Misled by Data? Review of Data Sources in National Intellectual Capital Research

Hannele Orjala
School of Business and Management, Lappeenranta-Lahti University of Technology, Finland
hannele.orjala@stat.fi

Abstract: This paper is a review of National Intellectual Capital (NIC) literature that focuses on the documented use of data and the data sources used in the NIC literature. The topic is important as the NIC research is largely based on data analysis and thus the use of data and the data sources used ultimately shape the reality around what is the big picture of national intellectual capital, as it is understood today. While this is the case, questions about use of data and the data sources used have not been in the core of the research tradition. The review focuses on 57 systematically collected NIC articles with a documented data source, published between the years 1991 and 2018. The results show that the majority of data-based NIC research is concentrated around a set of often-used data sources, while the rest of the data-based NIC literature uses a fragmented set of data sources. New data sources are rarely utilized. The documentation of data and data source use in the literature leaves room for criticism. Researchers and the users of NIC analyses benefit, if they are able to evaluate the quality, the coverage, and the relevance of the data sources used as a basis of NIC analyses. To the best of our knowledge, this is the first time that data sources used in the NIC literature are the main focus of a literature review.

Keywords: National Intellectual Capital, data use, data sources, literature review, data bias, data quality

1. Introduction

Over the past two decades, research on national intellectual capital (NIC) has received a lot of attention within the community around intellectual capital (IC) research. NIC research has highlighted the importance of information in the societal level of value creation that is, when nations target competitiveness, wealth, well-being, and performance gains on the aggregate level. Among other things NIC-research has created a conceptual platform and methodologies to improve and to benchmark the progress of nations. Importantly, NIC-research has widened the understanding of the dynamics behind aggregate performance (of nations) and lifted the status of researching the effects brought about by changes in the structures of the societies, such as the effects of globalization and of technological development.

As the review does not take a position or exclude articles based on the definition of NIC used in the studied articles, a strict definition of national intellectual capital is not necessary for this review. However, for the purposes of this review an in vein with Pedro, Leitão and Alves (2018b) NIC is understood as the variety of intangible assets that give long-term advantages to a nation and interact with organizations, regions, or globally and that are able to produce future and equal benefits measured at all levels. Importantly, NIC is understood as research that focuses on the national level.

Advanced national data-ecosystems and data-management (institutions) are key elements of modern data-driven societies and as such create the physical bedrock of NIC - they enable high-level value creation and ensure and enable fact-based decision-making for all sectors of the society at organizational, industrial, regional, national, and global levels. They also function as a mirror of successful value creation, when measures and indicators based on information gathered are used in this context.

The aim of this paper is to study and answer to the question "what data sources (data ecosystems) have been used as the basis for the academic literature on National Intellectual Capital during the period between 1991 and 2018?" and to discuss the use of data and the sources in connection with national intellectual capital research. In order to reach this aim, this work presents a review of the academic NIC-literature from the perspective of data sources by first identifying 111 articles published on NIC and selecting the 57 articles, where the data source used is documented and reviewing them in detail. While data is widely utilized as the basis of analysis in the majority of NIC articles and while the use of data has previously been the subject of some criticism (see, e.g. Alfaro, López and Nevado, 2011; Chew, Sharma and Bontis, 2014; Käpylä, Kujansivu and Lönnqvist, 2012; Lee, Lin and Lin, 2017), studies focusing on the use of data and data sources used in the NIC research are non-existent. While meritorious literature reviews on NIC exist (see, e.g. Dumay, Guthrie and Puntillo, 2015; Labra and Sánchez, 2013; Pedro, Leitão and Alves, 2018a and 2018b; Petty and Guthrie, 2000), these reviews do

Reference this paper: Orjala, H., 2021. Misled by Data? Review of Data Sources in National Intellectual Capital Research. *The Electronic Journal of Knowledge Management,* 19(1), pp. 43-53, available online at www.ejkm.com

not specifically concentrate on the use of data or the data sources used in the literature. To the best of our knowledge, this is the first time that data sources used in the NIC literature are the main focus of a literature review.

The approach taken here highlights the importance of the use of data and the data sources used as a part of the national intellectual capital discussion. The question of the effective utilization of data and the indicators used in the research- and policy-making levels is discussed together with a discussion around the challenges with access to data. As such, this review is relevant to the stakeholders in national intellectual capital questions that deal with data- and indicator-related questions. For researchers this review identifies potential problems with the use of data sources and the effects of these potential problems on research outcomes. Researchers and the users of NIC analyses benefit, if they are able to evaluate the quality, the coverage, and the relevance of the data sources used as a basis of NIC analyses. For the data ecosystem actors, at national, private, and global levels, such as national statistical institutes this review highlights their role in ensuring the operational conditions of scholars and the NIC research community, not to mention public data-based decision-making. Furthermore, this article highlights the importance of the dialogue between data ecosystem actors and policymakers in securing the availability and quality of data for research and offers support for evidence-based decisions concerning data management questions for policy-makers.

We present a discussion about the possible and apparent problems with regards to data source and content documentation in the NIC literature and shortly discuss the risks and weaknesses associated with opaque data handling and other data related "less than perfect" modes of operation in research.

The rest of this article is structured as follows: first we shortly present the methodology used in the literature review and then we present the review-results. Then follows a literature-based discussion on the use of data-sources in the NIC literature and on observed problems connected to data-use. The paper is closed with a section that offers a short summary, presentation of conclusions, and some future directions for research.

2. Methodology

This literature review follows the state-of-the-art practice proposed by Webster and Watson (2002) and uses a paper selection process to gather relevant literature. Four article databases were utilized in collecting the articles written in English used: *Emerald Journals, EBSCO (all databases), Scopus Elsevier, and Science Direct*. Depending on the database filters "article", "peer-reviewed", "scholarly" were used, where available. Articles related to Intellectual Capital at macro, or national levels were included. The review covered years 1991 – 2018 (until the month of September).

Articles were searched based on their abstract with a number of variants of the key phrase "National Intellectual Capital" that included strings such as: "national intellectual capital", "national" and "intellectual capital", "nation" and "intellectual capital" and "country" and "intellectual capital", also the abbreviation "IC" was used. In summary, the search was narrowed down by:

1. Including only select subject areas: Scopus (Business, Management and Accounting, Social Sciences, Economics, Econometrics and Finance, Decision Sciences); Ebscohost (IC, human capital, knowledge management, economic development and competition, globalization, international trade, intellectual property, industrialization, information economy and technology, management, strategic planning and assets), for other sources the original search words were used without further filtering.
2. Including only English language articles
3. Including scholarly articles (articles, peer-reviewed, scholarly; depending on the archive)
4. Searching keywords (combinations) only from the document abstracts

The result of the initial abstract-level search was 848 articles – all these articles were scanned and exclusions were made, based on the article title pointing directly to exclusion criteria with relation to the focus of the level of analysis (firm, organization, region, university, research institute) that is, non NIC research focus. The included set contains one article with multiple-level analysis, where, based on author judgment the focus was strongly on NIC. Also, some purely technology-oriented papers and articles that concentrate on ethical issues were excluded.

PROCESS	SOURCES / NUMBER OF ARTICLES			
Intellectual capital search	EMERALD Journals 1 033 articles	EBSCOhost 2 606 articles	SCOPUS, Elsevier 3 885 articles	ScienceDirect 2 654 articles
Criteria 1-4	189 articles	238 articles	137 articles	340 articles
Merging results	848 articles			
Exclusion of non-NIC articles	213 articles			
Exclusion by abstract and full paper scanning	111 articles			
Articles with documented data source	57 articles			

Figure 1: Selection process used with the quantity of remaining articles after each step

The remaining 213 articles were again manually scanned by reading the abstracts, which resulted in a final total of 111 articles on NIC of which 101 we had access to in full text. The 111 articles were published in 63 different journals, Journal of Intellectual Capital (JIC) being the single most important source with a 25 % share of the NIC articles. All the articles selected were cross-checked with the Finnish Publication Forum (www.julkaisufoorumi.fi/en) listing of high quality academic publications and 91 of the articles are rated at least level 1 (basic academic). The rest of the articles were either journals under the level 1 or conference articles.

Of the 111 articles in 57 the data-source used is documented. This set of 57 articles is the main focus of this review. Of these 57 data-based NIC articles 80% were rated at least level 1 in the Finnish Publication Forum classification. Figure 1 visualizes the selection process used.

For each article the following information was extracted:

1. Database where the article is stored (one of the four databases searched)
2. Year of publication (1991 – 2018)
3. Main focus / research question (shortly)
4. Article type ("general NIC", "country comparison", "country analysis" or "special NIC focus") done after a holistic overview of the articles and based on that the creation of the four classes
5. Data source(s) used in the article (the sources)
6. Number of citations the article had received by early 2019
7. Relation to public sector data (if and how public sector data is used)
8. Presence of public sector data providers (if and when data is used)
9. Geographical coverage (# of countries in the analysis)
10. Country of the university of the main writer of the article

A synthesis and an analysis of the information collected from the selected papers gives a holistic picture of the data-use and the data sources used in the NIC literature and reveals the what are the most often used data sources. The collected information (points 1-10 above) for the 57 studied articles is available as a table in Appendix 1.

3. Results

This section presents the results from the review. We look at the set of 111 identified NIC articles and observe that fifty-four of the total 111 NIC articles do not document the data source used, the data source is unknown, or no data is used, which leaves fifty-seven (57) articles that disclose the source of data used and of which, forty-seven specifically describe the data set used (in addition to mentioning the source). In other words, roughly half (51,4%) of the selected NIC articles fall under the focus of this review.

We start by taking a look at how the used data sources are indicated in the reviewed articles – we do this by article type for which purpose, we create a taxonomy of four different article types (or classes), according to the main focus of the article identified during review of all the articles. The four classes are: "General NIC", "Country comparison", "Country analysis", and "Special NIC focus" (articles with focus in a specific topic, e.g., environment, culture, crises).

Forty (40) of the reviewed 57 articles are comparative or country specific analyses, seven special NIC focus, and ten general NIC articles. See Table 1 for detailed information.

Table 1: Information of the data sources used according to the main focus of the article

	# articles	A (General NIC)	B (Country comparision)	C (Country analysis)	D (Special NIC focus)
Data source info available	57	10	26	14	7
- information on data sets used	47	7	26	7	7

In many of the reviewed articles the way the used data sources are identified and described is imprecise, e.g., detailed information about the reference year(s) in the data is not often available. A data source can have a wide range of designations that changes from article to article and the databases used are typically unclearly and vaguely specified. What is typically identified is the "department" or "division" from which the data comes from, within a larger data-providing organization. Referencing and listing of data sources as references is typically non-transparent or incomplete.

The above-mentioned problems with documenting data sources used are prevalent for all types of data sources, while some commonly used sources can be specifically identified with a less precise referencing, most often missing and imprecise documentation of data sources does not allow a complete understanding of what the actual source used had been. In the studied 57 articles, where information about the data source used was available, in 35% information detailed enough to allow the identification of the specific data source (within, e.g., a larger source organization) and the data-sets used was missing. Replication of previous research results and validation of results becomes remarkably easier, when the data sources used are documented properly and transparently. The lack of specificity in the information about the used data sources information may be an indication of a low attention paid to the impact of data on the research results in this field.

The typically and most often used data sources in the studied articles that indicate the data source used, include the various data bases housed within The World Bank organization. Other often used sources include The World Economic Forum with reports and indexes, and the United Nations with its various data providing organizations. More than half of the 57 articles use at least one of the above three as a data source. The three main data sources are often complemented with other data providers with index data. It can be observed that the most commonly used sources are also well-known by policy-makers (Salonius and Lönnqvist, 2012). Table 2 lists some of the organizations often used in the NIC literature as data sources and Appendix 1 offers a more complete listing of the used data sources with an article by article documentation.

Table 2: The most often used data sources

Organization	**Examples of data sources and indicators / indexes published**
The World Bank (WB)	World Bank (WB), World Bank group (WBG), WB's World development indicators (WDI), World Bank database (WB DB), Knowledge economy Index (WBG KEI), World Bank's Knowledge Assessment Method database (WB KAMD)
The World Economic Forum (WEF)	World Economic Forum (WEF), Global Competitiveness Report/WEF (GCR), Global competitiveness index (GCI), Network Readiness Index (NRI)
The United Nations (UN)	United Nations (UN), UN development Programme (UNDP), Human development index (HDI), Human development Report (UNHDR), UN Public Administration Country Studies (UN/PA), Millennium Development Goals (UN MDG), United Nations Educational, Scientific and Cultural Organization (UNESCO), World Health Organization (WHO), ILO survey data, International Labour Organisation (ILO)

Further statistics about the data sources used in the 57 articles are shown in Table 3, where one can see that the aforementioned three sources are the most frequently used data sources, while international and national

statistics organizations form the group of the second most used data sources. We note that the percentages in Table 3 do not add to one hundred as many articles use several data sources.

Table 3: Data sources used in National Intellectual Capital measurement

Main data sources in the articles; classification used in the Appendix 1 indicated in (•)	articles n=57	%
The World Bank (a)	24	42
The World Economic Forum (b)	21	37
The United Nation (c)	21	37
International Statistical offices (OECD and Eurostat) (e)	17	30
International Institute for Management Development (d)	15	26
National Statistical Offices (f)	14	25
European Commission and other European bodies (g)	6	11
Other banks and financial institutions (h)	5	9
Interview (j)	4	7
New data sources (i)	1	2
Other index and other sources (k, l)	18	32

It is common that multiple or combined data sources are used and information is collected from a number of sources – in the 57 articles reviewed the number of different data sources mentioned was 74. Table 4 shows statistics about the number of sources used per article in the studied population. Article specific information and classification can be found from Appendix 1.

Table 4: Number data sources used per an article

Number of data sources used	Number of articles (sample 57)	Percent of articles, %
1	13	22.8
2	13	22.8
3	13	22.8
4	5	8.8
5	5	8.8
6+	8	14.0
Total	57	100

It can be seen that 23% of articles used one, two, or three data sources – this already constitutes more than two thirds of all articles studied and means that about thirty percent of the articles used four or more data sources. Fourteen percent of the reviewed articles (8) used six or more data sources. It is common that the top data sources are used simultaneously, all three were used in eight articles and World Economic Forum and World Bank together in twelve articles.

The countries studied and the number of countries simultaneously studied varies between articles. The most common area studied in the reviewed articles is the European Union, which accounts for over one fourth of research, with 16% of the articles looking at all EU countries. Eleven percent of the articles studied a smaller group of EU countries, or one EU country alone. Fifty-six percent (56.1%) of the articles had more than 20 countries in their sample, while the largest number of countries simultaneously studied was 148 (Seleim and Bontis, 2013). Under ten countries were focused on in ca. thirty-seven percent (36.9%) of the studies. Table 5 shows statistics of the geographical coverage of the reviewed articles.

Roughly a quarter (24.5%) of the studied articles utilized data from more than 50 countries in their research. These studies were from nine different main authors with altogether fourteen co-authors, with López Ruiz being the most prolific single author of high-number multi-country analyses. According to a count based on first author affiliation, documented data-based NIC research is most active in Spain (10 articles) followed by Taiwan and by Finland (7 articles). It is somewhat interesting to note that USA, Australia, and the UK are not among the top countries.

Table 5: Geographical coverage of reviewed articles

Number of countries in the studies	Number of studies	Percentage of all
>100	4	7.0%
50-99	10	17.5%
20-49	18	31.6%
10-19	4	7.0%
3-9	5	8.8%
1-2	14	24.6%
unknown/none	2	3.5%
Total	57	100%

One area of interest to this research is to find out to what extent the studied literature uses public sector data and the (national) public sector data providers as data sources. Table 6 shows the public sector data use in the studied literature. It can be seen that the use of public sector data is prevalent in all types of articles and acts as the base of analysis in over two thirds of the articles overall. Use of public sector data is most pronounced in country specific and country comparison focused articles.

Table 6: Public sector data use in the surveyed NIC literature

	No public sector data used	Some public sector data used	Public sector data used as the base	Total
A (General NIC)	18.2% (2)	27.3% (3)	54.5% (6)	100% (11)
B (Country comparison)	7.7% (2)	15.4% (4)	76.9% (20)	100% (26)
C (Country analysis)	7.1% (1)	21.4% (3)	71.4% (10)	100% (14)
D (Special NIC Focus)	0% (0)	50.0% (3)	50.0% (3)	100% (6)
Total	8.8% (5)	22.8% (13)	68.4% (39)	100% (57)

The role of public sector data providers in the studied literature, in the light of the prevalent use of public sector data, can be expected to be important. All in all, this expectation is not fully met, as the public sector data providers have a strong visible role in less than half (45.6%) of the studied articles overall. Table 7 shows the role that public sector data providers have in the studied articles. It can be seen that one fifth (21.1%) of the articles do not mention public sector data providers at all and that they are given a strong visible role typically only in approximately half of the country specific and country comparison focused articles. Interestingly more than one third (35.7%) of the country comparison focused articles omit mentioning public sector data providers.

Table 7: Presence of public sector data providers in data-based NIC literature

	No role / not mentioned	Minor role / recognized	Strong visible role in an article	Total
A (General NIC)	27.3% (3)	45.5% (5)	27.3% (3)	100% (11)
B (Country comparison)	12.0% (3)	36.0% (9)	52.0% (13)	100% (25)
C (Country analysis)	35.7% (5)	7.1% (1)	57.1% (8)	100% (14)
D (Special NIC Focus)	14.3% (1)	71.4% (5)	14.3% (1)	100% (7)
Total	21.1% (12)	33.3% (19)	45.6% (26)	100% (57)

It seems that even if public sector data is often used in the literature, public sector data providers do not commonly have an important role outside providing the data. That is, there seems to be a low “state of interaction” between the NIC research community and the public sector data providers, which may be a tell-tale sign of weak collaboration between these “two camps” in terms of developing the public sector data collection and indicators that are important for research purposes.

The utilization of main data sources seems to of have been similar throughout the studied period in all types of articles with the exception of some special aspect focused articles that typically used indexes and special data sources, even new big data sources related to the chosen specific topic. Only 0.9 % of the articles recognize the options to renew also the data sources with big data/smart data (see Vahanyan, Vahanyan and Ghazaryan, 2018). Another “casual” observation about the Intellectual Capital literature is that what one can see is that the

majority of recent studies on intellectual capital (IC) concentrate on other areas than national, or macro level intellectual capital. The NIC proportion of all IC studies was only 1.1 % according to our search. Pedro and others estimate that the proportion of studies with an organizational focus is as high as 92.7% of the IC literature and that the proportion of NIC studies is only 4% and that the rest of the studies (3.3%) concentrated on a regional level (Pedro, Leitão and Alves, 2018a, 2018b). The difference between the findings can be attributed to differences in the way the searches were conducted. NIC literature is a niche of less than five percent of the overall IC literature.

While the overall weight of the NIC is not great, the selection and use of data sources is an important issue for NIC research as NIC research is often data-based and the NIC research has a strong history of using data. The theoretical formation of NIC is also largely based on data analysis, which makes the clarity and transparency of the sources used important.

4. Discussion about the data-related problems and relevance-issues observed in the literature

In this section, we discuss problematic issues connected to data that underlie the data-based research on national intellectual capital. As observed above, there are critical voices in the NIC research community that point to problems with data – main problematic issues identified in literature are the ill-availability of (relevant) data and data quality issues. Observed are also inconsistencies and biases in the use of data. The observed problems with data and the use of data have consequences with regards to the credibility and replicability of the results obtained. The real-world relevance of, e.g., indicators used in describing the state of and the progress (of NIC) in a specific nation and comparatively is also an issue of importance.

4.1 Observed data-related problems and consequences

One of the most pronounced observed shortcomings and limitations to NIC research (and reporting) with regards to data is the *lack of availability of data* for a number of nations (Alfaro, López and Nevado, 2011; Bounfour, 2003; Chew, Sharma and Bontis, 2014; Frederick and Mcllroy, 1999; Jednak, Dmitrović, and Damnjanović, 2017; Käpylä, Kujansivu and Lönnqvist, 2012; Lin and Edvinsson, 2008; Lin and Lin, 2008; Lin and Edvinsson, 2012; López *et al.*, 2011; Park and Oh., 2018; Piekkola, 2018; Stachowicz-Stanusch, 2013). The lack of availability of data has had serious consequences for NIC research in that it creates a *country bias and affects the research designs* used.

Country bias in this context means that the available data has guided data-based research to take place mostly and even only on countries for which there is ample data available and *consolidated the use of data from the incumbent data providers* (see Herciu and Ogrean, 2015; Stachowicz-Stanusch, 2013). Research designs used have been such that they are not necessarily optimal from the point of view of NIC research and reporting, but such that are able to utilize the data that is available for the purposes of NIC research and reporting (Käpylä, Kujansivu and Lönnqvist, 2012; Stachowicz-Stanusch, 2013). In other words, the availability of data and the type of data that is available may have (remarkably) limited the universe of usable methods (López *et al.*, 2011) and in effect steered the type of NIC research that has been conducted. It has been observed that the relatively small number of countries for which data is available often makes it difficult to apply advanced comparative statistical analysis methods to test data-based hypotheses (Lee, Lin and Lin, 2017). The statement has been made that "*the available data apparently determines the understanding of NIC*" (Käpylä, Kujansivu and Lönnqvist, 2012).

In situations, where data is available, there are often limitations to it, such as the lack of longitudinal data (Chew, Sharma and Bontis, 2014; Seleim and Bontis, 2013), or there might be significant year to year changes with regards to the data sources and in the available statistical data (Tomé, 2011) – both are issues that typically create problems in time series- and comparative analyses. It is typical that data sets reflect the past national IC performance, rather than the current or the future status of a nation (Chew, Sharma and Bontis, 2014; Lin and Edvinsson, 2008; Lin and Lin, 2008; Lin and Edvinsson, 2012; Pedro, Leitão and Alves, 2018a). Data limitations can also affect the reliability of any analyses that attempt to establish causal relationships (see, e.g. Lee, Lin and Lin, 2017; Seleim and Bontis, 2013). Similarly, the credibility of comparative analyses based on (single) country indicators that are composed of numbers that do not consistently (year to year) come from the same sources and that are not constructed in an identical way is suspect (Radjenovic and Krstic, 2017). This problem is similar to the common problem with the lack of *consistency* between quality criteria, accuracy requirements, and contents of data that different (national and regional) statistical systems require from the collected data (Lin

and Edvinsson, 2012). Because the manner in which (certain indicator) data is gathered varies from country to country, as a result there are typically inconsistencies and inaccuracies in comparisons and in analyses (Hervas-Oliver and Dalmau-Porta, 2007; Lin and Edvinsson, 2008; Lin and Lin, 2008) sometimes the differences can render comparisons pointless (Łukasiewicz, 2013; Navarro, Lópes and Nevada, 2011). Typically the institutions in each country can determine the set of (most suitable) indicators for which data is collected and the way in which the data is collected – although this practice may capture the unique traits of each country, it efficiently prevents (relevant) cross-country comparisons (see Käpylä, 2012).

Generally, it is noticed that it *is difficult to find the correct composition of variables to address and measure specific phenomena* (Lin and Edvinsson, 2008; Tomé, 2011). The selection of indicators used in indexes that try to capture the development of different phenomena most often also depend on data availability, rather than on (maximizing the) substantive content of the given index (Łukasiewicz, 2013). An easily measurable and definable format of knowledge is appreciated, as the dominant NIC discourse relies on measurement, numbers, indices, and quantitative comparisons and not on qualitative descriptions, hermeneutic understanding, or in-depth and detailed case-studies (Käpylä, 2012) and again, the typical source of information for different indexes are the few well-established and well-known sources.

It has been observed that in some cases the proposed composite-indexes may contain elements based on inherently subjective judgment that may create biases such as over- or underestimation of the results due to, e.g., the use of difficult to justify weights for variables (see Önsel *et al.*, 2008). Also the use of proxy variables in indexes is common (López *et al.*, 2011; Edvinsson and Bounfour, 2004; Bounfour, 2003); it is well known that proxies are not completely representative of the variables for which they are proxies, therefore, if a composite-index is composed of a large number of proxies, the results may become inaccurate. Proxies are likely under- or overestimating the phenomenon they are used to measure (Ståhle and Ståhle, 2012).

Generally, the use of data in form of measurements, indicators, indices, and international comparisons easily creates an *illusion of objectivity and universal applicability* (Käpylä, 2012; Käpylä, Kujansivu and Lönnqvist, 2012). If this is put in the light of data-related problems that NIC research and reporting is observed to have, it can be observed that completely trusting the reported results may lead to imperfect decisions.

Pedro and others (Pedro, Leitão and Alves, 2018a) estimate that due to the difficulty in obtaining suitable data, the growth in the number of empirical studies related to NIC is not as rapid as it would be under better circumstances. They further estimate that the lack of uniformity in indicators and data collection practices in different levels of IC is the reason for the relatively low number of empirical studies on NIC (Pedro, Leitão and Alves, 2018a). Ståhle and Pöyhönen state that the "consequences of the data and statistical indicators formation is seen as a drawback for the IC research tradition" (Ståhle and Pöyhönen, 2005).

4.2 Relevance issues related to data in NIC

The explanatory power of NIC indicators varies with time as a function of the developmental stage of the society that is being measured. For this reason it is important to review and update the NIC measures used for a given society regularly, or to perform measurement with a wide-enough selection of measures, in a way that guarantees that changes are captured, even if the structure and the drivers that drive a society forward change. Typically, the more mature an indicator is, the less explanatory power it has with regards to the development of NIC – this saturation is a testament to the constant development that takes place in what is considered to be a part of NIC and an indicator that is considered to be a growth driver at one point of time may "dry out" and outlive its usefulness as a measure of development (Salonius and Lönnqvist, 2012; Ståhle and Bounfour, 2008). When, for example, the penetration of mobile devices rises high enough, it no longer serves as an indicator of economic growth or performance, similarly circulation per capita of newspapers has lost its measurement value due to the media revolution.

Intangible goods are most likely the main type of IC that is making a large difference in the development gap between developed and developing countries. The impact of intangible goods on NIC is hard to measure, because a large part of that impact is in terms of hidden wealth (López et al., 2011). This also means that if we are unable to measure the impact intangibles have on NIC we most likely are not very good in measuring the NIC gap between countries either. The same kind of problem is associated more generally with all indicators that are only poorly able to capture the whole picture of what they endeavor to measure.

This boils down to noting that when societies change, the way the development in the societies is measured must change with it. That is, the way change takes place changes and the measuring instruments designed for the past ways change takes place may become irrelevant in measuring the new ways change takes place. This means that the way measurement is made should be reviewed and updated regularly to ensure the relevance of the measures. Changing the way measurements are made may cause problems of comparability of the gathered data over time, exactly as was observed above, within the discussion of the data-related problems and their consequences. It becomes increasingly important to understand that if there is an evolution in how change takes place and in the measures used to capture the qualities and size of the change, then documenting the measures and designing "backwards compatibility" becomes very important for the sake of allowing at least some credible longitudinal data to be collected. Put in another way, it is important to conserve core measures that are kept unchanged in how they are measured and reported throughout the changing forms of development. One must also note that intellectual capital is contextual and when the context changes the form and format of intellectual capital also and naturally changes (Ståhle and Bounfour, 2008).

What can be said is that the relevance of the measures used in understanding the state of NIC depends to a large extent to the fit of the NIC measures to the actual status quo in the country in which the measurement is made. It makes no sense to use the same measures in countries with vastly different developmental stage – where it makes sense to measure "literacy" in developing countries it makes sense to measure both "literacy" and "media- or data-literacy" in highly developed countries, because "literacy" is a saturated NIC measure for highly developed countries. To retain the relevance of NIC reporting and research there must be an evolution of measures and a backwards compatibility in the sense that also transitions from one developmental stage to another in a country could be observed also through the data collected. This also means that in designing measurement, the values of the future must be allowed to be different from the values that are accepted now (Käpylä, 2012), one can refer to the change of view with regards to strict environmental policies and how they today are considered to be indicators of positive development, where they were first considered to hinder economic development (Stavropoulos, Wall and Xu, 2018).

Maintaining NIC relevance most likely requires tools for making sure that relevant and "correct" indicators are used, there is however no consensus on evaluation models for NIC indicators (Radjenovic and Krstic, 2017). This is a clear path for further research and action in the NIC community and a challenge to the national and international data providers.

5. Summary, Conclusions, and Future Directions

This article presents a review of the academic National Intellectual Capital literature that focuses on the use of data and on the data-sources used. The review consists of presenting data collected from fifty-seven (57) articles from between the years 1991 and 2018 that have documented their data sources and discussion about the data-related problems observed in the reviewed literature.

The results show that number of different data sources used is rather high as a whole, but a small number of commonly used "preferred" data sources dominate in the literature. More than half of the studied articles use three or less data sources simultaneously, while only fourteen percent use six or more data sources. The number of countries analyzed in a paper varies between a set of papers that concentrate on one or two countries and a set of papers that study more than fifty countries – both of these groups represent about a quarter of the studied articles. Over seventy percent of all papers are either single country analyses or country comparisons. European Union is in the focus of a large number of articles, most likely due to the high availability of data.

The reviewed research is to a high degree based on data from public sector data providers, where about seventy percent of them are mainly based on public sector data and over ninety percent are at least partly utilizing public sector data. The role attributed to public sector organization in the studied is strong, while it is not as pronounced as the use of public sector data. There may be an opportunity for increased collaboration around research issues with the public sector data providers for the NIC community.

The review of literature shows that authors (in the literature) seem to feel that there is much to be desired with regards to the documentation connected to data use in both, documenting the data sources used and in documenting how the data used was actually used. Over one third of the data-based articles fail in documenting

the data source in a transparent way. The problems with documentation hamper possibilities of replicating research, which is an important issue with scientific data-based research.

In the literature observed problems regarding data use in NIC literature include the ill-availability of relevant data and data quality issues. Furthermore, inconsistencies and biases in the use of data may be found. The above may cause the applicability and credibility of obtained results suffer and in case the results are used as the basis for decision-making the problems may cause misleading decisions to be made. Availability of the data has likely determined data-based NIC research to a large degree and has therefore also affected the obtained results and the understanding of the NIC status quo.

A way to resolve the observed shortcomings is to require closer documentation of the data and data sources used in the research publication phase. If data is well documented then research that builds on well documented sources is able to update previous findings based on the same type of data used in the same way and by doing so ensuring longitudinal comparability, which at this time is also one issue that has been observed to hold challenges.

We want to point out some limitations to this work, first of all research on NIC is a marginal niche in within the otherwise quite abundant research on intellectual capital, therefore the results presented here can hardly be generalized beyond NIC research. While we have tried to ensure that all the most relevant articles are included we want to point out the possibility of having inadvertently excluded some relevant articles, however, we feel that the results are representative of the NIC research literature as a whole. We have not contacted the authors of the previous literature to study their data use any further – this may be an interesting avenue for further research. When choosing articles we have not used a multi-expert based (inter-rater) method, all choices are those of the authors.

Future research about the data-use and data sources of NIC that is needed from the point of view of better understanding the NIC status quo and development includes research into the development of well-functioning measures and indicators for the various NIC issues that the community studies and research that explores the state of NIC in different nations. As there is constant development in the world of intellectual capital, fueled by the ever-expanding new ways of using technology, the measures used to capture the status quo must constantly develop with the times, on the other hand, nations in different stages of development require different measures and indicators – mapping measure / indicator fit with stage of development would be a fine addition to NIC literature. There is also room for research concerning the connections between the NIC research community and the national data policies.

As an answer to the question “Misled by data?” we say “No”. Not misled, perhaps led; and while being led, we must be vigilant so that we are not being misled.

Acknowledgement

This research acknowledges the support from the Finnish Strategic Research Council project MFG40 grant numbers 335980 and 335990.

References

Alfaro, J., López, V. and Nevado, D., 2011, The relationships between economic growth and intellectual capital: A study in the European union, *Acta Oeconomica*, 61(3), pp. 293-312.

Bounfour, A., 2003, The IC-dVAL approach, *Journal of Intellectual Capital*, 4(3), pp. 396-413.

Chew, A., Sharma, RS. and Bontis, N., 2014, Intellectual wealth of nations revisited: Operationalising the value cycle of innovation, *Knowledge & Process Management*, 21(1), pp. 1-12.

Dumay, J., Guthrie, J. and Puntillo, P., 2015, IC and public sector: a structured literature review, *Journal of Intellectual Capital*, 16(2), pp. 267-284.

Edvinsson L. and Bounfour, A., 2004, Assessing national and regional value creation, *Measuring Business Excellence,* 8(1), pp. 55-61.

Frederick, H.H. and McIlroy, D., 1999, New Zealand and its competitors in the knowledge economy, *Telematics Informatics,* 16(4), pp. 177-217.

Herciu, M. and Ogrean, C., 2015, Wealth, competitiveness, and intellectual capital – sources for economic development, *Procedia Economics and Finance*, 27, pp. 556-566.

Hervas-Oliver, J. and Dalmau-Porta, J. I., 2007, Which IC components explain national IC stocks? *Journal of Intellectual Capital*, 8(3), pp. 444-469.

Jednak, S., Dmitrović, V. and Damnjanović, V., 2017, Intellectual capital as a driver of economic development *Economic Review: Journal of Economics & Business,* 15(2), pp. 77-84.
Käpylä, J., 2012, Towards a critical societal knowledge management, *Journal of Intellectual Capital*, 13(4), pp. 288-304.
Käpylä, J., Kujansivu, P. and Lönnqvist, A., 2012, National intellectual capital performance: A strategic approach, *Journal of Intellectual Capital*, 13(4), pp. 343-362.
Labra, R. and Sánchez, M.P., 2013, National intellectual capital assessment models: A literature review, *Journal of Intellectual Capital*, 14(4), pp. 582-607.
Lee, I., Lin, C.Y.Y. and Lin, T., 2017, The creation of national intellectual capital from the perspective of Hofstede's national culture, *Journal of Intellectual Capital*, 18(4), pp. 807-831.
Lin, C.Y.Y and Edvinsson, L., 2008, National intellectual capital: comparison of the Nordic countries, *Journal of Intellectual Capital*, 9(4), pp. 525-545.
Lin, C.Y.Y. and Lin, T. Y., 2008, National intellectual capital: Exploring Taiwan's standing, *International Journal of Learning and Intellectual Capital*, 5(3-4), pp. 311-331.
Lin, C.Y.Y. and Edvinsson, L., 2012, National intellectual capital model and measurement, *International Journal of Knowledge-Based Development*, 3(1), pp. 58-82.
López, V.R., Nevado, D.P., Navarro, J.L.A., Badea, L., Grigorescu, A. and Voinea, L., 2011, Measurement of national nonvisible wealth through intellectual capital, *Romanian Journal of Economic Forecasting,* 14(3), pp. 200-212.
Łukasiewicz, G., 2013, Macroeconomic determinants of intellectual capital quality, *Education of Economists & Managers*, 27(1), pp. 9-23.
Navarro, J.L.A., Lópes, V.R. and Nevada, D.N., 2011, Estimation of intellectual capital in the European Union using a knowledge model, *Zbornik Radova Ekonomskog Fakultet Au Rijeci*, 29(1), pp. 109-132.
Önsel, Ş., Ülengin, F., Ulusoy, G., Aktaş, E., Kabak, Ö. and Topcu, Y., 2008, A new perspective on the competitiveness of nations, *Socio-Economic Planning Sciences*, 42 (4), pp. 221-246.
Park, E.G. and Oh, W., 2018, Trust, ICT and income: Their relationships and implications, *Online Information Review*, 42(6), pp. 268-281.
Pedro, E., Leitão, J. and Alves, H., 2018a, Intellectual capital and performance: Taxonomy of components and multi-dimensional analysis axes, *Journal of Intellectual Capital*, 19(2), pp.407-452.
Pedro, E., Leitão, J. and Alves, H., 2018b, Back to the future of intellectual capital research: a systematic literature review, *Management Decision*, 56(11), pp. 2502-2583.
Petty, R. and Guthrie, J., 2000, Intellectual Capital literature review: Measurement, reporting and management, *Journal of Intellectual Capital*, 1(2), pp. 155-176.
Piekkola, H., 2018, Broad-based intangibles as generators of growth in Europe, *Economics of Innovation & New Technology,* 27(4), pp. 377-400.
Radjenovic, T. and Krstic, B., 2017, Measuring intellectual capital of national economies, *Ekonomika*, 63(2), pp. 31-43.
Salonius, H and Lönnqvist, A., 2012, Exploring the policy relevance of national intellectual capital information, *Journal of Intellectual Capital*, 13(4), pp. 331-342.
Seleim, A. and Bontis, N., 2013, National intellectual capital and economic performance: Empirical evidence from developing countries, *Knowledge & Process Management*, 20(3), pp. 131-140.
Stachowicz-Stanusch, A., 2013, The relationship between national intellectual capital and corruption: A cross-national study, *Journal of Business Economics & Management*, 14(1), pp. 114-136.
Stavropoulos, S., Wall, R. and Xu, Y., 2018, Environmental regulations and industrial competitiveness: Evidence from China, *Applied Economics*, 50(12), pp. 1378-1394.
Ståhle, P. and Pöyhönen, A., 2005, Intellectual capital and national competitiveness: A critical examination. Case Finland, in *6th European Conference on Knowledge Management proceedings in University of Limerick, Ireland, 2005*, pp. 575-583.
Ståhle, P. and Bounfour, A., 2008, Understanding dynamics of intellectual capital of nations, *Journal of Intellectual Capital*, 9(2), pp. 164-177.
Ståhle, S. and Ståhle, P., 2012, Towards measures of national intellectual capital: An analysis of the CHS model, *Journal of Intellectual Capital*, 13(2), pp. 164-177.
Tomé, E., 2011, Intellectual capital reporting: A conceptual study with application in the Portuguese case, *International Journal of Learning and Intellectual Capital*, 8(4), pp. 436-458.
Vahanyan, G.A., Vahanyan, H. and Ghazaryan, M., 2018, Interactive innovative tool for early diagnosis of global pre-crisis processes (based on measurement and assessment of the virtual intellectual capital, *Journal of Intellectual Capital*, 20(2), pp. 190-207.
Webster, J. and Watson, R.T., 2002, Analyzing the past to prepare for the future writing a literature review, *Management Information Systems Quarterly,* 26(2), pp. 13-23.

Appendix 1

Full listing of the reviewed literature with detailed classification is available on-line at https://lutpub.lut.fi/handle/10024/162353