GEOSPATIAL LINKED DATA PROLIFERATION IN NMCAs:

SYSTEMATIC LITERATURE REVIEW

Dr. Marjan Čeh
University of Ljubljana, Slovenia

Article:

RECENZIRANI ČLANKI | PEER-REVIEWED ARTICLES

UVAJANJE TEHNOLOGIJE LINKED DATA V GEODETSKIH UPRAVAH: SISTEMATIČNI PREGLED LITERATURE

GEOSPATIAL LINKED DATA PROLIFERATION IN NMCAS: SYSTEMATIC LITERATURE REVIEW

Marjan Čeh, Jernej Tekavec

UDK:
Klasifikacija prispevka po COBISS.SI: 1.02
Prispelo: 10. 4. 2023
Sprejeto: 31. 5. 2023

DOI: 10.15292/geodetski-vestnik.2023.02.222-243
REVIEW ARTICLE
Received: 10. 4. 2023
Accepted: 31. 5. 2023

SI | EN

IZVLEČEK

V raziskavi smo se osredotočili na sistematični pregled literature o objavljanju geoprostorskih podatkov v tehnologiji povezanih podatkov (linked data – LD) pri geodetskih upravah evropskih držav, odgovornih za zemljiško administracijo. Analizirali smo znanstvene študije, objavljene v letih od 2014 do 2023. Naša raziskava je bila usmerjena v odkrivanje vzrokov za razmeroma počasno uveljavljanje LD-tehnologij za objave geoprostorskih podatkov. Zato smo iskali razloge, najpogostejše težave, rešitve in izzive za objavo prednostnih tem geoprostorskih podatkov pri geodetskih upravah. Uporabljena raziskovalna metodologija temelji na dobro uveljavljenih pristopih, kot so analiza problemov, posegov, primerjav, izidov in okoliščin (PICOC) ter iskanje s tako imenovanim »vzvratnim valjenjem snežne kepe«. Postopek presoje upravičenosti za vključitev člankov v sistematični pregled literature je predstavljen v Smernicah za prednostne postavke poročanja za sistematične preglede in metaanalize (PRISMA). Pri branju celotnega besedila končnega nabora devetnajstih izbranih člankov smo analizirali vsebino in identificirali dvanajst ključnih področij. S frekvenčno analizo smo pridobili pregled LD objavljenih geoprostorskih tem, ki so najpogosteje objavljene pri geodetskih upravah

KLJUČNE BESEDE

povezani podatki, geodetske uprave, geoprostorska semantika, interoperabilnost, sistematični pregled literature

ABSTRACT

This research focuses on a systematic literature review of the geospatial Linked Data (LD) publication at National Mapping and Cadastral Agencies in Europe (NMCA) responsible for land administration. We analysed scientific studies published from 2014 to 2023. Our research aimed to find out the reasons for the relatively slow adoption of geospatial data publication as LD. Therefore, we searched for the most common problems, solutions, and challenges for publishing prioritised data themes in NMCAs. Applied research methodology relies on well-established approaches such as analysing Population Problems, Intervention, Comparison, Outcomes, and Circumstances (PICOC) and backward snowballing. The eligibility process is presented by Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines. We analysed nineteen extracted papers and identified twelve key issues from full-text reading. Frequency analyses resulted in the overview of the most frequently published geospatial themes as LD in several European countries. We addressed five research questions and concluded that in addition to systematic support for developing vocabularies and ontologies, trust, awareness, and knowledge about best practices could influence the success of LD publication at NMCAs.

KEY WORDS

Linked Data, National Mapping and Cadastral Agency, NMCA, geospatial semantics, interoperability, Systematic Literature Review, SLR

Geodetski vestnik

Geodetski vestnik

Geodetski vestnik is the publication of the Association of Surveyors of Slovenia (Zveza geodetov Slovenije) and it is an open access journal, issued quarterly in print and online versions. The journal is designed to serve as a reference source and an archive of advancements in the disciplines that make use of the geodetic, geospatial and spatial planning fields.

Current issue: 2023 (67), 2

CONTEXT for Systematic Literature Review

Research project V2 - 2295:

„Developing **guidelines**

to **improve** semantic **interoperability**

in spatial database management and

geoinformatics in **Slovenia**“

Surveying and Mapping Authority of the Republic of Slovenia

Project **aims**:

- Identify the **main obstacles** to **semantic interoperability** at the NMCA

- Find the **optimal combination** of the latest **technologies** and **approaches** of the semantic web for **geospatial LD publication**

Systematic Literature Review (SLR)

PICOC method:

1. Problem: siloed geospatial data

2. Intervention: publishing geospatial data as Linked Data

3. Comparison: web searchability, linkability

4. Outcomes: machine-readable, published geodata as open-linked data

5. Circumstances: NMCAs

RESEARCH QUESTIONS for LD implementation at NMCAs:

RQ1: The **main PROBLEMS** for publishing NMCA geospatial data as LD?

RQ2: **SOLUTIONS** and requirements **identified by experienced NMCAs**?

RQ3: **PRIORITY DATA THEMES** already published by NMCAs?

RQ4: What **CHALLENGES/ISSUES** have NMCAs identified?

RQ5: **REASONS** for the relatively **SLOW ADOPTION** of LD by NMCAs?

Existing studies **ANALYSES**

Databases: SCOPUS IEEE Google Scholar

Period: 2014 - 2023

Keywords: linked data (LD) AND geospatial AND NMCA

How to select studies for detail analyses SLR ?

6 criteria:

Table 1: Assessing studies before detailed data extraction with criteria/questions

Criteria	Eligibility question before extraction
C1:	Is the study relevant and essential to our research question?
C2:	Does the study focus on geospatial data?
C3:	Does the study clearly define the problems of LD publication?
C4:	Does the study explain the challenges of LD publication?
C5:	Does the study explain the best practices of LD publication?
C6:	Does the study cover the NMCAs domains?

Protocol of SLR

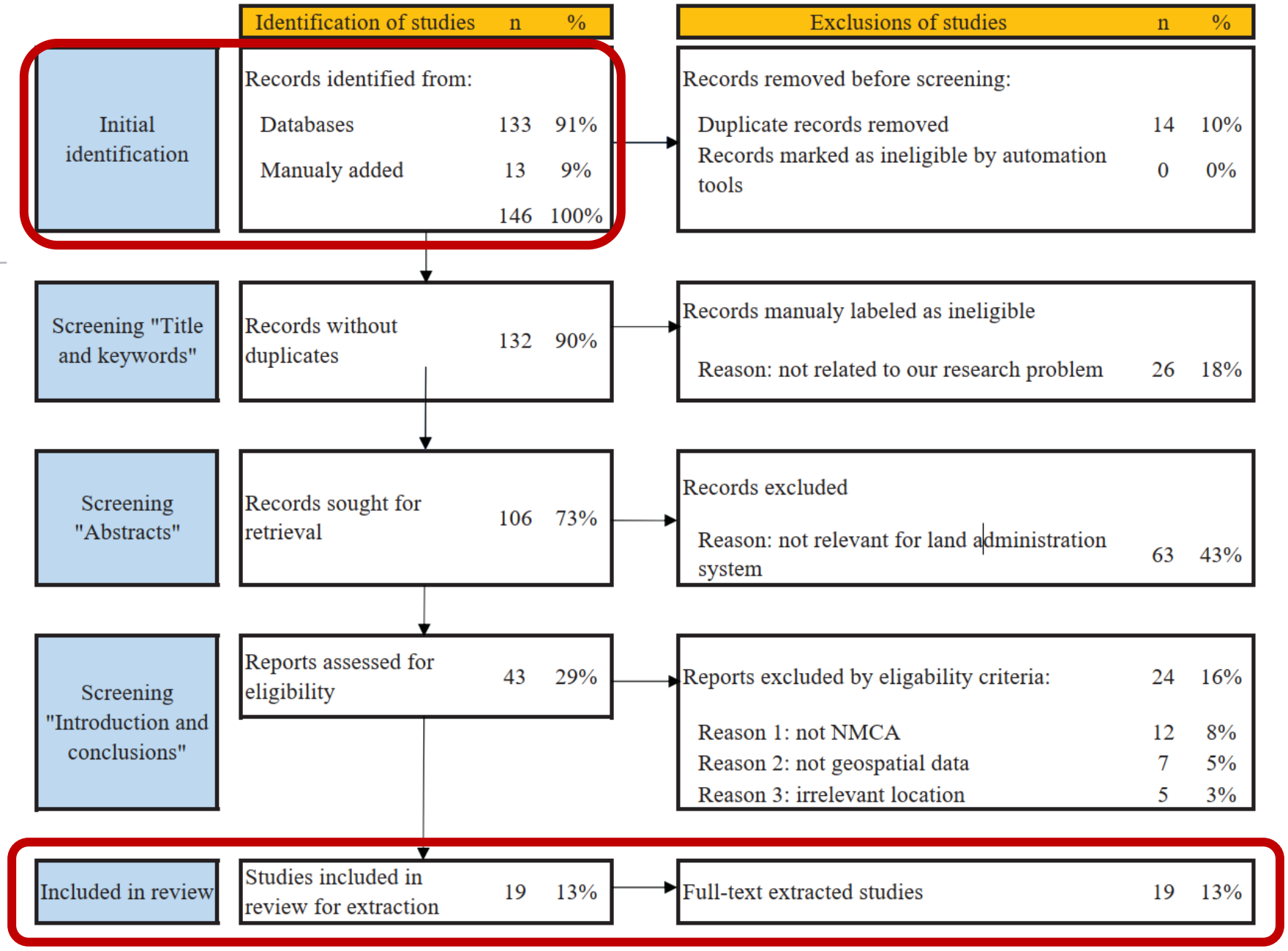


Figure 1: Results of studies selection in the form adopted from the PRISMA 2020 flow chart (Page et al., 2021)

19 most relevant studies according to 6 criteria (1 - 13):

Table 2: Detailed selection of studies and quality assessment criteria

ID	Author	C1	C2	C3	C4	C5	C6	Total score	%
		Relevance	Geospatial	Problems defined	Challenges defined	Best practices	NMCA domain		
1	Saavedra et al., 2014	0.5	1	1	1	0.5	1	4.0	67%
2	Çağdaş and Stubkjær, 2015a	1	1	0.5	0,5	0	1	4.0	67%
3	Çağdaş and Stubkjær, 2015b	1	1	0.5	1	0	1	4.5	75%
4	Hietanen et al., 2016	0.5	1	0.5	0,5	0	1	3.0	50%
5	Beek and Folmer, 2017	1	1	0.5	0,5	1	1	5.0	83%
6	Bucher et. al., 2020	1	1	0.5	1	0.5	1	5.0	83%
7	Debruyne et.al., 2017	1	1	1	1	1	0.5	5.5	92%
8	Kyzirakos et al., 2018	1	1	0.5	1	0	0	3.0	50%
9	van den Brink et al., 2018	1	0	1	1	1	0	4.0	67%
10	Folmer et al., 2018	1	1	0	0	1	1	4.0	67%
11	Ronzhin, 2019	1	0.5	1	1	1	1	5.5	92%
12	Forlmer et al., 2019	1	1	0.5	1	1	1	5.5	92%
13	Rowland et al., 2020	1	1	0.5	0	0.5	1	3.5	58%

19 most relevant studies (cont. 14 - 19):

ID	Author	C1	C2	C3	C4	C5	C6	Total score	%
		Relevance	Geospatial	Problems defined	Challenges defined	Best practices	NMCA domain		
14	Folmer et al., 2020,	1	1	1	1	0.5	1	5.5	92%
15	Yaman et al., 2021	1	0	1	1	0.5	0.5	4.0	67%
16	Issa et al., 2021	1	1	1	1	1	0	5.0	83%
17	Rowland et al., 2021	1	0	1	1	1	1	5.0	83%
18	Rowland et al., 2022	1	1	1	1	1	1	6.0	100%
19	Yaman et al., 2022	1	0.5	1	1	1	1	5.5	92%
		1.00	**0.79**	**0.72**	**0.82**	**0.65**	**0.79**	**4.6**	**76%**

Backwards snowballing studies references

From initial
146 studies

Picture Source: Cody Sperber

NVIVO

returned 1,706
reference studies

Studies published mostly in … journal

Table 3: Number of publications per journal from snowballing in present SLR

Journal	Publications
International Journal of Geographical Information Science	32
Computers & Geosciences	26
Semantic Web	18
ISPRS International Journal of geo-information	16
Environmental Modelling & Software	16
Transactions in GIS	15
Communications of the ACM	12
Computers	11
Semantic Web	9
GeoInformatica	9

~ 50%

LD studies, per author (snowballed from 43 eligable articles)

Table 4: Backward snowballing searches counted the references per author from the initial examination.

Author	Publications
Janowicz, K.	26
Lehmann, J.	23
Koubarakis, M.	23
Auer, S.	20
Kraft, M.	19
Bizer, C.	17
Corcho, O.	14
Hitzler, P.	14
Bereta, K.	14
Folmer, E.	13
Kyzirakos, K.	13

Most frequent concepts snowballed from 43 eligable articles (total of 25.511 concepts detected automatically)

Table 5: Frequency analyses of the words in eligible articles

Word	Count	%
data	7116	27.9%
spatial	2401	9.4%
linked	1491	5.8%
geospatial	1468	5.8%
information	1403	5.5%
web	1397	5.5%
semantic	1169	4.6%
RDF	946	3.7%
org	834	3.3%
level	778	3.0%

Most frequent concepts from 43 eligable articles (CONT.)

Word	Count	%
geo	756	3.0%
quality	674	2.6%
query	672	2.6%
datasets	659	2.6%
knowledge	658	2.6%
geometry	647	2.5%
GEOSPARQL	643	2.5%
ontology	624	2.4%
geographic	596	2.3%
model	579	2.3%

SYNTHESIS OF SLR

SMAs LINKED DATA PUBLICATION

12 key LD issues

from 19 studies

Issues

Table 6: Summary table for the synthesis of data collected from nineteen selected studies on issues of geospatial LD publishing

item	LD issues	description	study ID	count	%
1	interoperability	shared, open data accessibility, reuse, integration - combining different sources	3, 4, 8, 9, 12,15, 17	7	14,3%
2	vocabulary	common thesaurus for the domain terminology, multilingual terminology, cross-border terminology	2, 3, 4, 6, 7, 9, 18	6	12,2%
3	support	users support browsing, searching, querying, analyses, viewing	5, 9, 10, 12, 13, 17	6	12,2%
4	heterogeneity	disjoint semantically, structurally, and syntactically; diversity of contexts, terms definitions	1, 6, 9, 11, 14, 19	5	10,2%
5	ontology	knowledge, semantics, relations, explaining instances meaning to the outside world	4, 7, 9, 11, 18	5	10,2%
6	quality	LD publication quality metrics as completeness, updating, accuracy	1, 11, 16, 19	4	8,2%
7	readability	readability by humans and by machines	4, 6, 9, 15, 18	4	8,2%
8	big data	publishing large datasets, siloed data	5, 8, 9, 14	4	8,2%
9	efficiency	business issues of OLD, publication automation, ETL, time- and cost-efficient delivery, restrictions	8, 14, 17, 18	4	8,2%
10	data context	metadata - information about data	2, 4	2	4,1%
11	serialisation	provide data in different LD formats	4	1	2,0%
12	scalability	a scalable approach to linked data publication	18	1	2,0%

Answers to RQ

- RQ1 **Problems**: heterogeneity, siloed data, metadata, context.
- RQ2 **Solutions and requirements**: vocabularies, ontologies, metadata, standardised technologies.
- RQ3 **Priority data themes**: administrative units (7), place names (3), building identifiers (3), and addresses (2).

RQ4 NMCAs **Challenges** :

- setting development **priorities**,
- keeping pace with technological **trends**,
- compensating for the **financial costs of data sharing**,
- the **skills of staff** at NMCAs,
- the availability of **best practice guidance.**

RQ5: Reasons for the slow proliferation of LD

- **At public bodies slower adoption** of new technologies, than in private
- **Limited tooling**
- The **lack of knowledge** on managing **knowledge graphs (KG)**
- **Unable** to properly **advertise datasets' usefulness** to potential users, thereby **hampering reuse**
- The **lack of** sufficiently distinctive and functional **browsing and viewing facilities**
- **Publishers do not use LD themselves**, and
- LD fails to attract IT vendors due to **insufficient professional demand**.

Limitations of the study results

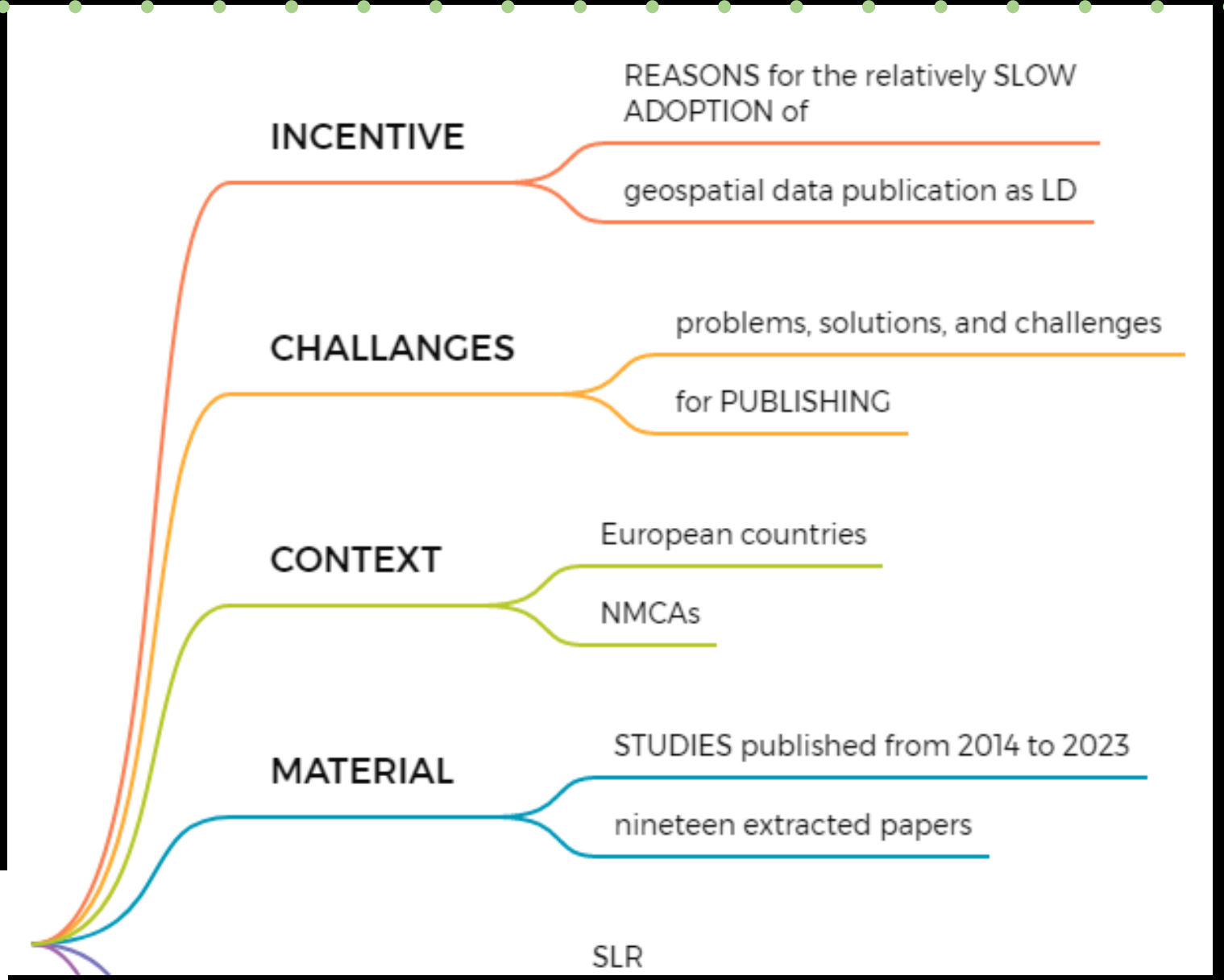
the study **did not address**:

- technologies,
- standards,
- processes, or
- legal issues

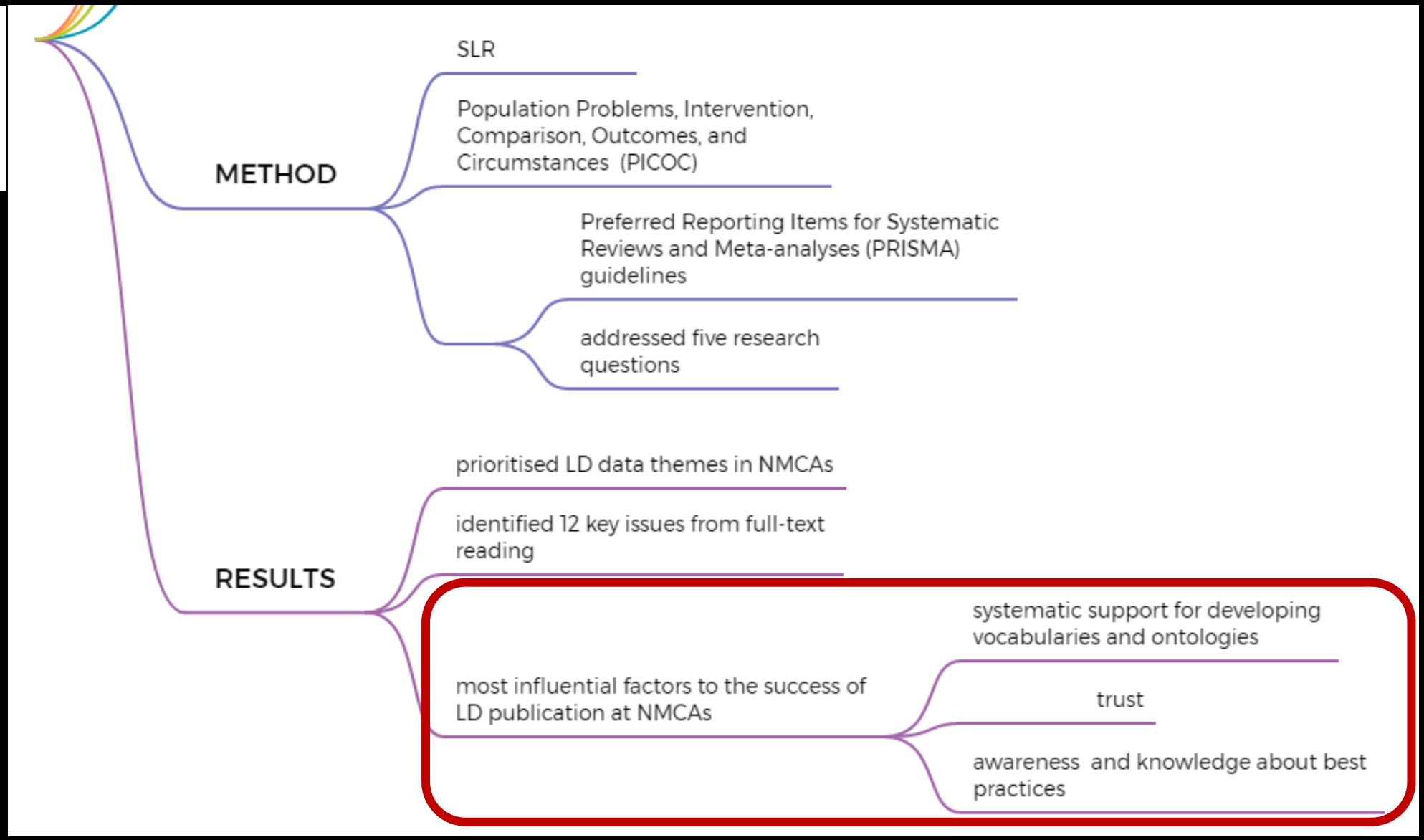
for LD publication at NMCAs.

CONCLUSIONS

GEOSPATIAL LINKED DATA PROLIFERATION IN NMCAS



GEOSPATIAL LINKED DATA PROLIFERATION IN NMCAS



GEOSPATIAL LINKED DATA PROLIFERATION IN NMCAS: SYSTEMATIC LITERATURE REVIEW

Slovenian research project V2-2295 (2022-2024)

Dr. Marjan Čeh
Chair of Geoinformatics and RE Cadastres