipo Ente: COMUNE	Codice Ente: 2050540010
Estrazione dati al 13/05/2016 19:12:32		

**QUADRO 8 - ACCENSIONE E RIMBORSO PRESTITI PER ENTE EROGATORE**  
(gli importi sono espressi in euro)

ENTI EROGATORI	CONSISTENZA AL 31/12/1999	ACCENSIONE ACCERTAMENTI	ACCENSIONE RISCOSSIONI C/COMP. + RESIDUI	RIMBORSO IMPEGNI	RIMBORSO PAGAMENTI C/COMP. + RESIDUI
Cassa DD.PP.	7.339.179,45	108.455,95	102.277,57	233.076,48	233.076,48
Istituti di previdenza amministrati dal Tesoro	0,00	0,00	0,00	0,00	0,00
Istituto per il credito sportivo	247.210,36	0,00	0,00	11.688,97	11.688,97
Aziende di credito	6.159.116,24	1.053.572,07	599.492,84	440.254,20	440.254,20
Istituti speciali di credito	129.736,04	0,00	0,00	54.261,03	54.261,03
Istituti di assicurazione	585.411,13	0,00	0,00	181.381,21	181.381,21
Stato (Tesoro)	0,00	0,00	0,00	0,00	0,00
Prestiti esteri	0,00	0,00	0,00	0,00	0,00
Prestiti obbligazionari	0,00	0,00	0,00	0,00	0,00
Altri finanziatori	0,00	0,00	0,00	0,00	0,00
Totale	14.460.653,21	1.162.028,02	701.770,41	920.661,89	920.661,89

**QUADRO 8 - BIS-CONSISTENZA DEI DEBITI**  
(gli importi sono espressi in euro)

DEBITI	CONSISTENZA Iniziale	CONSISTENZA Finale
Di finanziamento	14.460.653,73	14.702.019,35
Di funzionamento	3.937.787,60	4.698.427,91
Per iva	5.214,15	0,00
Per anticipazioni di cassa	0,00	0,00
Per somme anticipate da terzi	348.212,28	768.489,93
Verso imprese	0,00	0,00
Altri	137.125,50	237.748,35
Totale	18.888.993,27	20.406.685,53

Panel 8 is composed by two tables, but they have different columns. You should create two datafiles for each city and municipality with names panel\_8\_ **CODE\_YEAR**.xlsx and panel\_8bis\_ **CODE\_YEAR**.xlsx

There might be more than two tables! Just repeat the same procedure, creating one excel file for each table.

### **Panel 9**

Similar to panel 8, panel 9 might be composed by different tables with different columns. Create separate excel files.

### **Panel 10**

Similar to panel 8 and 9, Panel 10 is composed by two tables, but they have different columns. You should create separate datafiles.

### **Panel 11**

Usually, it is composed by only one table. It should be straightforward.

### **Panel 12**

Similar to panel 8, 9, and 10, panel 12 might be composed by different tables with different columns. Create separate excel files.

### **Panel 13**

This panel has the same structure of panel 1 and 2: different tables with the same columns and the name embedded in the first row of the table. Repeat the procedure used before.

### **Panel 14**

This panel has the same structure of panel 1, 2, and 13: different tables with the same columns and the name embedded in the first row of the table. Repeat the procedure used before.

### **Panel 15**

This panel has the same structure of panel 1, 2, 13, and 14: different tables with the same columns and the name embedded in the first row of the table. Repeat the procedure used before.

Starting in year 2007, this panel changes structure and becomes a series of tables with different columns. From 2007, create a separate excel file for each table with the following names: panel\_15s1\_**CODE\_YEAR**.xlsx, panel\_15s2\_**CODE\_YEAR**.xlsx, etc.

**Panel 16 (after 2007 might not be present)**

This panel has the same structure of panel 1, 2, 13, and 14: different tables with the same columns and the name embedded in the first row of the table. Repeat the procedure used before.

**Panel 17 (after 2007 might not be present)**

This panel has the same structure of panel 1, 2, 13, and 14: different tables with the same columns and the name embedded in the first row of the table. Repeat the procedure used before.

**Panel 18 (after 2007 might not be present)**

This panel has the same structure of panel 1, 2, 13, and 14: different tables with the same columns and the name embedded in the first row of the table. Repeat the procedure used before.

**Panel 19 (after 2007 might not be present)**

This panel has the same structure of panel 1, 2, 13, and 14: different tables with the same columns and the name embedded in the first row of the table. Repeat the procedure used before.

**Panel 50**

It might have more than one tables, but you should capture just the one titled “QUADRO 50 - PARAMETRI DEFICITARI”. It is usually the second.

**Final page →**



## **Deliverables**

I would like you to send me the excel files for each city and year in a zip file. Owing to the fact that the number of files will be very high, I would like you to structure the archive in the following way:

- different folders for each panel
- within the panel folders, different folders for each year
- and finally the excel files in the correct panel-year path

I would also like you to send me separate excel or text files for each panel, in which you include the observations for all municipalities and all years. For example, I would like to receive a file panel\_1 in which there are the rows of panel 1 for all municipalities and all years.

## **Attention**

The structure of panels might vary slightly between municipalities and years. Sometimes, the tables might have a different number of rows. Sometimes, a panel might have more or fewer tables. Sometimes, a panel might be missing altogether. Your code should be flexible enough to account for these circumstances.