Linked Open Data in Health and Clinical Care

A Review of the Literature

Stefano Bonacina[1]

[1] Health Informatics Centre, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

Abstract

Background: In the range of Semantic Web, the idea of linking and sharing the resources generated by different authors, like ontologies, knowledge bases, or datasets, is referred to "Linked data". Then, an ambitious project within the "Linked Data" paradigm is the "Linking Open Data" community project. It aims at publishing open data sets on the Web and semantically connecting data items belonging to different data sources.
Objectives: The purpose of this paper is to present a literature review on the subject of Linked Open Data in Health and Clinical Care. In fact, the availability of open data would increase evidence of the results of biomedical research, and consequently, of clinical practice.
Methods: Selection criteria have been defined and searching in PubMed/Medline and Scopus citation databases - for all years the database were available - journals papers have been retrieved. Finally, an evaluation grid has been defined for analysing the retrieved papers, to answer some defined research questions.
Results: Nine journal articles have been analysed according to the defined evaluation grid. In five out of nine papers, the main contributions are strategies and methodologies for the integration of systems, including bridging the information gap among forms for clinical research and the one for patient care. Then, in three papers the main contributions are the development of consistent triple stores according to the "Linked Data" paradigm. Finally, the last paper aims at building an open dataset for public health purposes.
Conclusions: The review was able to answer the research questions, despite the limited number of included papers.

Keywords

Linked Open Data, Semantic Web, Healthcare, Clinical Care, Ontology

Correspondence to:

Stefano Bonacina
Health Informatics Centre; Department of Learning, Informatics, Management and Ethics; Karolinska Institutet
Address: Tomtebodavägen 18a, 171 77 Stockholm, Sweden
E-mail: stefano.bonacina@ki.se

EJBI 2016; 12(2):en2-en11
received: June 30, 2016
accepted: November 22, 2016
published: December 31, 2016

1 Introduction

The term "Open Data" refers to the opportunity of using and distributing freely available data or databases generated and shared by third parties, for own business or research [1]. The concept of openness of data implies a certain number of issues to face. In fact, applications involving "Open Data" should assure interoperability, security of the transmission and storage of data, and the continuous accessibility. In addition, updates and maintenance should be accomplished in a way that prevents from malfunctioning, misunderstanding, or misinterpretation of the original meaning of data and preserves data quality [2].

In the healthcare domain, the availability of open data would allow including considerable amount of data in processing tasks, envisaging for great evidence of the results [3]. As above mentioned above, a prerequisite is the interoperability among systems. It does not mean that a unique electronic system should be used in different institutions/companies, but the data should be represented in a standardised format in order to be correctly understood. Different standards exist and are applied in the health IT domain. Without the claim to be complete, examples of standard are Health Level Seven (HL7) Clinical Document Architecture (CDA), version 2 [4], for the communication of information systems by messages; HL7 Fast Health Interoperability Resources (FHIR) [5], for boosting effective implementations compared to the previous one; Digital Imaging and Communication in Medicine (DICOM), version 3.0, for the transmission of medical images and their characteristics among digital imaging instrumentation [6, 7]. Finally, the ISO/IEEE 11073 is a family of standards for enabling the communication among devices and computer systems, in hospital settings or at patient home [8].

When data are expressed by different formats, but share the same meaning ontologies can be used for overcoming the heterogeneous representation formats and preserving the meaning. According to Gruber, an ontology is defined as an "explicit specification of a conceptualization" [9]. The main concepts of that definition are "conceptualization" and "specification". First, a "conceptualization" is the abstract representation of what we experienced and would like to represent in terms of concepts, objects, and their relationships - including their properties [10]. Then, a "specification" is for specify the conceptualization according to a language [10]. For specifying ontologies, natural language or other kinds of languages (formal) can be used [10]. The Web Ontology Language (OWL) [11], and the Resource Description Framework (RDF) Language [12] are applied for specifying ontologies on the web that need to be shared by software systems, also without human intervention (Semantic Web) [12, 13]. In other words, those languages allow delivering data and its meaning (semantics) among heterogeneous systems on the Web, also in a distributed way. It is achieved by means of representing data by triples. A triple is composed by a Subject, a Predicate, and an Object, and represented as follows, (Subject-Predicate-Object) [12]. The meaning of triple elements is similar to the one of the elements of a sentence in natural language, e.g. "A doctor visits a patient". As more triples can refers to the same entity - intended as the value of an element, e.g. "A doctor" – a graphical representation is adopted for depicting all the information about that entity. Generally, that graphical representation is a direct graph where the subject and the object are represented as nodes of the graph, while the predicate is represented by the edge from the subject to the object [12]. The RDF language is used for representing data set, ontologies, or knowledge bases by graphs composed of triples distributed on the Web. For accessing and retrieving data from that kind of information structures expressed by RDF language, the SPARQL query language has been developed by the W3C consortium [14]. The recursive acronym SPARQL stands for "SPARQL Protocol And RDF Query Language". The term "Protocol" in the acronym is for specifying that the language includes a protocol - i.e., a set of rules - for publishing the results of a SPARQL query on the Web. Then, a SPARQL Endpoint is a software service that receives queries expressed in SPARQL language, and returns the results in accordance with the rules of the protocol. A catalogue of currently available SPARQL endpoints is maintained by W3C [15].

In the range of Semantic Web, the idea of linking and sharing the resources generated by different authors, like ontologies, knowledge bases, or datasets, is referred as "Linked data" [13]. Examples of projects that can be considered in the range of "Linked Data" are the Gene Ontology [16], and the Open Biomedical Ontologies (OBO) Foundry [17]. Then, an ambitious project within the "Linked Data" paradigm is the "Linking Open Data" community project [18]. It aims at "publishing open data sets as RDF on the Web and setting RDF links between data items from different data sources" [19]. One achievement of this continuously running project is the "Linking Open Data" cloud [20]. It is a set of interconnected datasets and other data sources expressed in RDF language and accessible by SPARQL Endpoints [15].

As we mentioned above, the availability of open data would increase evidence of the results of biomedical research [3]. So, having a view of the research involving "Linked Open Data" (LOD) in the domains of healthcare or clinical care would be of great significance.

Unfortunately, a review of scientific research involving LOD in those domains is not yet available, at least according to the author's knowledge, at the time of writing. Therefore, this paper aims at presenting a review of LOD scientific research projects in Health and Clinical Care. To this end, searches in literature databases (i.e. PubMed/Medline and Scopus) have been performed and the retrieved papers analysed. In analysing the paper, the properties of interoperability [21], security of the transmission and storage (protection), and the availability (continuous accessibility) and maintenance have mainly been considered.

2 Methods

2.1 Selection criteria and literature search

The main questions guiding this review are as follows,

1) which research in Healthcare and Clinical care involves "Linked Open Data" sources?
2) How have the issues of interoperability, protection, accessibility (availability), and maintenance been faced?
3) Which are the main contributions and the envisaged directions for future research?

The following inclusion criteria were defined for including studies about "Linked Open Data" in this review:

1) "Linked Open Data" in "Healthcare" or "Clinical care" is the main subject,
2) journal articles – including the "in press" ones (historical articles, review articles, editorials, book chapters, conference papers, and grey literature were excluded),
3) articles published before the date of literature searches,
4) duplicates were included once, and finally,

5) articles written in English.

The sources considered for the online searches were Medline/PubMed and Scopus citation databases. In both PubMed and Scopus databases the "Linked Open Data" search string was searched in title, abstract, or keywords of a citation. In more detail, search strings were defined including "Linked Open Data" as a part of the title, a keyword, or a term in the abstract and combining "Healthcare" or "Clinical care" terms, respectively. The search of those terms was not limited to a specific section of a publication, e.g. "All fields" tag included for searching in PubMed database.

On April 20th, 2016, literature searches were performed in Entrez, the National Center for Biotechnology Information (NCBI) search and retrieval system (PubMed.com) and in Scopus database (Scopus.com) according to the criteria above mentioned. Table 1 presents the query strings used in both of the considered sources. Figure 1 depicts the flow chart of selection criteria. Both of them include the figures of the search results. Every article was examined by the author according to the methods described in the next section.

2.2 Classification grid for literature characteristics

Considering the three research questions above mentioned, we defined the attributes for a classification grid

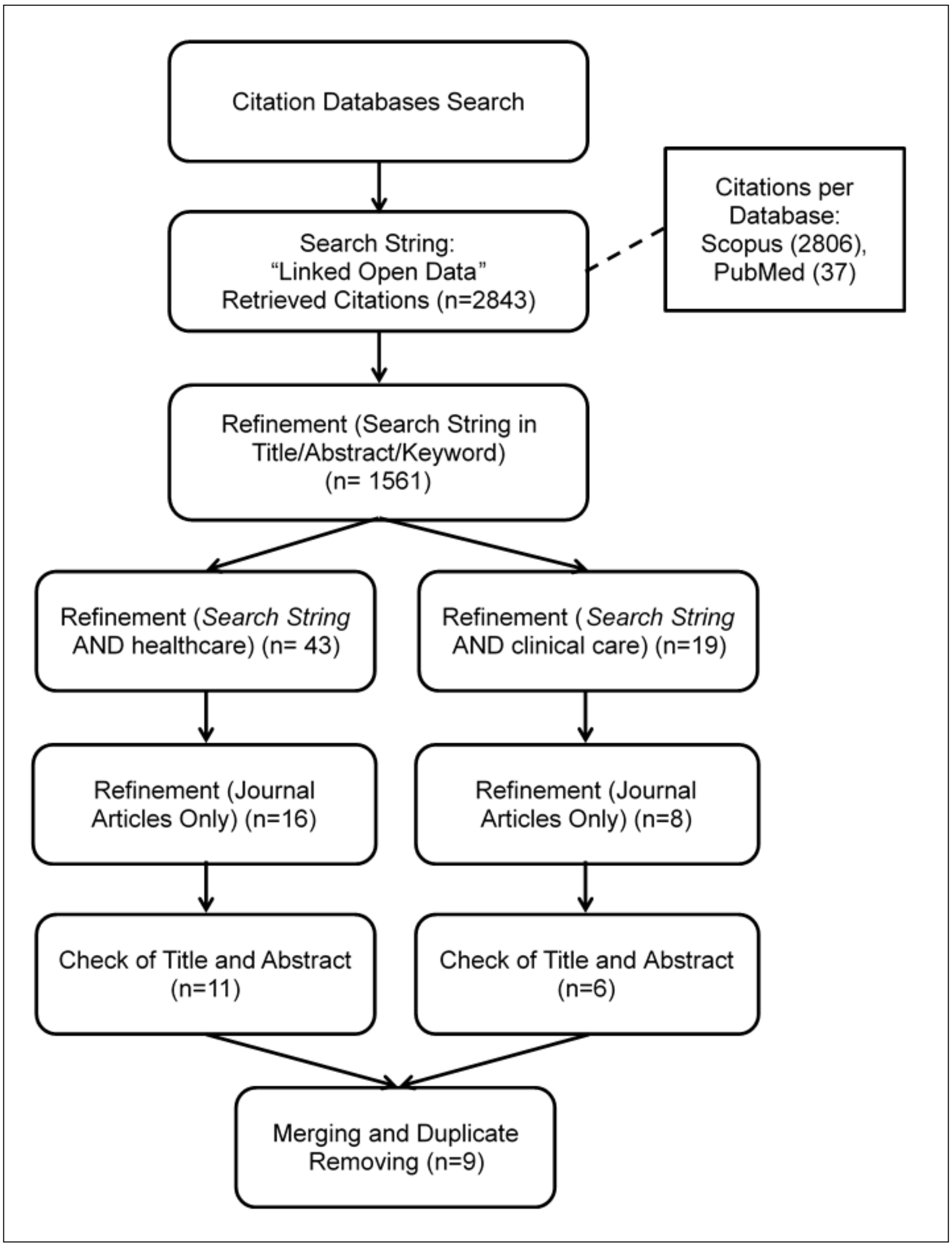

Figure 1: Literature search process according to the selection criteria.

Table 1: The search strings used and the number of identified citations.

No.	Search String	PubMed Results	Scopus Results	Sum of Results	Distinct Results
1	"Linked Open Data"[All Fields]	37	2806	2843	-
2	"Linked Open Data"[TIAB, KEYWORDS]	37	1524	1561	-
3	"Linked Open Data"[TIAB, KEYWORDS] AND healthcare	6	37	43	-
4	"Linked Open Data"[TIAB, KEYWORDS] AND healthcare (only journal articles)	4	12	16	13
5	"Linked Open Data" [TIAB, KEYWORDS] AND clinical care	1	18	19	-
6	"Linked Open Data" [TIAB, KEYWORDS] AND clinical care (only journal articles)	1	7	8	7

of the retrieved paper satisfying the inclusion criteria. For answering research question, 1) we defined the following attributes, "Article title", "Aim of the Research", and "Data Sources or Linked Open Data Involved" in the research. Generally, "Article title" and "Aim of the Research" attributes are easy to identify/extrapolate from a paper, while "Data Sources or Linked Open Data Involved" requires a scanning of the paper. That attribute refers to the data sources proposed or used in the research, including the ones available in the LOD cloud [20].

For answering research question 2), we defined the following attributes, "Interoperability / Accordance with Linked Open Data principles", "Protection", "Accessibility (availability), and maintenance". A brief explanation of their meaning is as follows. Interoperability is the ability of system to interact and understand each other, i.e., exchanging data, without any "special effort" [21]. In the LOD field, some "linked data" principles have been proposed by Berners-Lee [13, 22]. Those principles classify data according to a five-star scale, as follows [22]: 1) Data is available in the web according to the Open licence (no matter about the format); 2) as 1), plus data is expressed in machine-readable format (e.g. spreadsheets, tables as images); 3) as 2), plus data is expressed in non-proprietary format (e.g. Comma Separated Values – CSV format); 4) as 3), plus data is expressed according to W3C open standards (e.g., RDF and SPARQL) and they can be accessed by those standards; 5) as 4), plus "Linked RDF", "Link your data to other people's data to provide context" [22]. Then, "protection" refers to the tools or strategies adopted for protecting data, including the preservation of anonymity for data obtained from patients. Further, "Accessibility (availability), and maintenance" refers to the strategies adopt for ensuring a longitudinal service. In case of "5 stars data", it includes methods for updating data to maintain "Linked RDF" when data from others has been changed.

Finally, for answering research question 3), we defined the following attributes, "Main Contributions", and "Directions for future research". The "Main Contributions" attributes presents a summary of the findings considering the point of views of the healthcare domain and LOD domain; "Directions for Future Research" briefly presents next steps for contributing to the LOD domain. Furthermore, the "Year of Publication" has been considered for sorting the articles included in the review, and the "Reference Number" gives the citation number in the References section.

Summarising, the Classification Grid for literature characteristics is composed by the following attributes, "Reference Number", "Year of Publication", "Article Title", "Aim of the Research", "Data Sources or Linked Open Data Involved", "Interoperability / Accordance with Linked Open Data principles", "Protection", "Accessibility (availability), and Maintenance", "Main Contributions", and "Directions for Future Research".

3 Results

As reported in Table 1, and in Figure 1, the literature search identified 37 citations in PubMed and 2806 citations in Scopus applying ¡"Linked Open Data"[All Fields]¿ as a search string. Specifying the search strings according to the aim of the review, and the inclusion criteria (Figure 1), 9 papers have been considered and examined [23–31].

The results of the paper examination are presented in the Tables 2, 3, and 4, as the attributes for answering research questions 1, 2, and 3, were grouped as mentioned above. In Table 2, the attributes for answering research question 1 are presented for each paper. In 4 out of 9 papers, a framework – or system – supporting LOD datasets is developed locally for specific purposes (papers 2, 3, 5, and 7) [24, 25, 27, 29], see Table 2. In papers, 6, 8, and 9 the aim is the development of ontologies in RDF format [28, 30, 31]. Finally, in the paper 1 [23], a database has been developed with open data access, without Semantic Web technologies. As for the data sources involved, papers 2, 3, 5, and 7 [24, 25, 27, 29], include data collected locally and LOD sources to use for integrating or enriching those data in RDF format. In papers, 6, 8, 9, some LOD ontologies have been used as a base for the developed ones [28, 30, 31]. In paper 1, LOD sources have not been used (Table 2). Finally, the publication year for 3 out of 9 papers (papers 1, 2, and 3) [23–25] is 2013, for 3 out of 9 papers (papers 4, 5, and 6) [26–28] is 2014, for 1 out of 9 papers (paper 7) [29] is 2015, and for 2 out of 9 papers (paper 8, and 9) [30, 31] is 2016, Table 2.

In Table 3, the attributes for answering research question 2 are presented for each paper. In 4 out of 9 papers, five-star-linked open data requirements are met (papers 2, 3, 6, and 8) [24, 25, 28, 30] and a link available to the linked data datasets/system developed and presented in the paper have been found on the web, (Table 3, "Accessibility (availability), and maintenance" attribute). In three cases, papers 5, 7, and 9 [27, 29, 31], four-star-linked open data requirements would meet if the linked data datasets/system were published on the web, (Table 3, "Accessibility (availability), and maintenance" attribute). In one case, for paper 1 [23], three-star-linked open data requirements are met, as the developed Domesday dataset is available as Excel file and CSV format, (Table 3, "Accessibility (availability), and maintenance" attribute). Finally, paper 4 [26], does not meet linked open data requirements as a tool with just exemplary data is available (Table 3, "Accessibility (availability), and maintenance" attribute). As for patient data protection, it appears that it was properly addressed in 4 out of 9 papers (papers 1, 3, 7, and 8) [23, 25, 29, 30], while it appears not be faced in 5 out of 9 papers (papers 2,4,5,6, and 9) [24, 26, 27, 31].

In Table 4, the attributes for answering research question 3 are presented for each paper. In 5 out of 9 papers (papers 2,3,4,5, and 7) [24–27, 29] the main contributions are strategies and methodologies for the integration of systems, including bridging the information gap among forms for clinical research and the one for patient care (Table 4, "Main Contributions" attribute). Then, in 3 out of 9 papers (papers 6,8, and 9) [28, 30, 31] the main contributions are the development of consistent triple stores (Table 4, "Main Contributions" attribute). As for "Directions of Future Research" attribute (Table 4), performance improvement or increasing dataset size are the most mentioned.

4 Discussion and Conclusions

The presented review gives a snapshot of the current state of the research in the field of Linked Open Data for Healthcare and Clinical care. Tables 2, 3, and 4, describe the research performed in the field, how the issues of interoperability, protection, accessibility (availability), and maintenance have been faced, and main contributions and envisaged directions for future research, respectively.

According to the author's knowledge, at the time of writing, previous reviews on the subject of Linked Open Data in Health and Clinical Care have not been published. The number of considered papers – nine - could seem really limited; however, a rigorous and repeatable methodology has been followed in the literature search, as mentioned in the Methods section. Considering the publication year of the papers included in the review, two were published in the first third of the year, so more papers could be published on the subject, by the end of the year. Another reason for the limited number of papers respecting the inclusion criteria is the complexity of the Web Semantic subject and its limited spread in the clinical practice domain [18]. In fact, Web Semantic is still a research subject with main focus on ontologies [9, 10, 16, 17].

The results of this review revealed the main significant applications in the field developed so far [23–31]. In addition, the papers analysed have been evaluated according to the 5 star-based scales for linked open data requirements [22]. Then, the amount of papers analysed allowed answering the three research questions.

Those results can contribute to start new research initiatives in the domain of Linked Open Data for Healthcare and Clinical care. One direction is to initiate research in sub-domains that is not included in those mentioned in Table 2. Another direction is to improve current initiatives for fulfilling Linked Open Data principles comprehensively (see the Introduction), while protection, accessibility, and availability of data are maintained, as presented in Table 3. A third direction is to start research in the "Directions for Future Research", as described in Table 4.

The methodology applied in the presented review can help researchers in other fields in performing literature reviews. First, research questions have been defined. Then, literature databases have been selected according to the research field (LOD in Healthcare and Clinical care, in this literature review), search phrases have been defined, and searches executed. Then, a classification grid has been defined for being able to answer the research questions. The described process can be generalized and applied for other research fields.

Finally, this review presents some limitations. The authors decided to include only two citation databases (PubMed/Medline and Scopus), and including in the study only papers published in journals. However, considering conference proceedings would have led to the inclusion of the same project/system more than once, as conference proceedings are for presenting partial results of ongoing research.

In five years, the same literature review could be performed again and the results compared for assessing the spread of LOD in Health and Clinical Care.

Acknowledgements

The research here presented was supported by the Health Informatics Centre, Karolinska Institutet.

References

[1] The Open Definition [Online]. Available from: http://opendefinition.org Accessed: 15 June, 2016.

[2] Papavasileiou V, Flouris G, Fundulaki I, Kotzinos D, Christophides V. High-level change detection in RDF(S) KBs. ACM Trans Database Syst. 2013;38(1): Article 1, 42 pages. DOI=http://dx.doi.org/10.1145/2445583.2445584.

[3] Giglia E. Open access in the biomedical field: a unique opportunity for researchers (and research itself). Eura Medicophys. 2007 Jun;43(2):203-13.

[4] Raths D. Trend: standards development. Catching FHIR. A new HL7 draft standard may boost web services development in healthcare. Healthc Inform. 2014 Mar;31(2):13, 16.

[5] Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, Shabo Shvo A. HL7 Clinical Document Architecture, Release 2. J Am Med Inform Assoc. 2006 Jan-Feb;13(1):30-9.

[6] Kohn D. RSNA (Radiological Society of North America) participants embrace new imaging standard. DICOM 3.0 promises easy healthcare imaging communications. Health Manag Technol. 1994 Mar;15(4):26-9.

[7] Bidgood WD Jr, Horii SC. Introduction to the ACR-NEMA DICOM standard. Radiographics. 1992 Mar;12(2):345-55.

[8] International Organisation for Standardisation & The Institute of Electrical and Electronics Engineers. ISO/ IEEE 11073 family of standards. 2004. Available from: http://standards.ieee.org/ Accessed: 15 June, 2016.

[9] Gruber TR. A Translation Approach to Portable Ontologies. Knowledge Acquisition. 1993;5(2):199–220.

[10] Guarino N, Oberle D, Staab S. What Is an Ontology? In: Staab S, Studer R, editors. Handbook on Ontologies, 2nd Ed. Springer-Verlag: Heidelberg; 2009. p. 1-17.

[11] W3C OWL Working Group. OWL 2 Web Ontology Language Document Overview (Second Edition). 11 December 2012. W3C Recommendation. URL: http://www.w3.org/TR/owl2-overview/

[12] Klyne G, Carroll JJ, editors. Resource Description Framework (RDF): Concepts and Abstract Syntax. [Online, 2003]. W3C Proposed Recommendation 15 December 2003. Available at http://www.w3.org/TR/rdf-concepts/

[13] Berners-Lee T, Hendler J, Lassila O. The Semantic Web. Scientific American. 2001;284;34-43.

[14] Prud'hommeaux E, Seaborne A, editors. SPARQL Query Language for RDF. W3C Recommendation 15 January 2008. [Online, 2008]. Available from: http://www.w3.org/TR/rdf-sparql-query/ Last Access: 15 June 2016.

[15] W3C Wiki. SPARQL Endpoints. [Online, Updated 2016]. Available from: https://www.w3.org/wiki/SparqlEndpoints Last Access: 15 June 2016.

[16] Smith B, Williams J, Schulze-Kremer S. The ontology of the gene ontology. AMIA Annu Symp Proc. 2003:609-13.

[17] Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol. 2007 Nov;25(11):1251-5.

[18] Bizer C, Heath T, Berners-Lee T. Linked data—the story so far. International Journal on Semantic Web and Information Systems (IJSWIS). 2009;5(3):1–22.

[19] W3C SWEO Linking Open Data community project. [Online, 2007; Update, 2016] Available from: http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData. Last Access: 15 June 2016.

[20] Schmachtenberg M, Bizer C, Jentzsch A, Cyganiak R. Linking open data cloud diagram 2014. [Online, 2010. Updated 2014]. Available from: http://lod-cloud.net/ Last Access: 15 June 2016.

[21] IEEE. Interoperability. Standards Glossary [Online]. Available from: https://www.ieee.org/education_careers/education/standards/standards_glossary.html Last Access: 15 June 2016.

[22] Berners-Lee T. Linked data. [Online, 2006. Updated, 2009] Available from: https://www.w3.org/DesignIssues/LinkedData.html. Last access: 15 June 2016.

[23] Reddington J. The Domesday dataset: linked open data in disability studies. J Intellect Disabil. 2013 Jun;17(2):107-21.

[24] Sinaci AA, Laleci Erturkmen GB. A federated semantic metadata registry framework for enabling interoperability across clinical research and care domains. J Biomed Inform. 2013 Oct;46(5):784-94

[25] da Silva KR, Costa R, Crevelari ES, Lacerda MS, de Moraes Albertini CM, Filho MM, Santana JE, Vissoci JR, Pietrobon R, Barros JV. Glocal clinical registries: pacemaker registry design and implementation for global and local integration–methodology and case study. PLoS One. 2013 Jul 25;8(7):e71090.

[26] Samadian S, McManus B, Wilkinson M. Automatic detection and resolution of measurement-unit conflicts in aggregated data. BMC Med Genomics. 2014;7 Suppl 1:S12.

[27] Tilahun B, Kauppinen T, Keßler C, Fritz F. Design and development of a linked open data-based health information representation and visualization system: potentials and preliminary evaluation. JMIR Med Inform. 2014 Oct 25;2(2):e31.

[28] Xu B, Xu L, Cai H, Jiang L, Luo Y, Gu Y. The design of an m-Health monitoring system based on a cloud computing platform. Enterprise Information Systems. 2015; DOI: 10.1080/17517575.2015.1053416

[29] Saleem M, Padmanabhuni SS, Ngomo AC, Iqbal A, Almeida JS, Decker S, Deus HF. TopFed: TCGA tailored federated query processing and linking to LOD. J Biomed Semantics. 2014 Dec 3;5:47

[30] Bamparopoulos G, Konstantinidis E, Bratsas C, Bamidis PD. Towards exergaming commons: composing the exergame ontology for publishing open game data. J Biomed Semantics. 2016 Feb 9;7:4.

[31] Kawazoe Y, Imai T, Ohe K. A Querying Method over RDF-ized Health Level Seven v2.5 Messages Using Life Science Knowledge Resources. JMIR Med Inform. 2016 Apr 5;4(2):e12

Table 2: The filled in classification grid for answering Research Question 1.

Article No. / Reference No.	Year of Publication	Article Title	Aim of the Research	Data Sources or Linked Open Data Involved
1 / [23]	2013	The Domesday dataset: Linked open data in disability studies	To develop a database (The Domesday dataset) about the Augmentative and Alternative Communication (AAC) devices in use in the United Kingdom.	Data about provision of AAC devices was collected from different types of public body, mainly the National Health System (NHS) Trusts, and Local Education Authority (LEA), between 2006 and 2011.
2 / [24]	2013	A federated semantic metadata registry framework for enabling interoperability across clinical research and care domains	To develop a framework (a so called, federated metadata registry -MDR) for semantically linking the forms (Common Data Elements - CDEs) created for clinical research data with the ones used for patient care in clinical practice.	The MDR standard ISO/IEC 11179 has been translated according to the Linked Open Data principles and a triple store made available, as SPARQL Endpoint. The framework connects a number of CDEs (CDISC SHARE, HITSP, and OMOP), which are maintained by different MDRs, applying SKOS ontology. Simple Knowledge Organization System (SKOS) is for representing terminologies, thesauri and classifications according to Semantic Web technology (i.e. triple stores). Bioportal, which is an ontology repository and terminology server, offers common medical terminologies as triple stores.
3 / [25]	2013	Glocal Clinical Registries: Pacemaker Registry Design and Implementation for Global and Local Integration – Methodology and Case Study	To develop a device registry framework (applied to a pacemaker registry) for linking clinical research data with patient care data in clinical practice. Specifically, a database for reporting on pacemaker long-term outcomes has been developed.	Data is from the randomized trial ("Safety and the Effects of Isolated Left Ventricular Pacing in Patients With Bradyarrhythmias," ClinicalTrials.gov study ID NCT01717469). Data standards from ACC/AHA, CDISC, NCDR and Linked Clinical Trials (Linked CT) triple store, available in the LOD cloud.
4 / [26]	2014	Automatic detection and resolution of measurement-unit conflicts in aggregated data	To solve the problems of integration and harmonization among units of measurement, by a Semantic Web Service-based approach.	The clinical records of a patient cohort (536 unique patients) collected between 1986 and 1989 from a referral hospital in Nebraska, USA. GALEN compositional ontology, Semantic Automated Discovery and Integration (SADI) Semantic Web Service framework, Semantic Health and Research Environment (SHARE) mediator system, the Semantic Science Integrated Ontology (SIO).
5 / [27]	2014	Design and Development of a Linked Open Data-Based Health Information Representation and Visualization System: Potentials and Preliminary Evaluation	To develop a system for representing health information, querying, and visualizing them - the so called Linked Open Health Data (LOHD) - by integrating public health data coming from different sources, applying the Semantic Web technologies.	Data from United Nations program for HIV/AIDS (UNAIDS), and the Linked Open datasets DBpedia, Bio2RDF, and LinkedCT.
6 / [28]	2014	TopFed: TCGA tailored federated query processing and linking to LOD	To translate The Cancer Genome Atlas (TCGA) in a triple store (i.e. a Linked Data version of the atlas) and to develop a query engine, TopFed, offering SPARQL endpoints for improving remote query processing and virtual data integration.	TCGA, by the National Cancer Institute, makes available the characterization and sequencing of more than 30 cancer types, from 9 thousand patients at the molecular level.
7 / [29]	2015	The design of an m-Health monitoring system based on a cloud computing platform	To design and develop a Software as a Service (SaaS) cloud system for facilitating collection, storage, and processing of personal healthcare data for monitoring applications.	Diseasome, Depedia, and Drugbank triple stores for linking to disease definitions and drug descriptions. Dbpedia for disease descriptions. Personal data from patients in the antimicrobial resistance case study.

Article No. / Reference No.	Year of Publication	Article Title	Aim of the Research	Data Sources or Linked Open Data Involved
8 / [30]	2016	Towards exergaming commons: composing the exergame ontology for publishing open game data	To develop a model for the semantic representation of exergames (i.e. games for guiding in physical exercises) using OWL, including an exergame ontology that embodies game session concepts. A SPARQL endpoint is offered, too.	Game Ontology Project (GOP) ontology, Generic Component Model (GCM) architecture framework for describing component-based system formally. Ontology of physical activity (OPA), Ontology of physical exercise (OPE), Friend of a friend (FOAF) ontology for the player description, NCI Thesaurus for the description of the muscles involved in the exercises, and "Quantity, Unit, Dimension and Type" (QUDT) collections of ontologies for representing the units of measurements of the game metrics, vCard for the information about the locations of game sessions, and SKOS for creating a terminology of common terms for exergames.
9 / [31]	2016	A Querying Method over RDF-ized Health Level Seven v2.5 Messages Using Life Science Knowledge Resources	To develop a method for converting HL7 messages into RDF data, and use linked drug databases for enriching results of SPARQL query of clinical data, including adverse drug events.	Anatomical Therapeutic Chemical Classification System (ATC), United States Pharmacopeia Classification (USP), SIDER 2, KEGG, MEDIS DRUG. Medication orders and laboratory test results of 148 thousand of unique patients from The University of Tokyo Hospital.

Table 3: The filled in classification grid for answering Research Question 2.

Article No. / Reference No.	Interoperability / Accordance with Linked Open Data principles	Protection	Accessibility (availability), and maintenance
1 / [23]	Three-star-linked open data requirements are met	Collected data are related to AAC devices; in addition the dataset has been designed and developed in a way that prevents revealing the identity of the AAC device users.	The dataset is available at: http://joereddington.com/aac-and-the-domesday-dataset/
2 / [24]	Five-star-linked open data requirements are met, for the federated MDR.	It appears protection of patient data is not faced in the study.	The federated MDR standard ISO/IEC 11179 in RDF format is available at: https://github.com/srdc/semanticMDR SALUS Common Information Model Ontology is available at: http://www.salusproject.eu/ontology/salus-cim-ontology.n3 OMOP CDM Content Entity Model Ontology is available at: http://www.salusproject.eu/ontology/omop-cdm-ontology.n3 CDA/CCD Content Entity Model Ontology is available at: http://www.salusproject.eu/ontology/hl7-cda-ontology.n3
3 / [25]	Five-star-linked open data requirements are met.	In accordance with Good Clinical Practices (GCP) and The Health Insurance Portability and Accountability Act (HIPAA) of 1996.	The Cardiac Pacemaker Clinical Trials is available at LinkedCT: http://www.linkedct.org/resource/trial/nct01717469/ However, last update was made in 2012.
4 / [26]	Tools for demonstrating the BMI example presented in the paper are available. Linked open data requirements are not met, as data are not available.	It appears protection of patient data is not faced in the study.	The BMI example is available at: http://biordf.org/MeasurementUnitsDemo/
5 / [27]	At present, the system developed is not available for use. If the data were published on the web, four-star-linked open data requirements would meet.	It appears protection of patient data is not faced in the study.	It appears that the availability of the system is limited and not public.
6 / [28]	Five-star-linked open data requirements are met, as Linked TCGA database is available.	The Linked TCGA does not include patient data, so appears that protection issues are not faced in the study.	The Linked TCGA database is available at: http://tcga.deri.ie/ The TopFed's utilities and tools are available at: http://goo.gl/rtwm6q
7 / [29]	At present, the Cloud-MHMS is applied as a case scenario for allowing general practitioners (GPs) in China to access patient data and refer them to the appropriate hospital. If the data were published on the web, four-star-linked open data requirements would meet.	Multiple tenant access control is implemented to allow data isolation and data sharing.	It appears that the availability of Cloud-MHMS is limited to GPs in China.
8 / [30]	Five-star-linked open data requirements are met, as data from game sessions have been published on the web.	Data about game sessions was published carefully obscuring any information that may reveal identity.	Exergame ontology is available at http://purl.org/net/exergame/ns#.
9 / [31]	At present, the developed methods are applied only at The University of Tokyo Hospital. If the data were published on the web, four-star-linked open data requirements would meet.	It appears protection of patient data is not faced in the study.	It appears that availability of data is for clinicians and medical staff at The University of Tokyo Hospital.

Table 4: The filled in classification grid for answering Research Question 3.

Article No. / Reference No.	Main Contributions	Directions for Future Research
1 / [23]	The availability of a dataset including the provision data of AAC devices in the UK for research, policy definition, and public health purposes.	Continuing update of the dataset including data from public bodies (not included in the initial development). The assessment of the impact of AAC apps available on tablets.
2 / [24]	The developed federated semantic metadata registry. The definition of CDEs in machine-readable format.	Locally developed CDEs could be linked with CDEs developed by standardisation bodies.
3 / [25]	The developed Cardiac Pacemaker Clinical Trials based on LOD principles.	Implementing the integration with a platform for adverse events monitoring, defining protocols for data enrichment through natural language processing (NLP) methodologies.
4 / [26]	The definition of a Semantic Web-based solution for encoding units of measurement in a machine readable way and harmonising them in (clinical) databases.	Improving performance of the solution when large datasets are examined, and including more complex pattern of units of measurements, and temporal units, as well.
5 / [27]	The developed LOD-based health information representation, querying, and visualization system which uses Linked Data tools. The evaluation test of the query system – based on SPARQL – showed friendly interfaces should be introduced for users unfamiliar with Semantic Web technologies.	Improving the query engine and its interface for users unfamiliar with Semantic Web technologies.
6 / [28]	The Linked TCGA RDF dataset (20.4 billion triples), the TopFed query engine, and the SPARQL endpoints. Evaluation of the TopFed performances for 10 different SPARQL queries.	Applying the system in clinical practice as a decision support tool for suggesting drugs treatment for cancer patients.
7 / [29]	A cloud computing structure based on multiple layers for collecting data from patient monitoring devices, transfer them to the storage layer, and make them available to healthcare professionals. The cloud solution allows to memorize data one while GPs and clinicians can access them avoiding data duplication and cost for multiple collecting/storing.	Linked data model will be used for data analysis in Hospital Information Systems.
8 / [30]	The Exergame ontology enriches semantically data from exergames and allows semantic processing.	Enriching the ontology increasing the number of sessions and the number of exergames. Performance and scalability evaluation.
9 / [31]	The method for converting HL7 messages into RDF; a potential large-scale data federation for retrieving clinical information enriched by drugs information. About 650 million RDF triples for medication orders and 790 million RDF triples for laboratory test results were produced.	The application of the proposed methods in other hospitals that are using HL7 standard for exchanging messages.