

# Data quality evaluation: a comparative analysis of company registers' open data in four European countries

Janis Bicevskis  
Faculty of Computing  
University of Latvia  
Latvia  
Janis.Bicevskis@lu.lv

Zane Bicevska  
DIVI Grupa Ltd  
Latvia  
Zane.Bicevska@di.lv

Anastasija Nikiforova  
Faculty of Computing  
University of Latvia  
Latvia  
AN11093@lu.lv

Ivo Oditis  
Faculty of Computing  
University of Latvia  
Latvia  
Ivo.Oditis@lu.lv

**Abstract**—This paper is devoted to the analysis of open data quality of the company registers in four different countries. The data quality evaluation was obtained using a methodology that involves the creation of three-part data quality model: (1) the definition of a data object to analyse its quality, (2) data object quality specification using DSL, (3) the implementation of an executable data quality model enabling the scanning of a data object and detecting its deficiencies. All three components of the data quality model are designed as graphical language families, which allow formulating data quality specification for non-IT professionals. Validation of an open data published by company registers in four different European countries shows deficiencies in the published data and demonstrates the applicability of the proposed methodology for data quality evaluation.

**Index Terms**—data quality, executable models, Company register

## I. INTRODUCTION

THE open-world model is gaining ever more popularity [1]. The society calls for direct access to the information avoiding mediators and filters. The process of making data freely available to the public includes also providing access to the data in public company registers that serve public administration purposes. The state institutions, business and individuals are interested in facilitating the process of effective communication and receiving services. This is not possible without the accurate and timely registration of objects such as population, real estate, vehicles, taxes and other objects legally required to register in public registers.

In the situations when state information system data is made public, data quality is of crucial importance, i.e., can the open data be trusted and used, what are recommended purposes of data usage.

Guidelines and key principles for development of state information system in Latvia were defined in national program "Informatics" more than 17 years ago [2]:

- Public objects shall be accounted for and registered by a public agency operating under the supervision of the relevant ministry. The agency shall be responsible for the registered data including data precision, completeness, timeliness, etc. Data on each public object shall be registered by only one public agency.

- Duplicate entry or the same public object in another public agency is prohibited. All matters necessary shall be administered by the agency where the data object is registered.
- Information on public object shall be recorded at the time when information is generated avoiding temporary recording on paper and later data entry into an information system.
- Documents certifying public object registration shall be printed from the information system data base, thus ensuring compliance between the printed documents and the data stored in the information system.

Many state information systems were developed and implemented according to these principles that were meant to provide accumulation of qualitative information into public registers.

Assessing the state information systems in Latvia, we have to acknowledge that, indeed, many information systems operate in accordance with these principles. For example, personal identification documents are printed from the population register, vehicle registration certificates and driver's licenses are printed from the vehicle and driver register while using personal data from the population register, etc., Unfortunately, not all state information systems have implemented these principles, therefore it is important to analyze the data quality of these systems. Since state information systems have restricted accessibility, the task for researchers was to create an independent "external" mechanism for assessing the data quality without using the same information system that accumulated data.

Numerous studies have led to various definitions of data quality. For instance, data are of good quality if they satisfy the requirements imposed by the intended use [3]. The ISO 9001:2015 standard [4] considers data quality as a relative concept, largely dependent on specific requirements resulting from the data use. The same data may be sufficiently qualitative in one situation but completely useless under other circumstances.

This principle is confirmed by analyzing the data quality of the Latvian Population Register in 1999. In the Latvian Population Register a person is described by 7 data groups. Such personal identification data as registration number, name,

and surname contained a relatively small number of errors and were assessed as qualitative. While the data on place of residence, links to parental data or links to personal data of children were far from desirable quality. The data on place of residence available in the Latvian Population Register could not be used to contact the persons. Moreover, due to insufficient infrastructure and internet availability, the data entry into the register was not timely.

Due to the fact that the Population Register contains not freely published personal data, the study focused on an analysis of publicly available data of the company registers in 4 countries (Latvia, Estonia, Norway and the United Kingdom). Company records with the values of the parameters were "scanned", and deviations from data quality specifications were registered. The produced results are quite surprising showing that data accumulated and published for many years is of dubious quality.

This paper has two main objectives. The first objective is to clarify whether publicly available data provided by company register is trustful and what is the quality of these data for simple use, for instance, identifying a company and sending a message to this company. Our second objective is to verify quality evaluation methodology with the help of executable data quality models described further in detail [5].

The paper deals with following issues: overview about the methodology of evaluation of data quality (Section 2), an analysis of data quality of company registers in four countries – Latvia, Estonia, Norway, and the United Kingdom (Section 3).

## II. METHODOLOGY

So far, many studies have been devoted to data quality evaluation methodology and practice. All these studies can be divided into several groups:

- General studies on the data quality, in most cases, defining the data quality dimensions and their groupings [3], [6], as well as evaluation methodologies [7], [8]. The sources [6] and [8] provide a comprehensive overview on existing researches, methodologies and tools which can be explored and/or used in other researches on data quality problem.
  - The specific industry related data quality evaluation by analyzing industry-specific data and evaluation methodology [7], [9], [10]. Methodologies mentioned in [7], [10] are insufficiently industry-specific, as a result, it is difficult to apply their results to another area (it is difficult to re-use them for customizing to specific use-cases). Most of the proposed guidelines in the data quality evaluation methodology defined in [9] are difficult to use (especially for non-IT specialists) as it requires a lot of resources to complete required specification tables, for example, data quality parameter specification. Moreover, it could require involvement of the authors of this methodology. Therefore, it can hardly be used widely despite the fact that using this methodology could ensure comprehensive data quality specification and efficient data quality analysis.
  - The data quality evaluation of partially structured (semi-structured) data [11] and poorly structured data, such as Wikipedia [12]. Methodology for data quality analysis proposed in [11] covers even data quality improvement phase and allows involving stakeholders. Unfortunately, this methodology is not easy to understand for stakeholders. Moreover, the stakeholders are involved considering only their needs (finding out and satisfying them) and not trying to involve them in specific stages of the process of data quality analysis. Another method for quality evaluation is described in [12] and used to evaluate quality of Wikipedia articles and information contained in their info boxes; however, it is not applicable to specific datasets (e.g. data storing in the relational databases) without converting them into the appropriate format.
- The quality of open data published by the company registers in four European countries will be evaluated using the approach described in [5], which is used to evaluate the quality of fully-structured data. This chapter gives an insight into the methodology of data quality models according to the approach described in [5]. It is characterized by the following main characteristics:
- For each specific application, you can create your own data quality model and evaluate the quality of data for a particular application.
  - Data quality model can be described at different abstraction levels from informal text in natural language to an automatically executable model, SQL statements or program code.
  - Data quality model consists of three types of graphical charts describing data objects, data quality specification and data quality evaluation processes. The charts could be configured by creating and using domain-specific languages.
  - The data quality model is "external" to the information system that stores the accumulated data, i.e., the data quality model can be built without knowing about technologies used for accumulation of data.
- The proposed methodology is useful for both developers of information systems for defining data quality specification and for industry experts to assess the quality of the data published by different data suppliers.
- The proposed data quality evaluation solution consists of 3 main components: (1) data object, (2) quality requirements, and (3) quality evaluation process. These components form data quality model. The data object description defines the data which quality must be analyzed, the quality specification defines conditions which must be met to admit data as qualitative, and the description of quality evaluation process defines the procedure that must be performed to evaluate data quality.

A. Data Object

Traditionally the notion of a data object is understood as the set of values of the parameters that characterize a real-life object. The research results will be illustrated by simple examples from the Company Register of Latvia, the quality of which will be analyzed in the next chapter. In Fig.1 the data object Company is depicted with its attributes: Reg\_number – registration number of company, Name – name of company, Type – type of company, etc. The description of company is partly formal as rules for attribute values syntax are given.

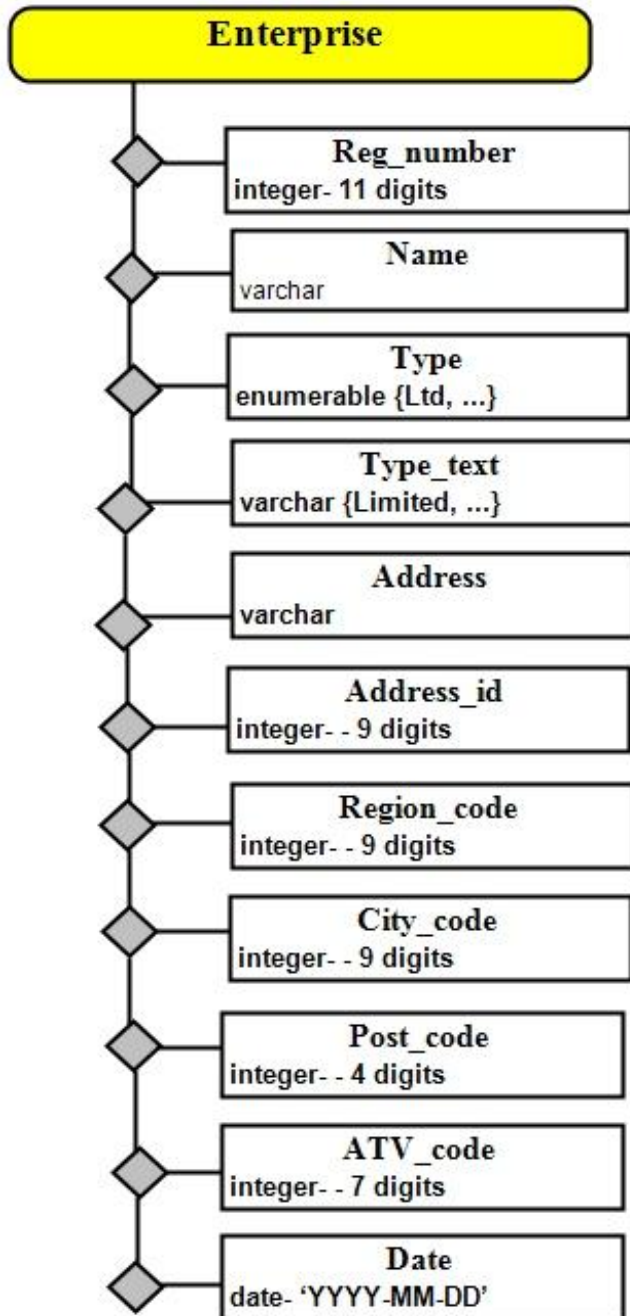


Fig.1. Data object “Company”

The quality checking of any data object parameters’ value is reduced to an examination of individual values’ properties, for instance, checking whether a text string may serve as a value of the field Name, or whether a value of the field Address is a correct address. Anyway, the checking of parameter values is a local and formal process. At the current stage of the research, it does not respect contextual interlinks with other data objects and does not check the compliance of data with the true characteristics of a real company.

The syntax rules for the permissible values of the data object’s fields can be formulated at different levels of abstraction including the formal language grammar and definitions of variables in programming languages. In the latter case, the data object model is closely related to the environment in which the model will be implemented.

A specific quality control of a particular data object usually is a part of the input data quality control in every information system. Data is usually entered into an information system by filling in screen form fields, followed by an information quality check and its retention in the database. In cases when the input fields are not filled correctly, the user receives an error message and may adjust the input data.

Information systems deal not only with individual data objects but also process many data objects in a unified way. In this case, the classes of data objects are used they represent many objects of the same structure. A data object class has a name, and its elements have the same structure and the same characteristic parameters. Each individual data object may contain parameter values fully or partially.

B. Data Object Quality Specification

A data quality specification contains conditions that must be met in order a data object is considered of high quality. The quality specification (Fig. 2) may contain informal descriptions of conditions, for example, in natural language or formalized implementation-independent descriptions. Data quality specification of a data object is defined by logical expressions. The names of data object’s attributes/ fields serve as operands in the logical expressions. The traditional means of programming languages, for example, the programming language C #, may be used as operations.

When processing the data object class, data object class instances are selected from the sources of information and written into a collection. All instances were processed cyclically by examining the fulfilment of a quality specification of each individual instance. The quality specification was similar to the specification used in the processing an individual data object. Thus, the quality problems of each individual instance were identified.

C. Quality evaluation process

The first step in the quality evaluation process describes the activities to be taken to select data object values from data sources. Thereafter, one or more steps are taken to evaluate the data object with a specific quality of the data, each of which describes one test for the compliance of data object “Company” with the quality specification.

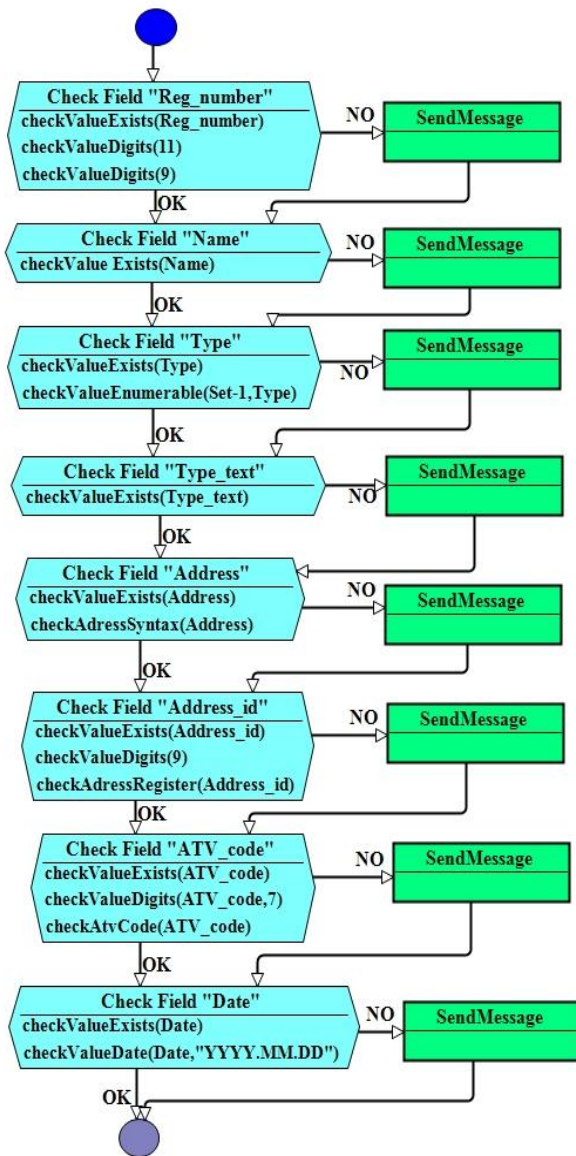


Fig.2. Quality specification of data object "Company"

In conclusion, steps to improve data quality can be performed by triggering changes in the data source.

The language describing the quality evaluation process involves verification activities for individual data objects which can be defined informally as a natural language text, or using UML activity diagrams, or in the own DSL. The Fig.3 contains separate field checks for the data object Company where each individual operation evaluates the data quality of the field by using a SQL statement. The SQL statement SELECT specifies the target data object, but WHERE specifies the quality specification. Such a data quality implementation is often used when data is stored in relational databases.

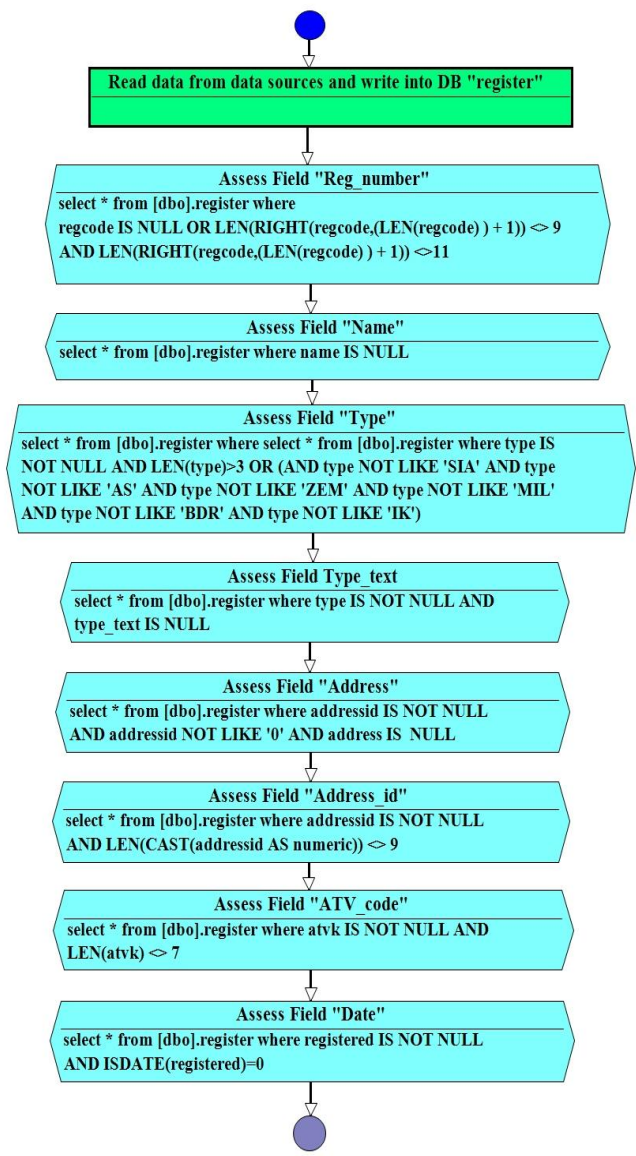


Fig.3. Quality evaluation process of data object "Company"

### III. EXPERIENCE OF OPEN DATA QUALITY EVALUATION

The data quality evaluation methodology was approbated on real data sets analyzing open data published by company registers in four European countries (Latvia, Estonia, Norway and the United Kingdom). The quality of company registers' data was evaluated setting two quite simple tasks (use case): identification of all recorded companies and verification of recorded business addresses by contacting companies.

A company was identified using the following parameters: registration number, name, type of activity and registration date. In the Latvian Company Register, this information is stored in the data fields Reg\_number, Name, Type and Date. If any of these fields was empty or did not correspond to the syntax rules, a company was not identified.

Company's business address with the postal code is needed to contact a company. The Company Register of Latvia stores

this information in fields Address and Post\_code. If any of these fields is empty or does not correspond to the syntax rules, the company cannot be contacted via mail.

Open data of company registers of four European countries were analyzed to fulfill the task of company identification and contacting via mail for all companies registered in the particular country.

The conclusions were drawn that data quality evaluation depends on specific data use case.

*a. Company Register of Latvia*

The Company Register (CR) is a public register partly available as open data. The open data set of the Latvian CR [13] contains 396 thousand records. A company is selected as the data object for evaluation. The structure of this data object class partly (11 fields) is described in the Fig. 1 and Table 1.

Each company in the CR is characterized by 22 parameters. The data quality checks showed that 13 of 22 fields have no syntactic errors. But as it is shown in the Table 1, data quality problems were detected in 9 fields. NULL values of the field Name in 10 records and NULL values of the field Date in 94 records are considered as severe data quality problems. The company name and registration date is primary information about company and may not be left empty. If these records relate to real companies, then identification of them will not be possible.

TABLE I.  
DATA QUALITY EVALUATION OF THE CR OF LATVIA

#	Field name	Field format	Error count	Comment
1.	Reg_number	11 digits NOT NULL	0	All companies have register number
2.	Name	NOT NULL	10 0.0025%	Company name is empty
3.	Type	NOT NULL Enumerable	0	Company type
4.	Type_text	NOT NULL Enumerable	1 403 0.35%	Company type (description) is empty
5.	Address	NOT NULL	366 0.09%	Company's legal address is empty
6.	Adress_id	9 digits, NOT NULL	4 523 1.14%	Company's address code is empty
7.	Region_code	9 digits, NOT NULL	280 662 70.7%	Region code is empty
8.	City-code	9 digits, NOT NULL	99 050 24.95%	City code is empty
9.	Post-code	4 digits, NOT NULL	20 498 5.16%	The postal code of company's address is empty
10.	ATV-code	7 digits, NOT NULL	5 521 1.39%	4 574 records - administrative territory code is empty 947 records - values are shorter than 7 digits
11.	Date	Date ('YYYY-MM-DD')	94 0.024%	Company's registration date is empty

However, if these records do not describe real companies, then these records should be removed from the CR. The number of incomplete records is not large and their processing would not require much work but this has not been done yet.

NULL values in the field Address in 366 records and NULL values in the field Post-code in 20498 records indicate potential data quality problems implying that this companies cannot be reached by mail. Other inaccuracies are not significant for the specific use case but they may be troublesome in other cases.

There are also 646 companies which according to their status are active but have NOT NULL value in "terminated" field which is not empty if only company's status is "closed" – liquidated or reorganized.

The blank values in other data fields, especially in the field Region\_code, lead to the conclusion that CR and users obviously lack the specification of open data or have incomplete information about filling fields, have to interpret the meaning of fields and acceptable formats themselves. The different interpretations of data formats and content lead to data quality problems.

*b. Company Register of Estonia*

The data quality of Estonian CR [14] was analyzed using the data set of 266171 records with 14 fields for each record. The data of Estonian CR seems to be of higher quality than the data of Latvian ER. All fields identifying companies were filled in.

The registration date of a company is not included in the set of open data, so it cannot be used to identify a company. The identified data quality problems were NULL values in address field of 29918 records as well as NULL valued in other address-related fields (see Table 2). Values of fields Etevoja\_address and ads\_ads\_oid are NULL in all records which suggest that the publication of values in these fields is unnecessary.

In addition to Ariregistri\_kood there also exists a KMKR number - Estonian value-added tax identification. 178 550 records do not have any KMKR number, however the most part of these companies should have this number although the given data set does not indicate it. This means that the Register of Companies of the Republic of Estonia does not provide complete data about companies.

To summarize the data quality problems were detected in 7 of 14 fields.

Unlike the Latvian CR published data, the Estonian CR open data can be used to identify companies. Though contacting a company to business address by mail may be difficult due to blank address fields.

TABLE III.  
DATA QUALITY EVALUATION OF THE CR OF ESTONIA

#	Field name	Field format	Error count	Comment
1.	Ariregistri_kood	8 digits NOT NULL	0	Company's registration number
2.	Nimi	NOT NULL	0	Company's name
3.	Ettevotja_staatus	Enumerable 'R', 'L', 'N'	0	Company's status is empty
4.	Asukoht_ettevotja_aadressis	NOT NULL	29 918 11.24%	Address is empty
5.	asukoha_ahak_tekstina	NOT NULL	19 964 7.5%	Text of address is empty
6.	indeks_ettevotja_aadressis	NOT NULL	22 621 8.5%	Company's address index
7.	ads_adr_id	NOT NULL	40 224 15.11%	Field is empty
8.	ads_normaliseeritud_taisaadress	9 digits, NOT NULL	40 099 15.1%	Field is empty
9.	ads_ads_oid			Field is empty
10.	ads_normaliseeritud_taisaadress	NOT NULL	40 099 15.1%	Field is empty
11.	Ettevotja_aadress			Field is empty

### c. Company Register of Norway

The Norwegian CR [15] was analyzed using the open data set of 1 100 993 records with 42 fields. Data analysis shows that all fields identifying companies are filled in according to the formatting rules (see Table 3). Also, legal addresses are available for all companies. There are no postal codes in 14683 records which could be seen as a data quality defect. Also in other fields that should contain addresses of a company related activities have blank values. Similar as in registers of other countries, the value of the field *forretningsadresse\_adresse* containing the legal address of the company is NULL in 68 128 cases.

Values of field *stiftelsesdato* containing date of company's liquidation are doubtful in 9 cases (1701-07-30, 1277-09-13, 1732-12-29, 1635-12-31, 1671-12-31, 1538-12-31, 1690-12-31, 1550-12-31).

TABLE IIIII.  
DATA QUALITY EVALUATION OF THE CR OF NORWAY

#	Field name	Field format	Error count	Comment
1.	organisasjonsnummer	9 digits NOT NULL	0	Registration number of Company
2.	navn	Varchar NOT NULL	0	Name of Company
3.	Registreringsdato	date NOT NULL	0	Date of registration of Company
4.	hjemmeside	NOT NULL	0	Address of Company
5.	stiftelsesdato	date	9	End-Date of Company is doubtful
6.	forretningsadresse_adresse	NOT NULL	68 128 6.2%	Field is empty
7.	forretningsadresse_postnummer	Enumerable 'R', 'L', 'N'	22 362 (2%)	Field is empty
8.	forretningsadresse_poststed	NOT NULL	14 683 (1.3%)	Post address is empty
9.	forretningsadresse_kommunenummer	NOT NULL	22 362 (2%)	Field is empty
10.	forretningsadresse_kommune	NOT NULL	22 362 (2%)	Region address is empty
11.	forretningsadresse_landkode	NOT NULL	14 863 (1.3%)	Company's address code is empty
12.	forretningsadresse_land	9 digits, NOT NULL	14 863 (1.3%)	Company's country code is empty

In general, it must be acknowledged that the information used to identify company via company's business is properly addressed. However, information necessary for postal communication with the company is missing in some cases. In total, data quality problems were detected in 8 of 42 fields.

### d. Companies House of United Kingdom

The CH of the United Kingdom [16] was analyzed using the open data set of 754 292 records which is about 20% of all registered companies.

Companies can be identified by values in fields *Company\_number*, *Company\_Name* and *IncorporationDate* (the date of foundation).

The quality of the data is very high, namely only one record does not have Company name and 3 records have dubious date values (see Table 4).

The stored addresses have quality defects since there are no addresses and/ or postal codes recorded for many companies.

TABLE IV.  
DATA QUALITY OF THE CH OF UNITED KINGDOM

#	Field name	Field format	Error count	Comment
1.	CompanyNumber	Varchar (8) NOT NULL	0	All companies have register number
2.	CompanyName	Varchar (160) NOT NULL	1 NULL	Company Name is empty
3.	RegAddress.AddressLine1	Varchar (300) NOT NULL	7 514 NULL (1%) 4 invalid	Company Address (Lin-1) 4 records contain: "XXXXXXX", "XXXXXX", "XXXXXXXXXXXX", "XXX XXX" values
4.	RegAddress.PostCode	Varchar (20) NOT NULL	12 151 (1.6%)	Company Address Post code is empty
5.	CompanyCategory	Varchar (100) enumerable NOT NULL	0	
6.	CompanyStatus	Varchar (70) enumerable NOT NULL	0	
7.	URI	Varchar (47) NOT NULL	0	Format: http://business.data.gov.uk/id/company/X', where X - CompanyName
8.	Incorporation Date	Date (DD/MM/YYYY), NOT NULL	3	invalid values - "16/06/1701", "09/08/1638", "25/04/1552"

Less significant data quality defects were found for the indicated company addresses: different names are used for one country in the register in RegAddress\_Country field (UNITED KINGDOM – 173 756 records and UK - 3; United States 447 and USA - 1, GREAT BRITAIN - 1, England - 5), certain listed values denote a particular area (WALES – 5 075, SCOTLAND, England & Wales – 184 097, England - 5, Virgin Islands - 41 and Virgin Islands, British - 22 and British Virgin Islands - 1) despite the fact that register contains countries value which unifies certain territories. Part of these values does not correspond with Companies House [14] policy which divides UK territory into Southern Ireland, England & Wales which companies are treated as a single entity and Scotland. Same tendency is observed in a country of origin: 464 records with country of origin of "United States" and 1 record – "United States of America"; 57 – "Great Britain", 3 – "England", 2 – "England & Wales" and 752 288 – "United Kingdom"; 143 – "Ireland", 9 – "Northern Ireland" and 2 – "Republic of Ireland"; 13 – "Nigeria" and 1 – "Republic of Nigeria". Moreover, 4 values does not corresponds with this field as they are not countries at all: "SW7" - South Kensington and part of Knightsbridge postcode, "EAST SUSSEX" which is a county in South East England, "BWI" - Baltimore/Washington International Thurgood Marshall Airport code, "DE 19901" Dover (city in the U.S. state of Delaware) postal code despite the fact that this field is supposed to contain only country names and there are specific

fields which are supposed to store postal codes and county names. There are companies from non-existing countries (Czechoslovakia, Jugoslavia, USSR which were registered after these countries have ceased to exist as political entities).

Moreover, there are 4 records where RegAddress AddressLine1, RegAddress AddressLine2, RegAddress PostTwown, RegAddress County and RegAddress Country values are "XXX".

In total, data quality problems were detected in 15 of 55 fields.

*e. Results*

This chapter analyzes the quality of data from company registers of 4 countries, which make some of their registry data open and available to public. A very simple example was chosen for data usage: (a) to search a company by its registration number or by its name, (b) once the company is identified, its address data is used to communicate with the company.

The performed data quality analysis is just one of the potential data usages. The data analysis was limited to the syntax analysis of data records, temporarily avoiding the analysis of interrelated (external) objects. As stated in [5], a deeper data quality analysis would lead to the analysis of links between records of the CR database and other data objects from external sources.

Despite the simplicity of the chosen data usage, inaccuracies were found in all four company registers. In total, percentage of columns with quality problems varies from 19% (in case of CR of Norway) to 50% (in case of CR of Estonia). Some quality problems could be easily solved thus improving the mentioned results significantly. Perhaps the authorities maintaining the registers are not even aware of this. This does not mean that the data from company registers cannot be trusted when a company must be identified since the number of defects is insignificant. The data in the company registers of Estonia and Norway can be used to identify companies without inconsistencies (all necessary fields are filled in). However, company register of Latvia (104 values - 0,0225%) and United Kingdom (4 values – 0,0005%) have several data quality problems which should be solved. However, correspondence with companies may fail, as the quality of address information is questionable. All analyzed registers had at least several data quality problems in the data fields containing the information be used contacting the company (address and postal code). The highest number of data quality problems were detected in the company register of Estonia (11,24% of address and 8,5% of postal code values were missing), the best results were shown by the company register of the United Kingdom (1% of address and 1,6% of postal code values were missing but also only 0,0005% of address values were invalid). These results show that data suppliers should inspect their data (authors of this paper recommend using the proposed approach) thus improving its quality. In addition, it can be concluded that small resources are needed to correct the few mistakes in the data to identify companies. But it would be much more difficult to complete address information.

#### IV. CONCLUSIONS

The paper has two main objectives: first, to verify authors' methodology of data quality evaluation using data quality models, and second, to determine whether or not the company register's published data are trustful, i.e., are these data of adequate quality for simple use. The study showed that:

- The data quality model allows describing the data quality independently of the information system that accumulates the data. Such an "external" data quality control mechanism enables the verification of "alien" data applicability within the context of our problem without the involvement of data holders.
- Using "alien" data from various sources of information, the author's had to reckon with the unprecise data definition in a meaningful sense. The data descriptions that were of interest to study were either not available or inaccurate. Thus, as in the case of company register, the authors had to interpret the data. This interpretation can be verified by analysing the data quality with the proposed methodology.
- The quality of the data can be described at least in two levels of abstraction: the one informally using natural language and another including executable program fragments or SQL sentences.
- Prior to use of the published by company registers in several countries, it is necessary to assess data reliability for a particular purpose. The open data quality analysis of the Latvian Company Register provided in this study serves as an example of data evaluation. The obtained results showed that the registers of all countries contain a small number of erroneous data. The most significant weakness of company data are fields describing company's legal or postal address; the values in these fields are just partially filled. In general, insufficient attention is paid to the quality of business company data. Even minor adjustments would make it possible to achieve significantly higher data quality.

This paper is a continuation of studies on information systems modelling [17], [18], [19]. Further research should focus on the creation of data quality models in the context of multiple data objects. It is also important to create a DSL to describe specific data quality activities.

#### ACKNOWLEDGMENT

This work has been supported by University of Latvia Faculty of Computing project AAP2016/B032 "Innovative information technologies".

#### REFERENCES

- [1] K. R. Popper, *The Open Society and its Enemies*, ([1945] 1966), 5th ed.
- [2] G.Arnica, J.Bicevskis, G.Karnitis, E.Karnitis: The mega-system: integration of national information systems : conceptual and methodological baselines . // Latvian Academic Library grey literature database. [Rīga], 2001
- [3] Redman, T.C.: *Data Quality. The Field Guide*, Digital Press, p. 74 (2001)
- [4] ISO 9001:2015: *Quality management principles*.
- [5] Bicevska, Z., Bicevskis, J., Oditis, I. Models of Data Quality. 12th Conference, ISM 2017, Held as Part of FedCSIS, Prague, Czech Republic, September 3-6, 2017, Extended Selected Papers. Lecture Notes in Business Information Processing, volume 311, pages 194-211 (2018)
- [6] Carlo Batini, Cinzia Cappiello, Chiara Francalanci, Andrea Maurino. Methodologies for data quality assessment and improvement. *ACM computing surveys (CSUR)* 41 (3), 16, 2009.
- [7] Freddie Bray, D. Max Parkin. Evaluation of data quality in the cancer registry: Principles and methods. Part I: Comparability, validity and timeliness *European journal of cancer*. March 2009 Volume 45, Issue 5, Pages 747–755.
- [8] Carlo Batini, Monica Scannapieco. *Methodologies for Information Quality Assessment and Improvement. Data and Information Quality Dimensions, Principles and Techniques*. Springer International Publishing Switzerland 2016.
- [9] Data Quality Evaluation Methods. International SEMATECH Manufacturing Initiative. Technology Transfer #08074943A-ENG, 2008.
- [10] Steven Van den Berghe, Kyle Van Gaeveren. Data quality assessment and improvement: a Vrije Universiteit Brussel case study. 13th International Conference on Current Research Information Systems, CRIS2016, 9-11 June. 2016, Scotland, UK.
- [11] Batini Carlo, Barone Daniele, Cabitza Federico, and Grega Simone. A data quality methodology for heterogeneous data. *International Journal of Database Management Systems (IJDMs-)*, Vol.3, No.1, February 2011.
- [12] Lewoniewski W. (2017) Enrichment of Information in Multilingual Wikipedia Based on Quality Analysis. In: Abramowicz W. (eds) *Business Information Systems Workshops. BIS 2017. Lecture Notes in Business Information Processing*, vol 303. Springer, Cham.
- [13] Company Register of Latvia. <http://dati.ur.gov.lv/register>
- [14] Company Register of Estonia. <https://opendata.riik.ee/en/dataset/http-avaandmed-rik-ee-andmed-ariregister>
- [15] Company Register of Norway. <http://data.brreg.no/oppslag/enhetsregisteret/enheter.xhtml>
- [16] Company House <https://www.gov.uk/government/organisations/companies-house>
- [17] J.Ceriņa - Bērziņa, J.Bičevskis, Ģ.Karnītis: Information systems development based on visual Domain Specific Language BiLingva, In: 4th IFIP TC2 Central and East European Conference on Software Engineering Techniques (CEE-SET), Krakow, Poland (2009).
- [18] J.Bicevskis, Z.Bicevska: Business Process Models and Information System Usability, *Procedia Computer Science* 77, pp. 72 – 79 (2015)
- [19] Bicevska, Z., Bicevskis, J., Karnitis, G.: Models of event driven systems. *Communications in Computer and Information Science* Volume 615, pp 83-98, (2016)