Besides Standards and Automation: an Experience with Census Databases

Bogdan-Florin POPOVICI, Ph.D.

National Archives of Romania, Brașov County Division, Brașov, str. G. Barițiu nr. 34, 500025, Brașov, România
e-mail: bogdanpopovici@arhivelenationale.ro

Besides Standards and Automation: an Experience with Census Databases

ABSTRACT

This paper came out of a practical experience, that did not offer the proper prerequisites indicated usually by standards. Censored databases of censuses, with partial documentation, was transferred to the National Archives and remained stored with no attempt to immediately process or make them accessible to the users. When examined, many issues and questions arose, both technically and conceptually. We share our experience and our reluctances, being aware of the fact that more advanced colleagues (using SIARDS or ADDML, for instance) may regard all as these primitive solutions. But, in the same time, we are convinced we are not the only ones lacking proper skills and expertise and our examples may be an example for other colleagues.

Key words: *census databases, archival processing*

Fra standard ed automazione: un esperienza con i database censuari

SINTESI

Questo intervento è nato da un'esperienza pratica, che non ha fornito i corretti requisiti necessari indicati di solito dagli standard. Le banche dati di censimenti, con documentazione parziale, sono state trasferite all'Archivio Nazionale e conservate senza alcun tentativo di elaborarazione immediata o per renderli accessibili agli utenti. Quando sono stati esaminati, sono emersi numerose questioni e domande, sia tecniche che concettuali. Si viole qui condividere questa esperienza e le sue criticità pur consapevolinche i colleghi più esperti (che usano ad esempio SIARDS o ADDML) possono considerare primitive tutte queste soluzioni. Allo stesso tempo, però, si è convinti di non essere gli unici a non disporre di competenze ed esperienze adeguate, e che questi esempi possono essere utili ad altri colleghi.

Parole chiave: database censuari, trattamento archivistico

Poleg standardov in avtomatizacije: izkušnje s podatkovno bazo popisa prebivalstva

IZVLEČEK

Ta članek je nastal na osnovi praktičnih izkušenj, ki niso zagotavljale ustreznih predpogojev, ki jih običajno predpisujejo standardi. Cenzurirana podatkovna baza popisa prebivalstva z delno dokumentacijo, je bila prevzeta v državni arhiv, in je ostala shranjene brez namena pristopa k takojšnji obdelavi ali pripravi gradiva za njegovo uporabo. Ko smo pistopili k pregledu podatkovne baze, so se pojavila številna vprašanja, tako tehnična kot konceptualna. V prispevku predstavljamo in delimo svoje izkušnje, pri tem pa se zavedamo dejstva, da lahko bolj napredni kolegi (ki na primer uporabljajo SIARDS ali ADDML) vse naše rešitve obravnavajo kot primitivne. Toda prepričani smo, da nismo edini, ki nimamo ustreznih veščin in strokovnega znanja, zato so lahko naši primeri in izkušnje primer za druge kolege.

Ključne besede: podatkovne baze, popis prebivalstva, strokovna obdelava arhivskega gradiva

Dincolo de standarde și automatizare. O experiență cu bazele de date ale recensămintelor

REZUMAT

Prezentul material prezintă experiența prelucrării bazelor de date ale recensămintelor din 2002 și 2011. Pornind de la cadrul legal, el însuși problematic, articolul urmărește modul în care datele primite au devenit entități funcționale, inteligibile pentru utilizator. Sunt descrise de asemenea alte acțiuni conexe ale SJAN Brașov de completare a documentației recensămintelor, precum și dilemele și răspunsurile noastre raportat la arhivarea bazelor de date de acest tip.

Cuvinte-cheie: prelucrare arhivistică, baze de date, recensăminte

1 Introduction

In 2002 and 2011 in Romania there were undertaken general censuses of populations and housings. The National Archives, by its departmental units, were required to consider for permanent preservation the documentation of these censuses. A certain procedure was followed and electronic data were transferred to the National Archives.

In 2013, we discovered that, apart from their transfer (one may read "ingest"), the files were not submitted to any the further archival processing in our institution[1]. The accompanying documentation was rather poor and, moreover, one may consider the records transferred as being partial relevant for the census. In the following, we shall discuss the topic of appraisal census data, we shall examine the data received and the process we performed, with some consideration on archival description of databases.

2 Legal framework

Between 18 and 27 March 2002, Romania undertook a census for population and housings. According to the first regulation on the census, after the publication of final results, the whole series comprising the individual paper forms shall be transferred for preservation to the National Archives and its territorial branches[2]. Six months later, a new government decision was issued, indicating that primary information from the data form delivered by surveyed persons will be anonymized, wrote on magnetic carrier and they will be preserved by the National Archives or its territorial branches. After publishing the final results, the original forms shall be destroyed[3].

Between 7 and 16 May 2011, another census was performed, on the same topic. According to a Government Decision in 2009[4], after the publication of final results, the database containing individual data shall be transferred for preservation to the National Archives or its territorial branches. A new Government Decision from 2011, changed again the provision, indicating that after the publishing of final results, the database in electronic form, containing anonymised personal data shall be transferred for preservation to the National Archives, while the original forms shall be destroyed[5]. As an extra ingredient, due to the fact the 2011 census was declared "the last paper-based census", the National Archives asked for its branches to take over also the original paper form, as historical relevant documents. In our cases, that lead to a juridical litigation, since the local Directorate for Statistics insisted in strictly application of the Government Decision, that is, refusing to transfer the paper forms and asking for destroying

---

1. Brasov County Division of the National Archives.

2. *Hotărârea nr. 680 din 19 iulie 2001 privind organizarea și desfășurarea recensământului populației și al locuințelor din România în anul 2002*, în *Monitorul Oficial*, nr. 439 din 6 august 2001, art. 17.

3.. *Hotărârea nr. 1505 din 18 decembrie 2002 pentru modificarea Hotărârii Guvernului nr. 680/2001 privind organizarea și desfășurarea recensământului populației și al locuințelor din Romania în anul 2002*, în *Monitorul Oficial*, nr. 19 din 15 ianuarie 2003.

4. *Hotărârea Guvernului nr. 1.502/2009 privind organizarea și desfășurarea recensământului populației și al locuințelor din România în anul 2011*, în Monitorul Oficial 860 din 10 decembrie 2009, art. 16.

5. *Hotărârea Guvernului nr* 922 din 21.09.2011 pentru modificarea și completarea *Hotărârii Guvernului nr. 1.502/2009 privind organizarea și desfășurarea recensământului populației și al locuințelor din România în anul 2011* în *Monitorul Oficial*, nr.689/28.IX.2011, art. 13.

the paper originals, on ground of data protection. The local division of the National Archives refuses the consent to their disposition, since they were declared of historical value, so the forms are still kept for preservation in the creator records centre.

In both cases, the Brasov County Division of the National Archives received for permanent preservation optical supports (CDs) carrying primary data (*micro-data*) of the censuses, in an anonymized form. In the following, we shall examine the content received, its technical and historically usefulness and relevance and we shall describe the archival processing we performed on them. Apart from the transfer process, there were no special regulations or indications about the method of archival processing, so the following have the character of a study case, with its good, bad and ugly parts. Many of the steps and procedures may look naïve for those archivists having expertise in dealing with databases preservation. But due to circumstances, the lack of professional guidance and expertise, we consider this presentation may be of interest to other colleagues, in similar situations.

3 The 'objects'

The data of the 2002 census was received on a compact disk, containing 16 files in `*.dbf` format, accompanied by narrative description, providing: names of files, their content and fields coded names, types and possible values. There were also list of values, serving presumably as sources for main tables.

The data was not authenticated in any way, nor at hash/checksum level, nor digital signature.

The data of 2011 was delivered also as on a compact disk, in package digitally signed. The package could have been open using a proprietary software belonging the company that issued the digital certificate (basically, it was a digitally signed and encrypted file `*.p7s`). Inside the package, there were 40 files in `*.dbf` and `*.csv` format (based on the file name, the same data in two formats) and files in `*.pdf` format containing a scanned census form, that contained the mapping of the database fields codified name and with the field name in clear from the form.

The compact disks remained kept as such until 2016, when it was raised the question about their archival processing and possible use.

MAPA	ULC	NRPCPH	RT	GOSP	PERS	P05	P08	P09	P10	P11	P12	P13	P14	P15	P16
80003	1	4	4		1	1	18527	1	0		0	3	2	1	1982
80003	1	4	4		2	1	18527	1	0		0	3	2	1	1973
80003	1	4	4		3	1	18527	1	0		0	3	2	1	1974
80003	1	4	4		4	1	23387	1	0	0	0	4	2	1	1974
80003	1	4	4		5	1	114319	1	0	0	0	26	1	1	1961
80003	1	4	4		6	1	16329	1	0		0	3	1	1	1977
80003	1	4	4		7	1	18527	1	0		0	3	2	1	1978
80003	1	4	4		8	1	46180	1	0		0	10	2	1	1979
80003	1	4	4		9	1	123013	1	13	2	1996	13	1	1	1982
80003	1	4	4		10	1	20821	1	0	0	0	4	2	1	1959
80014	1	73	4		1	1	9	1	8	2	2001	8	2	2	1980
80014	1	73	4		2	1	25692	1	4	2	2001	4	2	2	1980
80014	1	73	4		3	1	63580	1	14	1	2001	14	1	2	1979
80014	1	73	4		4	1	108696	1	24	2	2001	24	2	2	1977
80022	1	116	4		1	1	9	1	0		0	90	0	2	1917
80022	1	116	4		2	1	9	1	0		0	90	0	2	1915
80022	1	116	4		3	1	179132	1	40	1	1994	40	1	2	1924
80022	1	116	4		4	1	9	1	15	1	1994	15	2	2	1937
80022	1	116	4		5	1	9	1	4	2	1976	4	2	2	1949
80032	1	29	4		1	1	130534	2	29	1	1995	29	1	1	1985
80032	1	29	4		2	1	20876	2	4	1	2000	4	1	1	1985
80032	1	29	4		3	1	145934	2	32	2	2000	32	2	1	1984
80032	1	29	4		4	1	175055	2	39	1	2000	39	1	1	1985
80032	1	29	4		5	1	75	2	0	0	0	75	0	1	1986
80032	1	29	4		6	1	83632	2	0	0	0	19	1	1	1985
80032	1	29	4		7	1	75	2	0	0	0	75	0	1	1982
80032	1	29	4		8	1	75	2	0	0	0	75	0	1	1982
80032	1	29	4		9	1	75	2	0	0	0	75	0	1	1982
80032	1	29	4		10	1	1455	2	0	0	0	1	1	1	1985

4 Archival processing. Making objects understandable

In their original form, we had very little information about those data. As can be seen in the picture 1, opening the data would render tables of raw data, with no meaning whatsoever. Relying on some database knowledge, one could suspect some of the figures in the column are, in fact, IDs from the links to other tables and not figures with a meaning by themselves. In this regard, for making the data in the main tables understandable, the decision was taken to attempt to link the tables, decoding the meaning of fields and data. The possibility to keep the files as flat ones was rejected, since it was obviously the usability for research was better served if there was a possibility to filter / cross-query the various data.

4.1 Census in 2002

The actions for making "functional" the tables from 2002 census faced some technical difficulties. Firstly, the tool chosen for converting tables to a more functional framework was MS Access 2013, as a DBMS better known by archivists and having rather complex functionalities. Unfortunately, this version of the software did not have the capability to read files in `*.dbf` format anymore, which require the use of another tool, at least for conversion. A free tool was then used, DBF Plus, that could read and export the tables from their original dbf files to a tab-separated text file. After converting all tables. They were imported into MS Access database.

P30	Short Text	legat new
P31	Number	legat nou
P32	Number	legat nou
P33	Number	
P34	Number	legat COR
P35	Number	legat new
P36	Number	legat CAEN
P37	Number	legat new
P38	Number	legat new
P39	Number	legat new
P40	Number	legat new
P41	Number	legat new
P42	Number	legat new
P43	Number	legat new
VARSTA	Number	
JDOM	Number	legat new
MDOM	Number	legat new
STAOC	Number	
VIRD	Number	

Field Properties

General Lookup

Field Size	Long Integer
New Values	Increment
Format	
Caption	
Indexed	Yes (No Duplicates)
Text Align	General

During the export, the header of the tables was not exported, nor the correct page code. That had to be fixed in Access, by manually create a new field header and by setting the correct code page to reproduce Romanian regional settings. The coded headers remained as it were in tables, while the explicit meaning of the field was only rendered in the form (Figure 2 and 3). Then, field by field was checked for possible links with the existent list of values. This operation was, in many aspects, a fortune-teller endeavour, since there was no clear description of the source and the destination of data. In order to keep track of our intervention, we kept the imported tables untouched and new list of values were clearly indicated (Figure 4). In the main table, every field linked was annotated as to indicate the source of data (Figure 2). In this way, a user of the Access database is aware which is the "original" (in fact, the migrated/ imported tables) and which is the tables refactored by the archivist.

Tables	⨂
▦ 05 J08S7	
▦ 10 FWPD08--pop stabila din locuinte	
▦ 20 FWPCD08---pop_stabila din locuinte in comun	
▦ 30 WS08--cetateni straini	
▦ 40 WLF08--locuinte	
▦ 50 WCL08--cladiri cu locuinte	
▦ 60 WCHD08--unitati de locuit in comun	
▦ 70 GOSP08--gospodarii locuinte	
▦ 99 CAEN	
▦ 99 cetatenie	
▦ 99 COR	
▦ 99 ETNIE	
▦ 99 LIMBA	
▦ 99 RELIGIE	
▦ 99 SCOLI	
▦ 99 siruta02	
▦ copy 0 FWPCD08---pop_stabila	
▦ Copy FINAL Of FWPCD08---pop_stabila	
▦ new_P05	
▦ new_P08	
▦ new_P09-P20-P41	
▦ new_P10-P13	
▦ new_P11-P14	
▦ new_P15	
▦ new_P19	
▦ new_P28-P30	
▦ new_P29-P31	
▦ new_P32-P42	
▦ new_P35-P43	
▦ new_P37	

Linking the tables was again a very time consuming task, because one table could have 50-60 fields. Another issue was that some of the list of values (LoV) were, in fact merged as values in the main table, but separated as source, so a new merged LoV had to be created; basically, we reconstructed the source of values, having little evidence this was the original indeed. As if all these actions were not troublesome enough, we discovered that some of the fields, described in the explanatory fields as being: "not used", contained in fact positive values, whose meaning remained obscured.

4.2 Census in 2011

Most of the issued encountered when converting tables containing microdata from 2002 census were the same for the 2011 one. The data were also contained files with `*.dbf` format, coded fields names, values representing data or just pointers to LoV data and so on. The mapping between paper forms and table fields was done by the creator by indicating on a digital copy of paper forms the coded names of fields. In the process of de-coding names, we were able to find out that not all the fields were transferred in the tables, but they were altered by removing (at least) one field containing National ID Number.

Another issue for 2011 census files was the size of the database: due to large amount of information, it grew larger than 2 GB, (the limit of `*.accdb` file), forcing us to use separate databases for various tables (inhabitants, housing etc.) and related entities.

After the successful import and creation of tables, it was generated a set of forms and queries, in order to ease the reading and searching the data (Figure 3). Since the paper forms are considered a class in the Directorate of Statistics filing plans, it comes natural to consider these two databases as series. They were described as such in our archival management software-scopeArchiv.

But this archival processing was only a part of the process. As seen, our reality was far from a standard and proper SIP, metadata were poor enough not to be align to any of the professional ones. Basically, we received a bucket of files, in various formats, we migrated them and created a tool for access to data. We felt the need though to clarify conceptually these entities, their status (original, copy, derivate for research etc.) and also their archival nature and value.

5 Debating issues

5.1 About received packages

The files we received were produced by the creator. We had absolutely no information about the way those `*.dbf` files were generated, if they were altered (as we suspected), if they were the result of an export or copy, nor if they were subject of any other transformation. We requested information from the local Directorate of Statistics and, unfortunately, due to the distance in time, we could collect information only about the system used for the 2011 census. Due to this request for information, it was revealed the application used was a dedicated one, created specifically for the census in 2011. It was designated to run on Window XP operating system and it was programmed in Visual Foxpro. `*.dbf` files were, therefore, native, copied for us and not migrated. The architecture was client-server, each county units consolidating the date to a central server. Also, the application had a native export tool, in `*csv` or `*.xls` format (the alternative files sent to the archives; again, the files transferred seemed to be exported directly from the application, without other processing). We received no information whatsoever about anonymization actions. Since the files for transfer were prepared at central level, it is likely this information was not available at county level.

The files received were supposed to remain unaltered, as a proof of what we received. In fact, at least for 2011 census files, the package was encrypted and it could not be open except for using a software from the digital signature issuer. This is why we decided to keep as versions the package as-ingested (version 1), the package decrypted and compressed (by the creator, using Winrar) (version 2) and the package uncompressed (version 3). Since we do not have a digital archiving solution, versions of the ingested version ("SIP"), the `*.txt` and `*.pdf` files migrated ("AIP") and Access database ("DIP") were all attached to one archival description in scopeArchiv.

5.2 About archival value of the data

A maybe naïve question may arise concerning the value of data: which is the envisaged user, in what consist the archival value of these datasets? At first sight, the answer is clear: these are valuable micro-data about the people and housing, harvested during census. It is the micro-image of the society, important for historians, genealogists and so on.

This "obvious" picture should be a bit amended. Firstly, we have the proof that the data, at least in the case of 2011 census, was amended, anonymized. In this regard, strangely enough, it should be noticed that even the legal texts ignore the fact National Archives can protect the personal data, can control the release of sensitive information, and must receive original documents. In both census, the grounding legal texts require information should be first amended, then transferred to the National Archives. Moreover, on the Institute of Statistics, on its website, offers access to anonymized microdata "only for scientific research" (see INS-microdate pentru cercetare stiintifica, 2017), quoting European regulations (see Commission Regulation 557/2013/CE, 2013; Regulation (CE) nr. 223/2009), fulfilling, in this regard, also the mission of the National Archives. In other words, the documents publicly released by the Institute are, basically, what National Archives received, as another "user" of the data and not as a preserver empowered by the State to maintain original, fully accurate evidence of the information created by public institutions. Letting aside the institutional misrepresentation, the range of uses in time is heavily affected by the intervention and anonymization of data. Regarding retrospectively, the censuses in 18th or 19th century are very valuable today because, containing "personal data", it helps genealogists, family historians etc. to precise identify persons, houses, households and correlate these data for reconstruction of past events and situations. In our cases, anonymized data can only be used for statistical purposes, on streets, districts or communities, raising the level of information above specific persons. While data protection is an understandable and necessary measure in today's society, I think it should be examined the

impact of applying the same protective over historically-relevant documents. We consider that expanding the protection periods for census data may be a sufficient way to guarantee the right of individuals, without altering the historical information contained.

But a more serious question arose in what concern the data themselves. In direct discussions with statisticians, they indicated there were significant errors in collecting data by forms and, indirectly, the same issue affected the databases. This information was confirmed by official documents. According to the *Quality Report of the 2011 Population and Housing Census (2011, p. 21-22)*, the forms data collected were rather imprecise (*Recensământul Populației și al Locuințelor 2011, p. 5-6)* and needed indirect sources for data harvesting, correlation and corrections (at country level, 5.9% of data were collected from indirect sources, while for Brasov county the percentage was 6.1). The final results were generated by collecting data from other sources and by applying statistical corrections to the primary data. In other words, preserving by the Archives of databases with primary data, with no further documentation indicating the quality of data, the corrections applied and the final results raises serious questions about the quality of these data as historical sources. With this perspective in mind, we requested the creator to transfer also the files containing the final results, preserving in this way first and last data sets, but, however, with no firm guarantees they are indeed what we considered necessary for historical analysis.

In the same regard, an issue on appraisal was identified. We do not want to have here a review over extensive professional literature and ideas, attempting to identify the goals of appraisal, methods and so on. The idea we would like to emphasize is that the data can be relevant not only by themselves, but in the context they were collected. By context, we mean not only the general framework (that is, the census), but also: intentions, methods used, tools involved and so on. Therefore, *documenting the census* is not only about data, but also about the way Institute of Statistics did its job. Receiving only some tables with data gives only a slice of reality. In this regard, we requested also for transfer, for both censuses, the manuals for census field agents, sector maps, instructions and procedures applied, original forms etc. Despite this extra-documentation, is should be noticed the website for 2011 census contains more rich and relevant documentation than that transferred to the Archives (see Institatul Național de Statistică, 2017) and, even if we harvested that website, we may consider the National Archives failed to properly preserve the relevant documentation for 2011 census (and also, very likely, for 2002).

6 Lessons learnt. Conclusions

The experience of ingesting digital information about censuses in 2002 and 2011 to the National Archives and attempts to preserve them allows for a set of lessons, regarding both technical and archival aspects.

It is necessary for the archivists to be involved in the process of selecting data for preservation. It may be self-implicit in theory, but the examples we discussed here showed it is not always the case. For many reasons, the access of the archivists to the production system may be restraint and a selection based on data-meaning or based on consideration of the producer may not be good enough. The appraisal, in the same time, should be based on an extensive analysis of the workflow, in order to identify if the data intended to be preserved is indeed the data that would serve the most the users' interests in the future. The content, but also the context of creating those data should be documented, as a way to assess the quality of data, understand how data were collected, with what purposes, within which framework.

Due to the versatility of digital data, it may exist a temptation for the creating agency to manipulate the original data, as to apply various protection methods, for personal data, for instance. In this regard, we consider the Archives should promote more firmly the need for original (whatever that means in digital environment) and its capacity/obligation of ensuring the confidentiality-for as long as necessary-of the information transferred. Archives should preserve, as we all know, the evidence of the activity, not some altered copy of information resulted from that activity. This is why, in our opinion, the control of the archivists over the export of data from the working system may be a way to achieve the extraction of "original" data.

At a technical level, for the data have a meaning, it is necessary not only to preserve them, but to also have an extensive description of the relationships and constraints between various tables and queries and of the original working system. Basically, to identify the "records" which are composed of those

"data". In practice, the fields and relations may be coded and may not be self-explicit; description should make them clearly understandable. Also, the lists of values may be mixed in various way for certain cases, so not only "authoritative lists" are necessary, but the sets that are effectively source for the main table(s), used for the data to be explicit. If some intervention were made on original data, those should be explicitly indicated, documenting type of data removed and the methods used. All these would reveal not only the meaning, but also the technical context of creation/preserving the data.

In the final line, I would like to challenge the duty of archivist to create systems that would make the data understandable for the users (in other words, to create a Digital Information Package). The data may come to the Archives from a variety of systems, many of them complex enough to require high skills to restore the functionality, which would definitely exceed the competences of an archivist. What if we shall take the data and just preserve them, with all the semantic, provenance fixity information required-and only that? Looking to the past, the archivists preserved medieval charters even if some of them did not understand Latin; glass plates were kept even if devices for read the image properly were not available. What if we shall deliver to the user a zip package containing tables and documentation about them (if exists?). Of course, one of the duty of our profession is to make the holding available to the users. Moreover, these days, when low budgeting is a common issue, the higher visibility is one way for collecting more money-so, just keeping stuff without making them available, without promoting them may not be a good business. From this point of view, creating DIPs seems more and more like an archival professional mission, which emphasize a new dimension for archival management: archivists are no longer sufficient for processing historical archives.

References

Commission Regulation 557/2013/CE of 17 June 2013 and its implementation guide.Available at www.insse.ro/cms/files/eurostat/esds/Reg_EC_557_2013_EN.pdf *(last visited 1 April 2017).*

Hotărârea Guvernului nr 922 din 21.09.2011 pentru modificarea și completarea *Hotărârii Guvernului nr. 1.502/2009 privind organizarea și desfășurarea recensământului populației și al locuințelor din România în anul 2011* în *Monitorul Oficial,* nr.689/28.IX.2011, art. 13.

Hotărârea nr. 1505 din 18 decembrie 2002 pentru modificarea Hotărârii Guvernului nr. 680/2001 privind organizarea și desfășurarea recensământului populației și al locuințelor din Romania în anul 2002, în *Monitorul* Oficial, nr. 19 din 15 ianuarie 2003.

Hotărârea nr. 680 din 19 iulie 2001 privind organizarea și desfășurarea recensământului populației și al locuințelor din România în anul 2002, în *Monitorul Oficial,* nr. 439 din 6 august 2001, art. 17.

INS-microdate pentru cercetare stiintifica (2017). Available at http://www.insse.ro/cms/ro/content/ins-microdate-pentru-cercetare-stiintifica (last visited 1 April 2017).

Institatul Național de Statistică, 2017. Available at http://www.recensamantromania.ro/ (last visited 1 April 2017).

Quality Report of the 2011 Population and Housing Census (2011). Available at http://www.recensamantromania.ro/wp-content/uploads/2015/02/Raport-de-calitate_RPL2011_ENGLISH.pdf (last visited 1 April 2017).

Recensământul Populației și al Locuințelor 2011, p. 5-6. Available at http://www.recensamantromania.ro/wp-content/uploads/2013/07/prezentare-rpl-2011__Partea_I.pdf (last visited 1 April 2017).

Regulation (EC) nr. 223/2009 of European Parliament and Council from 11 MArch 2009. Available at http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32009R0223&from=EN (last visited 1 April 2017).

SUMMARY

In 2002 and 2011 in Romania there were undertaken general censuses of populations and housings. The National Archives, by its departmental units, were required to consider for permanent preservation the documentation of these censuses, in an anonymized form. The databases were archivally processed several years after their accession, which raised some issues, since the documentation available about the databases was not enough and it was necessary a contact with key persons from the creator who were aware of the technical context. The paper describes operations performed to make out of the plain data (submitted) a fully operational database (disseminated), with intelligible data. Further on, the paper approaches the quality of data and their usefulness for research, since the

information was anonymized and the quality of microdata was admitted of being poor even by the collector. A final topic approached regards the question of completeness of information about a census, since Archives are keeping the initial data, but nothing about the legal, procedural and institutional context of the census.