ACIG
APPLIED CYBERSECURITY & INTERNET GOVERNANCE

NASK

Structured Field Coding and its Applications to National Risk and Cybersecurity Assessments

William H. Dutton | Oxford Martin School, Oxford University, UK, ORCID: 0000-0002-0141-6804
Ruth Shillair | Department of Media & Information Studies, Michigan State University, USA, ORCID: 0000-0003-0341-9096
Louise Axon | Department of Computer Science, Oxford University, UK, ORCID: 0000-0001-5979-7630
Carolin Weisser | Harris Global Cyber Security Capacity Centre, Oxford University, UK

Abstract

Data on cybersecurity capacity building efforts is critical to improving cybersecurity at national levels. Policy should be informed not only by measures that allow internal assessment of strengths and weaknesses that enable cross-national comparisons. The International Telecommunications Union (ITU) and its Global Cybersecurity Index (GCI) has used a standardized survey that has been adapted and used in multiple national assessments by the Global Cyber Security Capacity Centre. This adaptation includes an addition of open field coding assessments that rely heavily on trained experts and interactions with national focus groups. These assessments are checked using multiple coders to increase reliability and reduce bias. This process of 'structured field coding' (SFC) is an approach to collecting and coding observations based on multiple methods, quantitative as well as qualitative. This approach differs from open field coding in providing a set structure for coding observations from the field based on established frameworks for assessment. The SFC process is explained along with a discussion of the origin and the advantages and limitations of this methodological

Received: 12.06.2023

Accepted: 18.10.2023

Published: 27.10.2023

Cite this article as:
W.H. Dutton, R. Shillair, L. Axon, C.W. Harris "Structured Field Coding and its Applications to National Risk and Cybersecurity Assessments," ACIG, vol. 2, no. 1, 2023, DOI: 10.60097/ACIG/162857

Corresponding author:
Ruth Shillair, Department of Media & Information Studies, Michigan State University, USA, ORCID: 0000-0003-0341-9096; E-MAIL: shillai7@msu.edu

approach. It can be used in a variety of studies but is presented here as a means to integrate data for cross-national comparative analyses. Its application to improving the reliability and validity of data collection across a region, such as the EU, would help stakeholders evaluate where they should invest resources to improve their cybersecurity capacity.

Keywords
cybersecurity capacity building; structured field coding; cybersecurity analysis; multi-methods security research

1. Introduction

Efforts to build a nation's capacity to withstand cyberattacks and other risks to cybersecurity contribute to a nation's security and economic vitality. Almost all our modern systems: communication, economic transactions, record keeping, and critical infrastructure are controlled through computerized systems. These systems increase efficiency and speed, yet simultaneously they can be an Achilles heel, as each system introduces additional surfaces that are vulnerable to attack. Thus, it is important to build an ecosystem that is robust and resilient to cyberattacks. This process of cybersecurity capacity building is systematic, touching many societal dimensions [1]. The investment in cybersecurity capacity building pays back in the function of critical infrastructure and economic vitality [1, 2] as well as new legal and policy frameworks. While efforts to proactively address security problems seem intuitively valuable, they are new, meaning there is relatively little research on whether they achieve their intended objectives. This paper takes a cross-national comparative approach to determine whether there is empirical support for investing in capacity-building. It reflects field research from 73 nations as well as comparative data analysis. These efforts recognize that improved cybersecurity capacity is a multi-dimensional effort, it includes not just technological improvements, but also improvements in education, awareness, and training [3].

The first step in capacity building efforts and cybersecurity policy initiatives is often establishing basic "norms" as to what are best practices across these dimensions [4]. Both regional [5] and international efforts have worked to develop norms based on input from interdisciplinary teams. One of the larger efforts is the International Telecommunications Union's (ITU) Global Cybersecurity Index (GCI) [6]. The GCI uses questionnaires and expert advice to rate countries on legal, technical, organizational strategies and plans, levels of

international cooperation, and other capacity building measures to create an overall ranking. A similar National Cyber Security Index (NCSI) has been developed by the non-profit consultancy organization, the e-Governance Academy Foundation, based in Estonia as a joint initiative of the Government of Estonia, the Open Society Institute, and the United Nations Development Program [7]. The NCSI is one project in its cybersecurity program of activities, which includes consulting projects for various nations along with its development of an index for a growing number of countries [7]. The index "measures the preparedness of countries to prevent cyber threats and manage cyber incidents" to compose a database that is publicly available [8]. The evidence is either provided by the nation's government officials, an organization or individual, or by the NCSI team through desk research on legal acts, official documents, and official websites. These sources are used by an expert group at NCSI to make summary assessments on multiple aspects for each nation [8]. In such ways, the selection of index items is a dynamic process that is founded on a solid understanding of published research, comparison to similar studies, and adaptation to emerging trends.

Once index items and norms of assessment are established the next challenge is the design of a methodological process. The research process should assess the performance of nations over time and allow comparison with other nations. Evidence is often based on indicators drawing from a multitude of sources, ranging from different institutions and departments as well as different methods, such as in-depth interviews, questionnaires and surveys, and the aggregation and interpretation of data collected for other purposes, such as national census records. Not only is the evidence collected from multiple sources, but also the outcomes of the assessments are critical to multiple stakeholders, each of whom have a strategic interest in how different sectors or nations are rated. There are also the challenges of assuring norms assessment is applicable to countries at different stages of economic and technical development, particularly as those with less experience and centrality of Internet use may be more vulnerable to cyberattacks [9, 10]. Furthermore, there are concerns when developed nations are assisting developing nations in cybersecurity development, as these efforts might be construed as supporting the geopolitical interests of developed nations [11], which could be branded as digital neo-colonialism rather than a win-win strategy. Thus, there is a need to develop tools that can be readily adopted by and scaled to small or large countries, validated and employed by the adopting nation, and yet reliable and standardized sufficiently to be comparable across nations for accurate and actionable insights.

Given limited resources, and many sources and types of data within each nation, the question is: How national assessments be done efficiently and in a reliable and valid manner that can be replicated and compared over time and with other nations? Structured field coding is an answer to this question. It is an approach which increases objectivity in cybersecurity capacity building assessments. Given its quantitative basis, it can also be linked with related data, such as from external risk assessments.

Every approach to the measurement of cybersecurity capacity building efforts of a nation or region has strengths and weaknesses. Structured field coding is offered as a means for addressing some of the key problems with developing reliable and valid indicators that are comparable across nations and over time. This paper explains, illustrates, and critically examines the concept of Structured Field Coding (SFC) and discusses limitations on its use.

1.1. Approaches to National Case Studies and Comparative Research

One of the first steps in measuring cybersecurity capacity building it to establish the basic parameters that are needed for capacity building. In the area of assessing the maturity of cybersecurity capacity building. There have been many approaches for developing comparable assessments. Many widely used scales have been developed across multiple societal dimensions that impact cybersecurity capacity, either directly or indirectly. These often include such aspects as technical norms, educational programs, legal protocols, and policy mandates. The ITU's Global Cybersecurity Index, is one example, along with a NIST cyber security framework [12], but these are only two of many other approaches that have been developed that overlap in their methodology and empirical indicators.

This paper focuses on one approach where structured field coding has been developed. Oxford University's Global Cyber Security Capacity Centre (GCSCC) has developed, in collaboration with over two hundred international experts, a Cybersecurity Capacity Maturity Model for Nations (CMM) and an approach to assessing nations that has been deployed in over 80 nations to date.

The CMM reviews cybersecurity capacity across five dimensions: (1) Cybersecurity Policy and Strategy; (2) Cyber Culture and Society; (3) Cybersecurity Education, Training, and Skills; (4) Legal and Regulatory Frameworks; and (5) Standards, Organizations, and Technologies. Each dimension consists of a range of factors that describe what

it means to possess cybersecurity capacity in that dimension, and aspects of each factor that enhance maturity. A set of indicators for each aspect of those factors is used to gauge cybersecurity maturity along a five-stage spectrum, ranging from (1) start-up; (2) formative; (3) established; (4) strategic; to (5) dynamic [13].

During the initial years of deploying the CMM, data-gathering involved in-country stakeholder consultation (typically 2–3 research staff visit over the course of three days), complemented remotely through desk research. The in-country consultations, which relied mainly on modified focus groups, was to yield evidence for assessing capacity building for each nation in ways that can be used both to recommend capacity-building initiatives for nations but also for making comparisons across nations. But as in the case of national comparisons, the regional assessments would ideally be comparable across multiple units to estimate capacity levels across the nation.

Investment, and policy decisions are inevitably made based on national assessments, whether these are limited to mere hearsay or anchored in systematically empirical and accountable evidence (Box 1). Thus, a rigorous and accurate assessment is advantageous for data-based decisions. Nations cannot avoid being challenged for their policy decisions, but the more reliable and valid the evidence is judged to be, the easier it is to demonstrate the foundations for prioritizing areas for investment.

VALIDITY: concerns the degree that an indicator is measuring what it is intended to measure. Are you measuring what you think you are measuring? Have there been multiple tests and expert input to support validity? Does cybersecurity capacity maturity indicate the resilience and status of a nation in responding to breaches and other attacks on cybersecurity?

RELIABILITY: refers to the degree that an indicator can replicate an underlying trait accurately or consistently. Will an approach to capturing a national level of cybersecurity maturity be capable of yielding the same results if replicated, such as by a different team of researchers?

Box 1. Reliability and Validity

1.1.1. Aggregate Data Collection

One approach that is common in relatively well-defined areas is to combine new or existing aggregate national indicators to assess performance, such as in the areas of economic development or freedom of the press. Cross-national comparative research is often based on field research or the use of available aggregate data that might have been collected for other purposes, but which can be used to extract empirical indicators of national similarities and differences.

For example, Freedom House rates countries or territories on 10 indicators of political rights, such as free and fair elections, and 15 indicators of civil liberties, including the rule of law, that are each ranked from 0–4, where 0 represents the lowest level of freedom and 4 the highest. In 2020, Norway and Finland were ranked 1 and 2 on their respective scores on the press freedom index [14]. Other aggregate data approaches have been used to develop indicators of governance [15], and cyber power [16], for example. Many of these kinds of studies or rankings are done by a single organization following developments and activities across multiple nations across the world. Aggregate data can be drawn from research conducted by other organizations for other purposes. The benefit of aggregate data is that it is often collected by well-funded and highly respected organizations and can be used to empirically test concepts that otherwise couldn't be tested at a large scale [2].

The use of the same data for multiple studies is almost demanded by the time and cost of developing national indicators. The challenge with aggregate data is that the data items collected limit the research questions that can be addressed. Thus, aggregate analysis is limited to the relationships between verified data in the sets under consideration, requiring some collection of original data.

1.1.2. Field Research and Data Collection

Original collection of data for national assessments allows for collection of more than just standard measures, it allows for customization and contextualization of the data collection tools to fit the specific country and phenomenon being studied. As in many domains, understanding the unique needs data sources within a country often takes time and expertise. For example, most national comparative research in the study of government and politics has been conducted by individuals who lived and/or worked for a sustained period in a particular nation other than their own. They had become expert participant-observers of activities in the nation that is their object of

study and most often develop their findings. A classic demonstration of this method is Alexis de Tocqueville's examination of democracy in America in the early 1800s (N = 1), or a comparative case study (N = 2 or more) [17]. A more recent example is Gabriel Almond and Sidney Verba's study of political attitudes in the United States, Great Britain, Germany, Italy, and Mexico, which drew on qualitative observations as well as comparative survey research data collection [18]. The World Internet Project (WIP)[1] collects data from a growing number of countries, focused on issues of the digital divide.

1 — https://www.worldinternetproject.com/

Given the intense commitment of time, labor, and expertise, it is rare for field research about cybersecurity to be conducted in many nations. This work has often been limited by the ability of individuals or a small team to be directly involved in observing, interviewing, and comparing nations. Even though this method can produce high quality reviews with many actionable insights, they can be prohibitively expensive and require commitment from many stakeholders.

Additionally, the need for strategies that might enable larger and more distributed teams to gather comparable data in the field, new methods are being developed that include the use of interviews and participant-observation. These have the potential to give deep insights to help guide policy changes that are needed in this dynamic domain. However, to develop these methods, that would allow both standardization for comparative value and flexibility to adjust to changing threats; it is important to first examine what has already been used in other leading cybersecurity capacity assessments.

1.1.3. National Risk Assessments and the UK's National Cyber Risk Assessment (NCRA)

National risk assessments are focused on identifying major risks facing a nation, developing estimates of how likely it is that the nation will experience each risk, and estimating the severity of the risk. Given the uncountable number of known and unknown risks that nations might face, even in a constrained area, such as cyber, these assessments rely heavily on the judgements of subject matter experts in a wide range of sectors within each nation. The Organization for Economic Co-operation and development (OECD) and other major international organizations view risk assessment as critical to managing these risks and achieving national economic and social goals [19].

The government of the United Kingdom has developed a Cyber assessment Framework from the National Cyber Security Centre

that has four objectives: managing security risk, protecting against cyberattack, detecting cyber security events, and minimizing the impact of cyber security incidents [20]. Each of these items are broken down into measurable items that focus on IT systems and network solidarity [20].

A cyber risk assessment captures the judgmental ratings or forecasts of relevant experts on the priority, likelihood, and level of cyber risks across many sectors and critical infrastructures of specific nations (Box 2). While the questions can be structured in similar ways cross-nationally, the answers are sufficiently unique to each nation that comparison is difficult. They are developed to help nations better anticipate and manage potential risks to their nation and not designed with an aim of cross-national comparison. As with cybersecurity capacity assessments, it is difficult to study such assessments across multiple nations in comparable ways.

National transportation and telecommunication systems, including the information and communication technologies (ICTs) that support them, are two huge critical infrastructure sectors. Depending on the level of analysis, the number of CIs in focus vary from four to nearly twenty viewed as necessary to the functioning of the nation. These are the systems, networks and assets that are essential to the functioning of a society [21].

Box 2. Critical Infrastructure Sectors.

Discussions between those involved in research on cybersecurity maturity and risk assessments identified ways to improve approaches to each area of research, also explored the potential efficiencies and synergies of integrating these two heretofore separate activities. The idea of an integrated cybersecurity maturity and risk assessment (Cyber-MRA) is attractive, creating one unified assessment to give insights and guide policy decisions. However, given the different organizations, traditions, and methods tied to each, their integration is challenging. Assessments that rely on qualitative analysis, focus groups, and interviews often produce deep insights, but these are not always quantifiable and comparable across sessions of data gathering. On the other hand, field surveys with strict "fill in the bubble" type approaches yield problematic data even though they allow quantitative analysis and comparison across data sets. A potentially strategic innovation for improving on and integrating these two approaches involves the use of what we have called "structured field

coding". The next section describes this approach and then moves to a discussion of how it can enhance and potentially integrate both maturity and risk assessments.

1.2. Structured Field Coding (SFC)

Structured field coding (SFC) can potentially advance the study of national cybersecurity maturity and national risk assessments as well as provide a means for better integrating multiple methods and indicators involved in both qualitative and quantitative approaches. The following is the history of the method, its growth and how it has been utilized in national cybersecurity assessments to provide rich insights and actionable items for building cybersecurity capacity.

1.2.1. The Origins of SFC

SFC was invented to solve a set of problems that arose in the study of the early use of computing in forty US cities in the late 1970s [22]. The study, entitled An Evaluation of Urban Information Systems (URBIS), was one of the first systematic studies of the use and implications of computer systems in American local governments. It was funded by the US National Science Foundation and based on what was then the Public Policy Research Organization (PPRO) at the University of California, Irvine. The principal investigator was Professor Kenneth Kraemer, who led the team of co-principal investigators, including James Danziger, William Dutton, Rob Kling, and Alex Mood. The study began with a survey of all 403 US cities with populations over 50,000. In 1975, circa the time of this study, this was about the size at which a city might have had one or more computers and associated applications, although nearly a quarter (23% or 93) of the cities over 50,000 inhabitants had such a negligible level of computing that they were dropped from the study.

The URBIS team devised a means for stratifying all US cities on key policy variables, which involved using standard demographic data to estimate scores on indicators of computer use for those cities for which data was not available. The team then randomly sampled cities such as to maximize variation on the major variables of interest to the study, such as the centralization or decentralization of computer facilities. This approach was called the "future cities research design" [23].

This analysis led to the selection of a stratified random sample of 40 US cities that were then studied more intensively. The research

formed the basis of numerous publications across all the investigators including two major academic books [22, 24], which together provide considerable evidence of the academic merit and acceptance of the study's methodology.

1.2.2. The Invention of SFC

It was in the pilot stages of this in-depth comparative study of the 40 sampled cities that methodological challenges began to arise. First, there were multiple teams going into the field, most composed of two researchers spending about two weeks in each city conducting interviews, visiting departments, and observing work around several foci of the study, such as in the use of computer-based data in policy analysis, detective investigative support, and other applications representative of different "information processing tasks". In pretesting our approach through a small set of "mini-cases", it was apparent that different investigators tended to confirm their preconceived notions, demonstrating the issue of researcher bias. This is a long known issue, even de Tocqueville is said to have had preconceived views on America, as one French critic said: "He had thought it all out before he learned anything about it [America]" [25]. This led to developing ways to make the research more objective versus relying too heavily on any one person's preconception, or other subjective or judgmental rating, and also to ensure multiple points of view and greater accountability – creating an ability to double-check the conclusions of those who did the field work.

Reducing researcher bias was accomplished using SFC. Essentially, each of the two or more researchers in the field would answer the same questions the entire team had considered critical to the study. They could use in-depth interviews, observations in the field, desk research, and informal discussions with staff and politicians, for example, to arrive at their answers. The team would then compare and contrast their answers and resolve differences of opinion across the members and explain in notes why the city was coded in the way it was finally determined. This method reduced the impact of bias from any particular researcher, while providing a means for integrating multiple observations into a single code. It also made every code more credible since it was the product of multiple observations and explanations for why a city was coded as it was, provided by those individuals who were observers in the field. Thus, it was a structured way to establish "inter-coder reliability" across data that was collected through multiple methods.

2. Methods: SFC in action

Structured field coding (SFC) refers to the development of predetermined questions and potential responses that are answered by the researchers while they are in the field. This is similar to a survey, but different in that the researcher completes the items based on evidence from the field while the data is being collected. Responses can be refined but are initially coded while fresh from interviews and observations in the field context. Using desk research, discussion group transcripts, and interviews with those informed about particular topics, the researchers in the field aim to be in a position to answer each question. The answers to each question are then used to operationally define each indicator. Supporting data is then available to others in the research team to test items for reliability and objectivity.

2.1. The URBIS Example

In the URBIS study, for example, a key question concerned whether the use of computing in cities would shift influence or power to one or another kind of actor [23]. One item asked about the use of data in the city: "In general, has the use or design of data banks [a term of the 1970s], their analysis or the distribution of findings tended to shift relatively more influence away from or to any of the following: [Manager or chief administrative officer; Mayor and staff; Council and staff; Departments; and Data bank custodians], with the following response categories: "Given less influence to; No discernable shift; Given more influence to" [23, p. 200]. Researchers were asked to provide notes to defend their response. A related question was: "Are various computer-based reports and special analyses generated from operational data used in responding to individual or citizen group requests or complaints?". Coding was provided for the manager, mayor, and council, with each coded separately with the following response categories: Cases cited that it could have been but was not used; Believe it is not used; Undecided, mixed; Believe it is used, Cases cited that it is used [23, p. 201]. This exact question was addressed in an interview one researcher had with the city manager of a large city with a city manager form of government. The manager said such analyses were never used, but the researcher went with him to the council meeting that followed, when the manager's office presented the results of systematic modeling where the data in fact were being used. In such ways, multiple sources are reconciled to arrive at a researcher's coding from observations and data collected in the field based on a pre-designed set of codes.

Another innovation tied to the use of SFC in URBIS was the identification of a set of information processing tasks (IPTs) that not only

indicated the levels of data use, but also flagged areas that could be studied in more depth. In the late – 1970s, the application of computing in governments ranged across at least six IPTs (Box 3). For example, by looking at a specific task, such as record searching in supporting detective investigations, or record-keeping in traffic ticket processing, it was possible to have a more concrete empirical basis for assessing the impact of computing. Looking across these IPTs within one city provided an overall picture that was more grounded in the government's use of computing [22]. It is possible that risk assessments might identify a set of comparable CIs that could be a focus for more concrete and detailed analysis in an analogous way.

1. Record-keeping, such as in traffic ticket processing
2. Calculating/printing, such as in budget control
3. Record-searching, such as in detective investigative support
4. Record restructuring, such as in policy analysis
5. Sophisticated analytics, such as in police patrol manpower allocation
6. Process control, such as in budget monitoring and control

Box 3. Information Processing Tasks Defined by URBIS.

2.2. The CMM Example

In the context of the CMM assessment, for example, there is a question with responses for whether a nation has a cybersecurity strategy (Box 3). Focus groups often have responses much like the URBIS experience of several decades ago. Many are unaware of efforts going on in other departments. Additionally, even experts seeking to find evidence of operationalizations in action might not find all the same evidence. If two or more researchers go to a country, they would code responses to this question, and together, they would reconcile any differences in order that they agree on the best code for the country, based on what they have learned. They assign a quantitative number to their qualitative judgements. In instances when they have fundamental differences, which cannot be easily resolved, they would go back into the field via an email, conference call, or use desk research to resolve the differences. For example, they might have been told a strategy was in development, but discover a published report online, and switch their coding to "4" for "yes, and it has been published".

The advantage of the SFC method is its ability to capture data that is not obvious, as well as items across domains, which might be missed

from relying solely on a survey of participants. Cybersecurity capacity building is a multi-faceted, multi-disciplinary effort that requires advanced application of expert knowledge. Measuring progress to evaluate efficacies is a complex, yet critical effort. Just as the example of the URBIS application, the SFC allows stakeholders to see where they are not utilizing resources available to them to improve policies and processes.

Q1. Does the country have a national cybersecurity strategy? (Circle Response)

5... Yes, the country's strategy has been cited as "world leading"
4... Yes, and it has been published
3... Yes, but not published
2... No, but it is in development with a draft or outline
1... No, but the processes for strategy development have been initiated
0... No, strategy does not exist

Evidence, Examples:

Box 4. A Question from the CMM

2.3. A Critical Perspective on the Approach

SFC is a relatively simple idea that provides a flexible approach that increases research validity and reliability when dealing with a complex array of methodological approaches. This is especially important when researching cybersecurity capacity building efforts as it touches on technological, educational, legal, communications, and societal domains.

2.3.1. Strengths of SFC

1. *Reliability of Multiple Observations and Codes:* SFC embeds the use of multiple observers, and coders for each question. This is designed to enhance the reliability of the code agreed across researchers. As differences that emerge across coders will lead to notes and explanations of how the code was resolved, the notes also enhance the reliability attributed to the resulting data.

2. *Areas of Uncertainty or Lack of Awareness.* Capturing multiple (independent) observations at the indicator level also enables

the researchers to identify the specific indicators on which there was uncertainty (on which two or more observers significantly disagreed, for example), and which therefore need continued attention, such as in follow-up calls to the field work, in order to resolve. In many online or remote collection of data these discrepancies would not be known. Discrepancies across coders might also be interesting evidence of conflicting viewpoints in the country or differences in the knowledge base of the participants interviewed. In the example of the CMM, this method helps paint an even richer picture of cybersecurity maturity in the country and identify problematic or uncertain areas of capacity building.

3. *Capturing Detail at the Indicator Level.* For the CMM, as one example, each question represents an indicator of one or more aspects of maturity. Having an agreed coding for each indicator provides more variation and evidence at the indicator level for each nation. Individual indicators can then be used alone or in some descriptive and comparative analyses, as well in calculating maturity levels of their respective aspects.

4. *Evolvability of Operational Definitions.* Coding at the indicator level makes it possible to refine and revise any operational definition of any variable including each respective question. In the case of the CMM, the model can be evolved simply by revising the operational definition of aspects, such as moving an indicator to be grouped or combined with a different aspect. In addition, the team can operationally define why a country is given a particular maturity code on any given aspect. Since each indicator related to a maturity code for any given aspect is recorded, any change in the definition of an aspect can be accommodated by changing the operational definition – how different indicators are combined. In defining an aspect in a new way, using the existing indicators, the new aspect can be quickly recalculated to obtain a new maturity score defined by the model. Researchers can draw from the existing indicators. That is, the model can evolve, and the existing indicators can be used to recalibrate maturity levels.

5. *Precision of Comparisons Over Time and Cross-Nationally.* The operational definition of indicators and maturity or risk levels will enable more reliable and operationally defined variance across countries, and more reliable and valid measurements of maturity or risk over time and cross-nationally. For instance, by relying primarily or only on modified focus groups, the researchers make judgmental ratings of the maturity levels of each aspect given the

observation of particular „indications" of their level. But this does not capture variation on the actual indicator – only that it might have been observed. More precise, operational indicators would enable less dependence on judgmental ratings and better able to capture minor differences cross-nationally and longitudinally. For instance, in the example question above, you might find that a country had published its strategic plan, another which has not. So small variations would be more visible and subject to analysis.

6. *Transparency and Accountability.* SFC leads to national ratings that are more transparent and accountable as anyone could see and question the operational definitions of ratings, and the indicators used. So SFC would enhance transparency of the data and analyses based on cross – national or over-time comparisons.

7. *Integration of Data from Multiple Sources.* One of the most valuable advantages is that by enabling the use of multiple data sources in coding, it is possible to draw not only from multiple data on the same nation but also multiple studies, if conducted concurrently by the same or even different research teams, if they used the same SFC. A later section of this paper will illustrate its potential for integrating the study of national cybersecurity capacity building with the study of national cyber risks.

8. *Integration across Sectors or Infrastructures.* Finally, many studies of governments or nations cannot study all activities, sectors, or infrastructures. A pragmatic but also a valid approach is to identify a sample of individuals, departments, sectors, or infrastructures to study in more depth, but in ways that can be compared and/or aggregated to a higher level of analysis. SFC could be developed to ask similar questions about different objects of analysis in ways that the answers are more comparable and less problematic to aggregate.

2.3.2. Limitations of SFC

There are weaknesses or limitations of SFC – it is not a silver bullet for resolving major challenges in national and cross-national comparative research. These include:

1. *Limits on Independence of Coders.* Each indicator defined by a maturity model or risk assessment should be coded by two or more independent observers. With at least two individual researchers going into the field there is nevertheless the likelihood of some bias of individuals to confirm their

preconceptions – a confirmatory bias – but also the potential for interpersonal influence to undermine the independence of the coding. Of course, at the end of the day, researchers need to compare codes and resolve differences of opinion, which demands some role of interpersonal influence and compromise. However, these sessions can be conducted in light of concerns over avoiding any confirmatory and group think biases, and most other research approaches face the same threats, such as how different individuals code group discussions

The potential for any lack of independence is addressed in several ways. First, desk research is likely to involve researchers beyond the field team. Secondly, the explanations of codes could indicate a lack of independence, which would be apparent to those beyond the field team drafting the report. Finally, the codes and the report based on them will be reviewed by experts outside the field team, including experts within the respective nations. Judgmental ratings and SFC will be sufficiently transparent that they will be subject to several stages of accountability.

2. *Time Demands of Coding.* In some respects, the use of two independent coders might be viewed as doubling the workload on the research team, but this is a compromise that will lead to more reliable and valid indicators (Box 1). As discussed in the section on the origins of SFC, the risks of a single coder appear greater than threats that two coders, but one more coder will not eliminate such a risk. Nevertheless, clarifications of codes by two coders will add more texture to the meaning of the code and the evidence behind it.

3. *Pressure to Reduce the Number of Indicators.* Time demands do exert pressure on the study team to minimize the number of questions or indicators included in the study. It is a natural expectation that subject matter experts in cybersecurity or cyber risk will want to be as comprehensive as possible and include every conceivably important question. However, there reaches a point when the time required in the field surpasses that allotted, which threatens the care and precision of the coding process. The research team needs to include enough indicators to get a reliable estimate of aspects related to those indicators but avoid temptations to be comprehensive. It is easy to write questions but difficult to answer and code them. This creates an inherent problem with the team creating too many questions in ways that inadvertently reduce the quality of the research. A survey has limits imposed by the time that respondents are willing to spend answering questions. This

places severe constraints on the number of questions asked in surveys, for example. Likewise, a specified time in the field places similar constraints on the number of interviews, discussion groups, and participant – observation that is possible in a single nation. The research team must therefore exercise considerable discipline in reducing redundancy, tangential questions, and exceedingly complex coding issues to ensure that the field research is completed. Just as a set of survey questions does not need to be comprehensive to provide an indication of a behavioural or attitudinal propensity, neither do the indicators included for SFC need to be comprehensive. What indicators are necessary to make a judgement on the relative maturity of a nation in a particular area of cybersecurity?

4. *Risk of Failing to Gain Multiple Codes.* It is possible a researcher might fail to get evidence about all indicators, so two observers will enhance the likelihood of at least one researcher collecting evidence from interviews or observations that can be used for coding the indicator. This is a pragmatic reality of field research. It is not ideal, but the effort would strive to obtain evidence from each researcher on each question, recognizing that this will not always be possible in the time allotted to field research, and the strategies for gathering data in the field, such as in dividing in-depth expert interviews up between the two or more researchers. Desk research and post-field research interviews, such a via video conferencing, can be used to address any doubts raised by the lack of double coding.

5. *Limits to the Detail and Precision of Rapid Field Research.* Surveys are blunt instruments, seldom capable of capturing the precise level of detail many journalists, public officials, and other subject matter experts expect from them. Likewise, any research based on multi-methods conducted over a very short period of time – such as a few days – cannot be expected to be as precise as one would wish. For example, any data collected today, might be different tomorrow. Any evidence uncovered by two researchers over 2–3 days might miss additional evidence that could have affected the coding of indicators. Any period of time chosen for the research might be influenced by events in the national context or even the personal situation of the person interviewed that would bias the observations, such as a change in administration. These limitations need to be recognized and efforts should be made to identify any problematic data, but the team also needs to realize that they cannot be totally overcome.

6. *Unknown Knowns.* By having a pre-defined "structure", or sets of questions and codes, the study could fail to capture information that is pertinent to cybersecurity maturity assessment that was not already identified and added to the SFC's codes. In the case of the CMM study, this is addressed in part by continuing revisions of the underlying model, based on lessons learned from past assessments and new technical and legal approaches. But SFC does lean heavily on the research team having the right pre – conceived notions of what are the best indicators of cybersecurity maturity, and not missing any key developments. However, discussions gained in in-depth interviews and modified-focus groups are recorded and can inform each case but would be less likely to be valuable for comparative study. Employing a more "grounded theory" approach, where the participant-observation and related data are collected in a more open-ended manner [26], would focus more attention on researchers reviewing the interviews, notes and discussions in an iterative manner to identify the codes to be applied to it. That said, this approach is still framed by the less structured ideas that frame the questions and observations of the researchers and tend to develop more unique frameworks for each case study that could be compared cross-nationally but in different, broader, and more thematic ways. Moreover, the use of SFC adds numbers to qualitative data. It does not erase or substitute for qualitative and other quantitative data and observations. What it does do is insist that the researchers cover areas defined by the SFC and in this respect it steers data and observations in ways that might not be incorporated in a more open – ended approach to following the evidence.

2.4. SFC Enhanced by Modified Focus Groups

The use of structured field coding (SFC) could complement and augment the use of modified-focus groups for field research on cybersecurity capacity building, or the use of discussion groups focused on cyber risks. Past CMM reviews relied greatly on what the GCSCC team has called modified – focus groups, which have several limitations that can be reduced using SFC.

The term "modified-focus groups" (MFG) is meant to convey the divergence of this approach from traditional focus group methods per se, a process invented by a famous sociologist, Robert Merton, in the 1950s to study opinion formation, such as study of why people support a policy. Standard approaches to focus groups are generally used to surface a wide range of opinions, through open-ended

questions, such as seeking to understand what people understand by the concept of cybersecurity or cyber risks. They are excellent approaches for understanding how to design a questionnaire, for example, a focus group discussion of how the government thinks about cybersecurity capacity might help us design more structured questions to which individuals could respond.

However, standard focus groups are not designed to reach a consensus on a question or an issue, but to foster a range of opinions. The exact opposite aim is the normal rationale for a MFG. MFGs are designed to elicit a range of opinions that lead to some consensus, such as whether a nation follows a particular practice. Moreover, MFGs bringing together individuals from government, business and industry, civil society, and academia violate some assumptions that underpin the value of collective intelligence. MFGs can bring in rich insights beyond the more objective SFC and even though it is a challenge to combine the data, the process has led to success in the CMM national evaluations based on such criteria as construct validity – judged by empirical relationships with other indicators expected to be associated with the indicator being measured.

One of the challenges of MFG it that they are also very difficult to validly replicate. If a specific field researcher moderating the discussion chose to kick off discussions with their own inspired prompts and questions, based on the specific context and informal discussions, it could skew the entire group. Each focus group could be primed somewhat differently and would therefore possibly react to somewhat different sets of questions and prompts. They can be replicated only in the broadest sense of doing multiple focus groups, with each likely to be composed of different sets of individuals and with the likelihood of being primed by the early statements and questions raised by the participants and moderator. Thus, despite such challenges, MFGs work well enough, as reflected in the face validity and construct validity of measurements, while also playing an important role in awareness raising and networking. In addition, the MFGs provide important indications of the state of knowledge in the country and where knowledge gaps might be, which can be important to influencing practices. It can be important to gain a sense of whether people from a range of sectors are generally aware of various strategies, legislation, activities (e.g., awareness-raising activities, to understand whether they have a good reach) – and not just to know whether these things exist by asking the experts. For example, if awareness-raising activities exist, but very few people know about them, this could explain any lack of success.

Results

Structured field coding is supported by multi-method, multi-sourced data collection, thus seeks to increase validity and reliability in critical research. Figure 1 illustrates how multiple data sources can feed into judgements made on the coding of a nation on any number of criteria. SFC would be used to code each indicator of both studies and in doing so, it would convert data from any source into a comparable national indicator. That said, some data would not need SFC, such as the population of the nation and other demographic indicators that would directly fall into the data set if they are at the national level of analysis. Since major approaches to cybersecurity capacity assessments and cyber risk assessments use many of the same data sources, it is feasible to integrate the conduct of both assessments to create an integrated national data file (IND).

An example of how operationalization measures of indicators would work is given in the following examples. Firstly, an analysis of an indicator of the quality of cybersecurity education in a nation. This might be feasible to gauge through desk research using existing reports, news, and the web and related social media. On the other hand, an indicator of Internet use in each respective nation could be measured through existing surveys [6], or bespoke surveys created by the study team. Rating the indicator of the risk of cyber-attacks, along with their likelihood and severity, could be gained through NCRA surveys, governmental, business and industry reports, and all followed up with expert interviews. The multiple sources, collected through multiple methods pool together and strengthen insights.

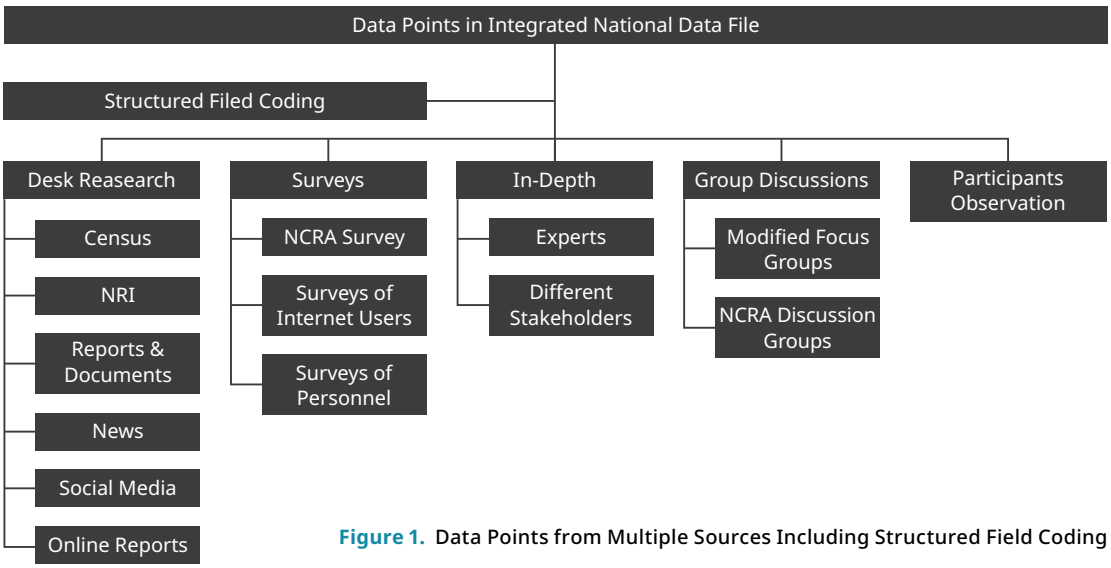

Figure 1. Data Points from Multiple Sources Including Structured Field Coding

3. Discussion

The development of quantitative data from qualitative research is a challenge in the best of circumstances. However, using SFC as part of a holistic data collection process, is a challenging process as it takes funding of trained researchers and a long term commitment to support the required levels of data collection and analysis.

Structured field coding, even though it has been foundational in measuring cybersecurity capacity maturity across many nations through various programs, has to potential to do even more as it is scalable and comparable. A complete assessment relies on multiple methods of data collection. Additionally, this technique could also be used to integrate the data collection process in ways that reduce duplication (each assessment has some common indicators, such as demographics) and create an integrated national data (IND) file that would facilitate analysis of the relationships between aspects of cybersecurity and capacity building. Policy measures that encourage the use of robust measures such as SFC allow nations to measure progress in their capacity building efforts. It is possible to maximize reliable variance across nations in ways that would better support cross-national and longitudinal analysis. These types of analysis are essential to better understand the impacts of less direct measures of capacity building (e.g., legal changes or educational efforts) impact long term outcomes.

4. Conclusions

Structured field coding (SFC) provides a robust technique for reducing redundancy while enhancing the efficiency and effectiveness of cross-national comparative studies. It provides a structured way to enhance inter-coder reliability across data collected through multiple methods. At the same time, does not lose any of the virtues of multiple methods, such as focus groups or in-depth interviews. And it allows the research to amalgamate data in a documented and transparent way across multiple methods to move into a simple structured frame. A promising potential application is the integration of cybersecurity maturity assessments done in conjunction with cyber risk assessments. The resulting integrated national data file would be more powerful than either one data file on its own in supporting a nation's self-assessment and help bring together a wide range of analytical approaches to key questions.

Funding

This work was supported by the Oxford Martin School | Global Cyber Security Capacity Building Center https://gcscc.ox.ac.uk/home-page and their sponsors.

References

[1] S. Creese, W. H. Dutton, P. Esteve-González, R. Shillair, "Cybersecurity capacity-building: cross-national benefits and international divides, *Journal of Cyber Policy*, vol. 6, no. 2, pp. 214–235, 2021, doi: 10.1080/23738871.2021.1979617.

[2] W. H. Dutton, S. Creese, R. Shillair, M. Bada, "Cybersecurity Capacity: Does It Matter?, " *Journal of Information Policy*, vol. 9, pp. 280–306, 2019, doi: 10.5325/jinfopoli.9.2019.0280.

[3] R. Shillair, P. Esteve-González, W. H. Dutton, S. Creese, E. Nagyfejeo, B. von Solms, "Cybersecurity education, awareness raising, and training initiatives: National level evidence-based results, challenges, and promise," *Computers & Security*, vol. 119, p. 102756, 2022, doi: 10.1016/j.cose.2022.102756.

[4] R. Collett, "Understanding cybersecurity capacity building and its relationship to norms and confidence building measures," *Journal of Cyber Policy*, vol. 6, no. 3, pp. 298–317, 2021, doi: 10.1080/23738871.2021.1948582.

[5] M. Górka, "The Cybersecurity Strategy of the Visegrad Group Countries, "*Politics in Central Europe*, vol. 14, no. 2, pp. 75–98, 2018, doi: 10.2478/pce-2018-0010.

[6] ITU, "Global Cybersecurity Index." 2018. [Online]. Available: https://www.itu.int/dms_pub/itu-d/opb/str/D-STR-GCI.01-2018-PDF-E.pdf. [Accessed: Aug. 8, 2023].

[7] EGA 20, "The National Cyber Security Index Ranks 150+ Countries' Cyber Security Status." [Online]. Available: https://ega.ee/news/national-cyber-security-index-ranks-150-countries/ [Accessed: Nov. 11, 2023].

[8] NCSI, "National Cyber Security Index Methodology 3.0," 2023. [Online]. Available: https://ega.ee/wp-content/uploads/2023/08/NCSI-3.0_Methodology.pdf. [Accessed: Aug. 9, 2023].

[9] L. Muller, "Cyber Security Capacity Building in Developing Countries: Challenges and Opportunities," Norwegian Institute of International Affairs, Oslo, Norway, 3, 2015. [Online]. Available: https://cybilportal.org/wp-content/uploads/2020/06/NUPIReport03-15-Muller.pdf. [Accessed: Nov. 13, 2023].

[10] S. Creese, W. H. Dutton, P. Esteve-González, "The social and cultural shaping of cybersecurity capacity building: a comparative study of nations and regions," *Personal and Ubiquitous Computing*, vol. 25, pp. 941–955, 2021, doi: 10.1007/s00779-021-01569-6.

[11] Z. Homburger, "The Necessity and Pitfall of Cybersecurity Capacity Building for Norm Development in Cyberspace," *Global Society*, vol. 33, no. 2, pp. 224–242, 2019, doi: 10.1080/13600826.2019.1569502.

[12] S. Almuhammadi, M. Alsaleh, "Information Security Maturity Model for Nist Cyber Security Framework," *Computer Science & Information Technology* (CS & IT), Academy & Industry Research Collaboration Center (AIRCC), pp. 51-62, 2017, doi: 10.5121/csit.2017.70305.

[13] Global Cyber Security Capacity Centre, "Cybersecurity Capacity Maturity Model for Nations (CMM)," 2021. [Online]. Available: https://gcscc.ox.ac.uk/the-cmm. [Accessed: April 11, 2023].

[14] Freedom House, "Freedom in the World 2023," 2023. [Online]. Available: https://freedomhouse.org/report/freedom-world/2023/marking-50-years. [Accessed: Nov. 13, 2023].

[15] D. Kaufmann, A. Kraay, M. Mastruzzi, "The Worldwide Governance Indicators: Methodology and Analytical Issues," *Hague Journal of the Rule of Law*, vol. 3, no. 2, pp. 220–246, 2011, doi: 10.1017/S1876404511200046.

[16] J. Voo, I. Hemani, D. Cassidy, *National Cyber Power Index 2022*. Cambridge, MA: Belfer Center for Science and International Affairs, Harvard Kennedy School, 2022.

[17] A. de Tocqueville, *Democracy in America.* Washington, D.C.: Regnery Publishing, 2003.

[18] G. A. Almond, S. Verba, *The Civic Culture: Political Attitudes and Democracy in Five Nations*. Princeton, NJ: Princeton University Press, 2015.

[19] OECD, "Recommendations of the Council on Digital Security Risk Management," Paris, 2022. [Online]. Available: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0479. [Accessed: Apr. 12, 2023].

[20] National Cyber Security Centre, "NCSC CAF Guidance Principles and Related Guidelines," 2019. [Online]. Available: https://www.ncsc.gov.uk/collection/caf/table-view-principles-and-related-guidance. [Accessed: Apr. 12, 2023].

[21] NIST, "Framework for Improving Critical Infrastructure Cybersecurity," National Institute of Standards and Technology, pp. 1–41, 2014, doi: 10.6028/NIST.CSWP.04162018.

[22] K. L. Kraemer, W. H. Dutton, A. Northrop, *The Management of Information Systems*. New York: Columbia University Press, 1981, doi: 10.7312/krae93774.

[23] K. L. Kraemer, J. N. Danziger, W. H. Dutton, A. M. Mood, R. Kling, "A future cities survey research design for policy analysis," *Socio-Economic Planning Science*, vol. 10, no. 5, pp. 199–211, 1976, doi: 10.1016/0038-0121(76)90029-X.

[24] J. N. Danziger, W.H. Dutton, R. Kling, K. L. Kraemer, *Computers and politics. High technology in American local governments*. New York: Columbia University Press, 1982.

[25] A. Ryan, *On Tocqueville: Democracy and America*. New York: W. W. Norton & Company, 2014.

[26] J. Corbin, A. Strauss, "Grounded Theory Research : Procedures , Canons and Evaluative Criteria," *Zeitschrift Für Soziologie.*, vol. 19, no. 6, pp. 418–427, 1990, doi: 10.1515/zfsoz-1990-0602.