An Approach to Characterize Open Government Data: A Case from Three Brazilian Public Datasets

The adoption of open government data encourages transparency, participatory governance, and accountability. Achieving these goals depends on the quality of the data. For example, outdated and uncompleted data may hide accurate analysis. In a previous study, we faced the challenge of manually analyzing Brazilian open government data quality. Manually selecting, cleaning, and transforming the raw data into an appropriate format for analysis requires a representative effort that is error-prone. In this document, we present the details of a semi-automatic approach, including the script for validation and the target metadata from which we characterize the Open Government Data portals.

1 Python Script and Metadata in JSON Format

```
1  {
2  def validarurl(par1):
3    try:
4        response = requests.get(par1)
5        if response.status_code == 200:
6            return 'url validada'
7        else:
8            return 'url inv lida!'
9    except:
10       print("Ocorreu um erro!")
11       return 'erro'
12
13 if kdoc == 'results':
14 for idoc in vdoc:
15     cont = cont + 1
16     print('-----')
17     print('Nr Recurso:', cont)
18 for k2doc, v2doc in idoc.items():
19 if (k2doc == 'description'):
20    if v2doc is not None:
21       if v2doc == '':
22          cont_description_branco
23          = cont_description_branco+1
24          print('Descricao em Branco')
25       else:
26          cont_description_valid =
27          cont_description_valid+1
28          print('description:'+v2doc)
29 else:
30  if (k2doc == 'url'):
31   print('url:'+v2doc)
32   inicio=v2doc[0:4]
33    #print(inicio)
34    if inicio=='http':
35      texturl=validarurl(v2doc)
36      cont_url_valid=
37      cont_url_valid+1
38      print(texturl)
39    else:
40      cont_imposs_valid=
41      cont_imposs_valid+1
42      print('nao foi possivel, validar
43      devido protocolo disponibilizado')
44    else:
45      if (k2doc == 'name'):
46          if v2doc is not None:
47              if v2doc == '':
48              cont_name_branco =
```

```
49            cont_name_branco + 1
50            print('Nome em branco')
51    else:
52            cont_name_valid =
53            cont_name_valid + 1
54            print('name:' + v2doc)
55    else:
56     if (k2doc == 'format'):
57         if v2doc is not None:
58            if v2doc == '':
59                cont_format_branco =
60                cont_format_branco + 1
61                print('formato em branco')
62    else:
63         cont_format_valid =
64         cont_format_valid + 1
65         print('format:' + v2doc)
66    else:
67         if (k2doc == 'created'):
68             if v2doc is not None:
69                if v2doc == '':
70                    cont_date_branco =
71                    cont_date_branco + 1
72                    print('data em branco')
73                else:
74                    cont_date_valid =
75                    cont_date_valid + 1
76                    print('created:' + v2doc)
77  }
```

Listing 1 – Python Script to Validate Metadata

```
1  {
2  {"help":"https://dados.gov.br/api/3/action/help_show?name=resource_search",
3  "sucess":true,
4  "results":{
5       "count": 19,
6       "results":{
7               {
8                   "mimetype":null,
9                   "cache_url":null,
10                  "state":"active",
11                  "hash":"",
12                  "description":"Recompra de CFT-E das entidades mantenedoras com
                        ades o ao FIES. De 2010 a 2022",
13                  "format":"Formul rio de Consulta",
14                  "url":"https://www.fnde.gov.br/olinda-ide/servico/Fies_Compra/versao
                        /v1/aplicacao#!/recursos/FiesCompra",
15                  "created":"2022-09-15T11:45:50.297907",
16                  "cache_last_updated": null,
17                  "package_id":"1eae9959-9304-466c-85ba-6ab68fa3325e",
18                  "mimetype_inner":null,
19                  "last_modified":null,
20                  "position":0,
21                  "revision_id":"c3d3b232-f5ea-49dc-b86d-63e1612a8078",
22                  "size":null,
23                  "url_type":null,
24                  "id":"cb770c50-ee8f-4be2-b75a-9033f9aa5236",
25                  "resouse_type":null,
26                  "name":"Recompra FIES"}
27  }
```

Listing 2 – Metadata from the FIES Repository

```
1  {
2  {"help":"https://dados.gov.br/api/3/action/help_show?name=resource_search",
3  "sucess":true,
4  "results":{
5       "count": 7,
6       "results":{
```

```
 7                    {
 8                         "mimetype":"text/csv",
 9                         "cache_url":null,
10                         "state":"active",
11                         "hash":"",
12                         "description":"",
13                         "format":"csv",
14                         "url":"http://www.geoservicos.inde.gov.br/geoserver/wms?service=WFS&
                              version=1.0.0&request=GetFeature&typeName=MPOG:MCMV&outputFormat
                              =CSV",
15                         "created":"2018-06-12T05:08:08.632076",
16                         "cache_last_updated": null,
17                         "package_id":"2e1a5ade-d1c7-45da-b953-c6e8a65f9e83",
18                         "mimetype_inner":null,
19                         "last_modified":null,
20                         "position":0,
21                         "revision_id":"5ea9baf0-dba2-4590-8a7a-510cad518072",
22                         "size":null,
23                         "url_type":null,
24                         "id":"f595714a-0356-40f2-aeae-bf7695a372e4",
25                         "resouse_type":null,
26                         "name":"Programa Minha Casa Minha Vida"}
27  }
```

Listing 3 – Metadata from the Minha Casa Minha Vida Repository

```
 1  {
 2  {"help":"https://dados.gov.br/api/3/action/help_show?name=resource_search",
 3  "sucess":true,
 4  "results":{
 5       "count": 384,
 6       "results":{
 7                    {
 8                         "mimetype":"application/zip",
 9                         "cache_url":"",
10                         "state":"active",
11                         "hash":"",
12                         "description":"Todas as tabelas da base de dados em abril de 2015",
13                         "format":"zip+csv",
14                         "url":"https://transparencia.gov.br/download-de-dados/historico/
                              historico-copa/201504",
15                         "datastore_active":false,
16                         "created":"2013-01-30T09:17:05.337191",
17                         "cache_last_updated": null,
18                         "package_id":"0018277e-675f-4a6e-ba97-13f4a1e9f9b9",
19                         "mimetype_inner":"application/zip",
20                         "last_modified":"2012-12-31T00:00:00",
21                         "position":0,
22                         "revision_id":"37b752ff-bc67-461b-b090-5ff232ee014a",
23                         "size":null, "url_type":null,
24                         "id":"19862cf6-e5df-44ca-8759-61d7c1d9f05d",
25                         "resouse_type":"example",
26                         "name":"Toda a base de abril de 2015"}
27  }
```

Listing 4 – Metadata from the Portal da Transparência Repository

2   Configuration file:

```
 1  URLS_REPOSITORIOS#
 2  [https://dados.gov.br/api/3/action/package_searchq=fies,
 3  https://dados.gov.br/api/3/action/package_searchq=mcmv,
 4  https://dados.gov.br/api/3/action/package_search?q=transparencia]
 5  TERMOS#[fies,mcmv,tranparencia]
```