International Journal of the Commons
Vol. 10, no 2 2016, pp. 440–466
Publisher: Uopen Journals
URL:http://www.thecommonsjournal.org
DOI: 10.18352/ijc.652
Copyright: content is licensed under a Creative Commons Attribution 3.0 License
ISSN: 1875-0281

Challenges and opportunities in coding the commons: problems, procedures, and potential solutions in large-N comparative case studies

Elicia Ratajczyk
Center for Behavior Institutions and Environment and School of Human Evolution and Social Change, Arizona State University, USA
elicia.ratajczyk@asu.edu

Ute Brady
Center for Behavior Institutions and Environment and School of Human Evolution and Social Change, Arizona State University, USA

Jacopo A. Baggio
The Department of Environment and Society, Utah State University, USA, and the Center for Behavior, Institutions and Environment, Arizona State University, USA

Allain J. Barnett
Department of Anthropology, University of New Brunswick, Canada

Irene Perez-Ibara
Center for Behavior Institutions and Environment, Arizona State University, USA

Nathan Rollins
Center for Behavior Institutions and Environment and School of Human Evolution and Social Change, Arizona State University, USA

Cathy Rubiños
Center for Behavior Institutions and Environment and School of Sustainability, Arizona State University, USA

Hoon C. Shin
Center for Behavior Institutions and Environment and School of Human Evolution and Social Change, Arizona State University, USA

David J. Yu
Center for Behavior Institutions and Environment and School of Sustainability, Arizona State University, USA

Rimjhim Aggarwal
Center for Behavior Institutions and Environment and School of Sustainability, Arizona State University, USA

John M. Anderies
Center for Behavior Institutions and Environment, School of Human Evolution and Social Change and School of Sustainability, Arizona State University, USA

Marco A. Janssen
Center for Behavior Institutions and Environment and School of Sustainability, Arizona State University, USA

Abstract: On-going efforts to understand the dynamics of coupled social-ecological systems and common pool resources have led to the generation of numerous datasets based on a large number of case studies. This data has facilitated the identification of important factors and fundamental principles thereby increasing our understanding of such complex systems. However, the data at our disposal are often not easily comparable, have limited scope and scale, and are based on disparate underlying frameworks which inhibit synthesis, meta-analysis, and the validation of findings. Research efforts are further hampered when case inclusion criteria, variable definitions, coding schema, and inter-coder reliability testing are not made explicit in the presentation of research and shared among the research community. This paper first outlines challenges experienced by researchers engaged in a large-scale coding project; highlights valuable lessons learned; and finally discusses opportunities for future comparative case study analyses of social-ecological systems and common pool resources.

Keywords: Codebook development, coding, common pool resources, complexity, coupled infrastructure systems, content analysis, intercoder agreement, intercoder reliability testing, social-ecological systems

Acknowledgement: The team would like to acknowledge project support from the National Science Foundation, grant number GEO – 1115054.

1. Introduction

Long-term efforts to understand social-ecological systems (SES) involving the management of common pool resources (CPR[1]) has led to the generation of a large body of data composed primarily of case studies (Wade 1984; Berkes 1989; Ostrom 1990; McKean 1992; Baland and Platteau 1999; Bardhan and Mookherjee 2006; Cox 2014; Epstein et al. 2014). If we are to understand CPR governance, we must be able to make comparisons across case studies but are challenged to develop reliable methods of making this often complex and messy data comparable. Meta-analysis, in the field of environmental social science is a mixed methods approach involving data extraction from case studies, through the coding of texts, for use in statistical or other comparative data analysis techniques (Hruschka et al. 2004; Rudel 2008; Cox 2015). As the essential activity of meta-analysis, coding involves the classification and quantification of texts or other media segments, preserved in a form which can be subjected to formal analysis (Hruschka et al. 2004). In this paper, we will contribute to understanding the challenges of coding case studies in environmental social science by critically exploring the experience of a team of researchers at the Center for Behavior, Institutions and the Environment (CBIE) at Arizona State University (ASU) while coding the 69 cases that form the data for Baggio et al. (2016) and Barnett et al. (2016).

In the next section (Section 2), we will briefly discuss the overall opportunities and challenges inherent in the coding of case studies for large-N meta-analyses and why this is a particularly important methodology in the field of environmental social science. We will discuss three primary challenges which we find can hamper meta-analysis efforts: 1) methodological transparency; 2) coding reliability; and 3) replicability of findings. In Section 3, we discuss our coding methodology in some detail and compare it to recommendations in the methods literature, including: preliminary decisions, codebook development, coding protocols, and intercoder reliability testing. We explore ways of increasing methodological rigor in these areas by adopting certain techniques and strategies from other disciplines in the social sciences and compare the approaches used by the CBIE team to approaches, or "best practices", recommended by a number of leading authorities within the methods literature. In Section 4, we utilize our findings from this comparison to develop a recommended coding protocol which we think could be widely applicable and easily adaptable to others using a comparative or meta-analysis methodology for research on SESs and the commons. We conclude the paper by sharing some ideas for future research in Section 5. We hope that by sharing these key methodological challenges and opportunities, we will stimulate a broader platform for communication and collaboration among scholars which will lead to better, more transparent research designs, opportunities in

[1] CPR theory is based on the assertion that there are many ways in which people are able to cooperate to solve social dilemmas involving shared, or common pool, resources and that there are some fundamental similarities which help people do this (Schlager 2004).

meta-analysis and data synthesis, and discoveries that will enhance our understanding of SESs.

2. The challenge

Meta-analysis, comparative analysis, and synthesis rely on the use of a rich resource of case studies which have been collected by numerous researchers over a long period of time. Secondary analysis of data of this kind, gathered for other purposes using diverse measures and variables, is inherently subjective and it is therefore important to take measures to increase coding reliability and replicability. This can present challenges in research design and implementation. Secondary analysis of existing case studies, however, has the advantage of being a relatively low-cost approach, compared to primary data collection, and can enable larger scale comparative analyses (Kelder 2005; Savage 2005). Meta-analysis offers the opportunity to refine findings within a wider community, discover what the dominant discourses are and generate new knowledge through the validation of previous findings. In addition, the use of synthesized datasets allows for the use of existing data in new ways and analyses across multiple time periods, scales and sectors, thereby potentially improving researchers' ability to understand complex system dynamics and adaptation (Ostrom 1990, 2012; Kelder 2005; Poteete et al. 2010; Cox 2014). Araral (2014) and Agrawal (2014) characterize this type of work in the study of the commons as the "emerging third generation" of research within the legacy of Elinor Ostrom, and see these efforts to generalize and extend her arguments across scales and with increased complexity as being of "fundamental importance" (Agrawal 2014, 87).

Relying on secondary data, however, is often difficult (Poteete et al. 2010) as existing data are often limited in their scope and scale, and are separated into independent databases using unique coding schema[2] and storage structures which are not always made publicly available. These limitations and divisions hamper synthesis efforts and comparability. For example, there are a number of data repositories (Table 1 in Supplementary Material) based upon the work of Elinor Ostrom and her collaborators on CPR theory.[1] These libraries of data represent a rich and mostly unexploited resource for increasing our understanding of CPRs via meta-analysis and comparison with contemporary data (Corti et al. 2005). These databases, however, each possess their own idiosyncrasies, sometimes leading to diverse interpretations of theory, coding schemes, organization, variables, and definitions. Researchers often do not disclose sufficient methodological information to replicate, verify or compare findings, including access to the codebooks, information on case or variable selection, theoretical assumptions, or intercoder reliability testing approaches. Problems associated with ambiguous or missing

[2] The term "schema" is defined as the organization and structure for a database as often used in computer programming literature (Roberts, 2005).

information based on unreported assumptions hamper the replicability of study findings and undermine the reliability and validity of such research.

Research is always a work in progress and case studies and comparative analysis done in isolation may be disputed or later found to be wrong. In addition, there may be issues of confirmatory bias or non-representative sampling involved in the selection of cases for secondary analyses, even when they contain sufficient levels of information. Thus, intercoder reliability testing and reporting is critically important, as is the disclosure of coding variables and codebooks. In order to advance the intra- and inter-institutional analysis of data, more rigorous standards should be established, such as common standards and protocols and the explicit reporting of assumptions. Even without consensus on standards or protocols, however, selection criteria should be made transparent by research teams in order to facilitate the emergence of common practices and increased methodological rigor in environmental social science in general.

Access to the resource of SES and commons data that currently exists can, itself, be viewed as a public good which is currently underprovided due to lack of transparency and coordination. Institutions which govern the proper and productive use of these resources could effectively reduce issues which private property dataset approaches now generate. The differences in databases and lack of transparency by researchers limit synthesis efforts and the ability to conduct broader, large-N case comparisons. Agrawal (2014) asserts that furthering this research will require methodological innovation, better theoretical sophistication and improved data. Furthermore, he states that the use of new methods involving more qualitative analysis and experimentation are the current drivers pushing the field forward. However, the successful use of these new methods will depend upon substantial amounts of new data, better integration of data, a sophisticated hierarchical organization of datasets, and increased analytical rigor (Agrawal 2014). In order to increase coding replicability, reliability, and transparency, some scholars assert that explicit identification and alignment of the coding rules, organization and work-process knowledge (or coding schema) used in coding may be important in mitigating problems of missing data and interpretations of concepts (MacQueen et al. 1998; Stemler 2001; Medjedović and Witzel 2005). Because meta-analysis of this type is a relatively new methodological approach in social science research (Corti et al. 2005), some authors argue that there has not yet been enough published research looking at the issues it may raise (Corti and Thompson 2004). In this paper, we critically explore our experience in answer to these challenges. We hope to offer some guidance and identify valid issues of consideration in the coding of secondary data for meta-analysis, thereby contributing to the dialogue in this area.

3. Coding methodology

In order to increase the replicability and the transparency of our coding process we have created a detailed Coding Manual (see Supplementary Material) and a

Recommended Coding Protocol (see Section 4). A coding protocol is the common set of systematic procedures that a research team agrees to follow during the coding process (Rourke and Anderson 2004) and a coding manual typically contains the coding questions, answer codes, and information to aid in clarification and coder alignment which embody the research questions being explored in a study (MacQueen et al. 1998). Our coding manual was developed incrementally throughout the coding process and our recommended coding protocol outlines the way that we would conduct the project in retrospect, resulting from the analyses and comparison to the methods literature as detailed in the following sections. Figure 1 (below) illustrates how our process compares to practices recommended in the methods literature (MacQueen et al. 1998; Mayring 2000; Hruschka et al. 2004). We then discuss the comparison between the recommended "best practices" model synthesized from the methods literature (left side of Figure 1) and the process used by the CBIE team (right side of Figure 1), focusing on the challenges raised during the coding process and how the recommendations from the methods literature could potentially address them.

3.1. Formulate research agenda

The formulation of the research agenda for the original meta-analysis project at CBIE (Baggio et al. 2016) was related to three objectives. The primary objective of that study was to examine case studies to determine whether particular configurations of Ostrom's (1990) design principles (DPs) were indicative of successful CPR governance. The second objective was to replicate and then expand upon a previous study conducted by Cox et al. (2010), which provided some empirical support for the claim that there is a higher chance for each of Ostrom's (1990) individual DPs to be present in successful cases of CPR management across a range of contexts. The third objective was to link the expanded DPs (Table 1) found in Cox et al. (2010) with variables found within the existing database for the Common Pool Resources (CPR) Project (Ostrom et al. 1989). Since the DPs and the variables used in the CPR database are both founded on CPR theory, we thought it would be possible to link them, thereby facilitating the synthesis of two separate datasets that use similar concepts but different coding schema. Larger datasets of comparable cases improve meta-analyses and researchers' ability to use mixed qualitative and quantitative methods, as well as improve analyses across multiple sectors, scales, and time periods. In doing so, our ability to understand complex system dynamics and adaptation in these system types is potentially enhanced (Poteete et al. 2010).

3.2. Identify dataset

Decisions about case selection and subsequent text segmentation are extremely important steps in the identification of the dataset to be used for meta-analysis (Hinds et al. 1997; Stemler 2001; Weed 2005). Cases should typically be

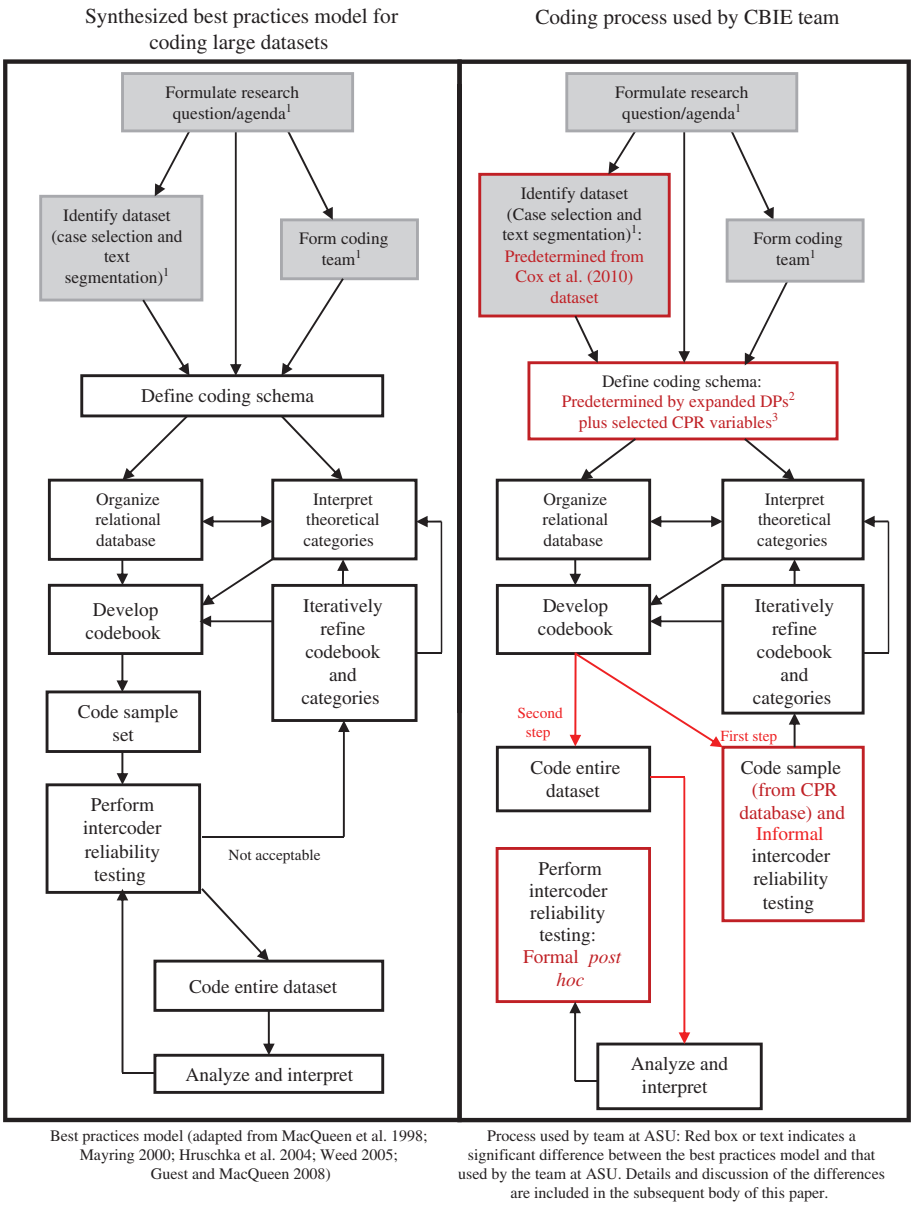


Figure 1: Coding process comparison illustrating the process utilized by our team compared to the "best practices" model described above and discussed in further detail in the following sections. [1]Preliminary decisions; [2]Table 1, this document; [3]Ostrom et al. 1989.

screened and analyzed for fit based on both their applicability to the research questions and data completeness (Hinds et al. 1997; Stemler 2001; Weed 2005). Longer texts, like the case studies used in this study, should be segmented into

smaller units of text (e.g. a sentence or a paragraph) to increase intercoder agreement and reliability (Krippendorff 2013) and decrease coding discrepancies (Hruschka et al. 2004). A coding protocol generally includes guidelines as to how a text should be segmented for data analysis and coding (Hruschka et al. 2004; Bernard and Ryan 2010; Bernard 2011). Inclusion and exclusion criteria formally clarify the reasoning behind the selection of cases and segmentation of texts (Hruschka et al. 2004). Ostrom et al. (1989) found exclusion criteria to be extremely important and included careful screening criteria for the cases included in the original CPR database.

Because the primary and secondary objectives of the CBIE team's research agenda were to replicate and extend upon the findings of a previous study, the selection of cases was predetermined by the dataset used in the study by Cox et al. (2010). Consequently, this limited our ability to select cases for fitness and data completeness. We did, however, limit our selection of cases to a sub-set of the Cox et al. dataset by sector (irrigation, fishery, and forestry), based on our third objective of synthesis with the existing CPR dataset (Ostrom et al. 1989). This resulted in the coding of 69 out of the 77 cases presented in Cox et al. (2010). During the coding process, our team experienced some difficulties with the fitness of the dataset due to missing data. For example, there were some cases which we found had sufficient social outcome data but not enough biological data, or vice versa, making the overall determination of success or failure in these cases difficult. Without explicit information on the inclusion/exclusion criteria used by the Cox team, it was more difficult for us to replicate and validate findings of success or failure across cases. We also found that some cases had ample data on one or two specific DPs but lacked information on the presence or absence of others. The Cox study may have been less sensitive to missing data on DPs because they were analyzing individual DPs against success, rather than looking for combinations of DPs as in the CBIE approach (Baggio et al. 2016). While analyzing combinations of DPs may present increased issues with missing data, Baggio et al. show the potential advantages of this approach.

Cox et al. (2010) segmented text by dividing longer documents into individual cases representing a single geographical location and temporal period. The text segmentation for the CBIE study was pre-determined by the divisions made in Cox study, and inter-related with case selection and the issues previously described. We found that the segmentation of texts contributed to the issues of missing data and fitness because some cases might include a single paragraph within a larger document or might instead include a number of sentences or excerpts related to a specific location scattered throughout the document which were considered one segment. Since criteria for the segmentation of texts into cases from larger regional studies was not explicitly reported in the Cox et al. publication, the CBIE team initially debated whether to include or exclude cases based on our own screening criteria, but ultimately decided to use the same cases that were also evaluated by the Cox team.

3.3. Form coding team

The use of two or more coders is important for assessing the replicability and reliability of coded data (MacQueen et al. 1998). The number of coders sufficient to establish reliability is not agreed upon in the literature, but in general, the more coder inference required and/or the rarer that codes appear in texts, the greater the number of coders that should be utilized (Bernard and Ryan 2010). We divided all 69 cases among the entire coder team assuring that there were generally three coders per case. This resulted in eighteen distinct coding team combinations. Since our coding project involved case studies that reported on SES conditions from a variety of perspectives requiring a certain amount of coder inference, utilizing three coders, rather than just two, was an appropriate and beneficial design feature.

3.4. Define coding schema (categories and organization)

Definition of the coding schema for a comparative or meta-analysis project involves the theoretical interpretation of categories and organization of the relational database (MacQueen et al. 1998; Mayring 2000; Hruschka et al. 2004; Weed 2005; Guest and MacQueen 2008). The theoretical interpretation of categories refers to a deductive approach to specifying themes, codes, or variables which will be searched for and coded within the texts and which are based on a defined body of theory (Weed 2005). The organization of the relational database simply refers to the way that the data will be organized in the database.

The primary coding categories used within our study were derived from the expanded design principles defined by the Cox et al. (2010) study (Table 1). Araral (2014) argues that there are two specification problems in the Cox et al. (2010) study that may also apply to our study. Araral's (2014) first concern is the re-specification of Ostrom's (1990) DP for clear boundary rules (DP1) into two distinct DPs for user boundaries (DP1A) and resource boundaries (DP1B) (Cox et al. 2010). Araral (2014) asserts that Ostrom (1990) intentionally did not separate the original design principle in this manner because within the "context of collective action in the commons" (p. 18), boundaries refer to enforceable property rights, not spatial boundaries. He also points out that the relevant critical literature has previously illuminated that spatially based definitions of community are problematic because the "overlapping, fuzzy and temporal nature of rights" can lead to difficulties in defining community across scales (Araral 2014). This issue has been previously illuminated in the relevant literature, with claims that spatially based definitions of community are problematic because the overlapping and temporal nature of rights can potentially lead to difficulties in defining community across scales (Brewer 2012; Araral 2014; Barnett 2014). Others, however, have suggested that this is a faulty argument and that the distinction made by Cox et al. (2010) is a helpful tool in defining clear agent boundaries (Pitt et al. 2012). Ostrom (1990) stated that "Without defining the boundaries of the CPR and closing it to 'outsiders', local appropriators face

Table 1: Expanded design principle questions (adapted from Cox et al. 2010) as basis of coding variables and questions.

Design principle	Description
1A	The presence of the design principle 1A means that individuals or households who have rights to withdraw resource units from the common-pool resource must be clearly defined. Is this design principle present?
1B	The presence of the design principle 1B means that *the boundaries of the CPR must be well defined*. Is this design principle present?
2A	The presence of design principle 2A means that *appropriation rules restricting time, place, technology, and/or quantity of resource units are related to local conditions*. Is this design principle present?
2B	The presence of design principle 2B means that *the benefits obtained by users from a CPR, as determined by appropriation rules, are proportional to the amount of inputs required in the form of labor, material, or money, as determined by provision rules*. Is this design principle present?
3	The presence of design principle 3 means that *most individuals affected by the operational rules can participate in modifying the operational rules*. Is this design principle present?
4A	The presence of design principle 4A means that *monitors are present and actively audit CPR conditions and appropriator behavior*. Is this design principle present?
4B	The presence of design principle 4B means that *monitors are accountable to or are the appropriators*. Is this design principle present?
5	The presence of design principle 5 means that *appropriators who violate operation rules are likely to be assessed graduated sanctions (depending on the seriousness and context of the offense) by other appropriators, officials accountable to these appropriators, or both*. Is this design principle present?
6	The presence of design principle 6 means that *appropriators and their officials have rapid access to low-cost local arenas to resolve conflicts among appropriators or between appropriators and officials*. Is this design principle present?
7	The presence of design principle 7 means that *the rights of appropriators to devise their own institutions are not challenged by external governmental authorities*. Is this design principle present?
8	The presence of design principle 8 means that *appropriation, provision, monitoring, enforcement, conflict resolution, and governance activities are organized in multiple layers of nested enterprises*. Is this design principle present?

the risk that any benefits they produce by their efforts will by reaped by others who have not contributed to those efforts" (p. 91). The definition of the CPR boundary can be seen as the definition of the spatial boundary (DP1A), while the exclusion of "outsiders" can be seen as the definition of the user boundary (DP1B).

Araral (2014) also points to the definition of a "successful CPR" as the second specification error of concern. Our team found that the definition of success and failure are complex, and ended up using a different approach than that reported by Cox et al. (2010). Cox et al. (2010) defined "success" in cases that "reported successful long-term environmental management" (Cox et al. 2010, 40), while we define success according to a number of dimensions defined by social and ecological outcomes variables (Table 2) drawn from the CPR project coding schema

Table 2: Coding variables/questions and categories.

Outcomes variable categories					
Resource sustainability (12 variables)	Process of collective choice arrangements (4 variables)	Equity among users (4 variables)	Overall success/failure of the CIS (1 variable)		
1a and 1b: Quality of units being withdrawn 2a and 2b: Maintenance of public appropriation infrastructure 3a and 3b: Maintenance of public distribution infrastructure 4a and 4b: Maintenance of public production infrastructure 5a and 5b: Balance of resource availability and withdrawal 6a: Changes in condition of natural infrastructure 6b: Changes in condition of human-made hard infrastructure	7a and 7b: Levels of trust among appropriators 8: Changes in trust level 9: Rule following	10: Disadvantaged appropriators 11: Harm to those who are worst off 12: Distance between least and most advantaged 13: Changes in the levels of equity among appropriators	14: Success or failure		
Expanded design principle variable categories					
DP1A (2 variables)	DP1B (2 variables)	DP2A (2 variables)	DP2B (2 variables)	DP3 (7 variables)	
15: Well defined group 16: Presence or absence of DP	17: Spatial boundary construction 18: Presence or absence of DP	19: Rule flexibility 20: Presence or absence of DP	21: Rule fairness 22: Presence or absence of DP	23: Options to express needs to decision makers 24, 24.1 and 24.2: Chief exec. position 25 and 25.1: Proposed collective choice rules 26: Presence or absence of DP	
DP4A (5 variables)	DP4B (4 variables)	DP5 (3 variables)	DP6 (2 variables)	DP7 (4 variables)	DP8 (3 variables)
27 and 27.1: Records of use 28 and 28.1: Records of resource condition 29: Presence or absence of DP	30: Self-monitoring 31 and 31.1: Official guard 32: Presence or absence of DP	33 and 33.1: Sanctions vary 34: Presence or absence of DP	35: Arenas for exchange of info 36: Presence or absence of DP	37, 37.1 and 37.2: Right to participate in management 38: Presence or absence of DP	39: Chief exec. report externally 40: More than one organization 41: Presence or absence of DP

(Ostrom et al. 1989), including: 1) resource sustainability (variables 1a–6b); 2) process of collective choice arrangements (variables 7a–9); 3) equity among users (variables 10–13); and 4) overall assessment of Success or Failure for the case (variable 14). Overall success (used in Baggio et al. 2016 and Barnett et al. 2016) was then coded as "success" when the resource was utilized sustainably, and there was an absence of conflict among resource users. We also used CPR variables to augment each DP variable, making each DP a theoretical category. Fifty-seven variables, in total, were specified and divided into 15 categories; one for each of the four dimensions of outcome "success" and the 11 expanded design principle categories (Table 2).

The specification of success may be a fundamental issue in our field (Araral 2014). Ostrom (1990) defined "success" within CPR governance as those "institutions that enable individuals to achieve productive outcomes in situations where temptations to free-ride and shirk are ever present" (p. 15). "Institutions" are the rules, norms, and shared strategies that people use to organize all forms of repetitive and structured interactions at all scales (Ostrom 2005). When Ostrom talks about "success," she is referring to successful collective action. Cox et al. (2010) used this definition, stating that cases were coded as unsuccessful if there was a "clear failure in collective action and management" (p. 40). Both the Cox et al. definition and the outcomes variables, which we used to construct our definition of success, capture this part of Ostrom's (1990) definition. The major difference in Cox et al., however, comes from including the idea of "long-term environmental management" (Cox et al. 2010, 40) which is not included within the outcome variables used in our study. While the idea of long-enduring CPR institutions is well founded within the literature (Ostrom 1990, 2005; Anderies et al. 2004; Cox et al. 2010; Poteete et al. 2010), we found this to be a difficult concept to assess within the meta-analysis of secondary data. Most cases in the dataset only captured a limited snapshot in time and did not include adequate longitudinal data to indicate the longevity of success within the case. In addition, Cox et al. divided some texts into separate cases for a single location but different time periods, which further limited any temporal analysis of success.

Agrawal (2014) has argued that commons scholars have not clearly differentiated between different measures, dimensions, and outcomes but have relied upon relatively vague terms like "sustainability", "success", and "long-term viability" instead. This raises fundamental questions within our field about what constitutes appropriate longevity for an assessment of success in a case and/or across comparative cases. Ambiguities involved in the specification of variables and problems with the definition of success and longevity assessments in cases made it difficult to reproduce the results of the Cox et al., study and hindered our synthesis and meta-analysis efforts. Specification problems, like these, are often key drivers of the missing data problem in studies which can plague both analysis efforts and intercoder agreement and require further dialogue within the field of research (Araral 2014).

3.5. Develop codebook and code sample set

According to the methods literature, sample coding should typically be performed on a random sub-set of the dataset and coding questions should be iteratively refined until intercoder reliability testing results are deemed satisfactory (Mayring 2000; Hruschka et al. 2004). Sample coding is the testing of the coding schema on a small random sample of the data to facilitate iterative refinement prior to the coding of the full dataset. The variables described in Table 2 (above) were initially documented in a set of preliminary coding questions and were pre-tested on a sample of three cases representing each sector (fisheries, forestry, irrigation) randomly selected from the existing CPR database. This allowed us to compare current coding[3] results with those of the original coding conducted by Ostrom's team (1989) and determine consistency in the interpretation of the CPR variables. Although the three sample cases from the CPR database were not a part of the dataset for the meta-analysis project, this allowed us to more accurately assess alignment with the CPR variables thereby providing a measure of inter-coder agreement. Coding results from the pre-test sample coding were subjected to formal intercoder reliability testing by one of the primary investigators of the project before coding the entire dataset commenced. Any questions related to further interpretation of variables were discussed and clarified by the entire research team during periodic meetings as an informal means of increasing intercoder alignment. Issues clarified in project meetings were then incorporated into a preliminary coding guide which included the questions for each of the original 57 coding variables supplemented with explanations and answers derived from coder questions and team discussions.

3.6. Perform intercoder reliability testing and iteratively refine

The best practices model (Figure 1) recommends formal intercoder reliability testing on a subset of the dataset, as well as iterative intercoder agreement testing throughout and after the formal coding process (MacQueen et al. 1998; Mayring 2000; Hruschka et al. 2004; Guest and MacQueen 2008). We have found that this step is often missing from reports on studies of CPRs using meta-analyses (Netting 1976; Wade 1984; Berkes 1989; Ostrom 1990; McKean 1992; Baland and Platteau 1999; Bardhan and Mookherjee 2006; Cox 2014; Epstein et al. 2014; Fleischman et al. 2014; Villamayor-Tomas et al. 2014). Hruschka et al. (2004) explain that a reluctance to assess coder agreement is common in some branches of social science because: (1) researchers may generally believe that the quantification of qualitative data is unnecessary because qualitative research is a "distinct paradigm" that cannot or should not be subject to a quantitative evaluation; and

[3] Results from the sample coding of the three CPR cases were compared to the original results for those cases contained within the relational database for the CPR Project (Ostrom et al. 1989) and so were comparable with only those variables extracted from the CPR project (45 variables), not including the "Success" variable or any of the 11 expanded design principle variables.

(2) there is a general skepticism about the ability to actually measure subjective data and reproduce coding results. We believe the latter argument to be the most viable reason for the apparent lack or under-reporting of intercoder reliability testing in our field but have found that it would potentially be helpful when iteratively included throughout the coding process.

Our team only tested intercoder agreement on the initial sample set of CPR cases and did not test for intercoder reliability again until the analysis and interpretation phase of the project. Our informal coding guide development process was aimed at establishing an informal feedback loop of intercoder alignment, refinement of theoretical interpretations and iterative adjustments to the coding questions based on ambiguities and questions that arose during the coding process. Assessment of coding conducted in other studies (Ostrom et al. 1993; Wollenberg et al. 2007; Cox 2014) suggests that this is a more common practice in our research community than the more formal methods. Hruschka et al. (2004) recognize this consensus-based approach toward "interpretive convergence" (p. 321) as a potentially useful method for increasing intercoder reliability, but state that more analysis may be needed to determine the validity of this approach.

3.7. Code dataset

Coding is the essential activity of the content analysis methodology and requires the identification of themes or categories that appear in text or other media segments (Hruschka et al. 2004). Coding can be done in a number of ways ranging from highlighting pieces of text by hand to the use of sophisticated Qualitative Data Analysis (QDA) software packages. While QDA software is sometimes expensive and requires training, some studies have found that use of QDA software has been found to aid in increasing rigor and intercoder reliability during the coding process (Denzin and Lincoln 2000; Rambaree 2007), allowing coders to identify and tag specific text segments and associate them with a particular category or memo. Texts coded by individual coders can later be combined and analyzed, thus allowing for easier identification of coding discrepancies (Bernard and Ryan 2010). In contrast, hand-coding and/or use of spreadsheet software is inexpensive and requires little to no additional coder training.

Individual coders on the CBIE team coded text segments which they felt exhibited explicit evidence supporting their answer to each of the 57 coding questions and documented the answer to the question and the supporting text segment(s) in spreadsheet format (Table 3). QDA software was not used in the CBIE project due to time and cost constraints. Each team of three coders then met to compare answers and decide upon a single group code, reducing the subjectivity of codes and generating more reliable coding (Hruschka et al. 2004). Where there was consensus on the answer to a coding question between the individual coders on any variable, the same answer was given as the group code for that variable. Selected text segments were then utilized as "evidence" of an appropriate code when mitigating discrepancies between team members to arrive at an

Table 3: Example of coding results by case study (column SECDESC) and coding group (AEN).

SECDESC	Cox case	Group	Coder	1a.BEGQUAL	1b.ENDQUAL	NotesQUAL	2a.BEGCONDA
Rural coastal fishing village 50 km south of Mombasa, Kenya	1	AEN	A	−1	−1	It says the stocks decline but does not mention the quality of the fish	−2
Rural coastal fishing village 50 km south of Mombasa, Kenya			E	−1	4	pg. 2773 mentions the decline of fish stocks in the area	−2
Rural coastal fishing village 50 km south of Mombasa, Kenya			N	−1	3		−1
Rural coastal fishing village 50 km south of Mombasa, Kenya			Group	−1	-1		−1

The coding results displayed are the codes for individual coders "A", "E", "N", as well as the agreed-upon "Group" code. The blue color of column "1b.ENDQUAL" indicates a disagreement between coders which was resolved by group agreement for the resulting group code of −1 MIC, indicating that the group decided that there was not enough information in the text to make a decision.

agreed-upon group code. Any coding disagreements were resolved through group discussion among the coding team members and during project meetings where study PIs addressed unresolved issues. Final coding results for all cases were later combined into a single master spreadsheet.

3.8. Analyze and interpret results – *post hoc* intercoder reliability testing

Analysis of coding team dynamics and formal *post hoc* intercoder reliability testing[4] (see Supplementary Material) were conducted along with other analyses for the meta-analysis study (Baggio et al. 2016; Barnett et al. 2016). Results showed potential inconsistencies in intercoder agreement and coding team dynamics may have developed from the informal consensus-based process used by the CBIE team. The informal methodology may have resulted in distinct advantages for coders who were able to more forcefully argue their positions or better document all instances of text that led them to code a variable in a

[4] Reflexive analyses, select social network analysis, and intercoder reliability testing were performed to better understand the coding team dynamics and coding processes.

certain way, highlighting the need for explicit rules of coding and for increased attention to both intercoder agreement and reliability (MacQueen et al. 1998; Stemler 2001).

Post hoc intercoder reliability ratings were calculated to examine the overall intercoder agreement by team, but also to determine which coding variables were more difficult to identify within the texts (see Baggio et al. 2016). We found that inconsistencies the challenges discussed above contributed to low intercoder reliability ratings, but that these challenges are not insurmountable. They should be considered part of a normal coding process and are typical of many similar projects within our field of study. Coder agreement is generally expected to be low initially, particularly when coding "focuse[s] on identifying and describing both implicit and explicit ideas" (Namey et al. 2008, 138), such as inferring the presence or absence of DPs in case studies. The fact that many case studies in our dataset were lengthy texts may have further contributed to marginal intercoder agreement. These challenges can be decreased through more formal methods, like the "best practices" model presented here (Figure 1). For example, to address discrepancies in coder interpretation, the literature recommends coding several iterations of subsets of the data, followed by formal reliability testing (percent agreement and a Kappa statistic that takes chance into account) and iterative codebook revisions until acceptable intercoder reliability ratings have been reached (Hruschka et al. 2004; MacQueen et al. 2008; Bernard 2011). Once acceptable intercoder agreement has been reached, coding of the entire dataset proceeds which is supplemented by continued random sample intercoder reliability testing to prevent "coder drift" or "code favoritism" (Carey and Gelaude 2008, 251).

3.8.1. Data preparation

Post hoc intercoder reliability testing required considerable data preparation in order to unify coding data, minimize bias due to incompatible comparisons, and transfer complex coding values into a format that could be analyzed by intercoder reliability statistical software. Details of these processes are outlined in the Supplementary Material.

3.8.2. Intercoder reliability testing

For coding projects involving >2 coders and coding values that are nominal and multiple, Feng (2014) recommends Krippendorff's alpha, Fleiss' kappa, and/or percent agreement. Krippendorff's alpha is a reliability coefficient that is a "generalization of several known reliability indices" (Krippendorff 2013, 1). Its advantage lies in its ability to calculate intercoder agreement among an indefinite number of coders and any number of scale values. It can handle missing and incomplete data, as well as large and small sample sizes and is considered a robust measure of intercoder reliability (Bernard and Ryan 2010; Krippendorff 2013). Fleiss' kappa is a variant of the popular Cohen's kappa statistic which allows for more than two coders (Bernard and Ryan 2010). Similar to Krippendorff's

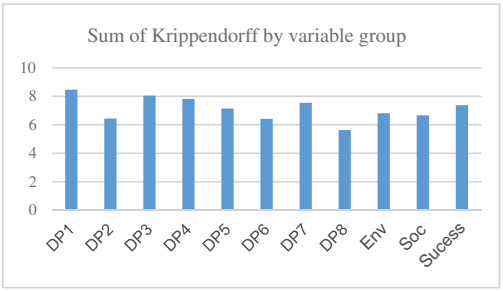
alpha, Fleiss is a statistic that measures coders' agreement with respect to chance (Bernard 2011). Finally, although simple percent agreement tends to overestimate intercoder reliability because it does not account for chance agreement (Hruschka et al. 2004; Feng 2014), it is appropriate to utilize this technique in conjunction with other measures if the variables analyzed are nominal (Feng 2014). Simple percent agreement provides a good yardstick to determine whether the intercoder reliability ratings obtained through Krippendorff and Fleiss may be skewed due to particularly high agreement or missing variables.

Utilizing the irr-package in R (Gamer et al. 2012), intercoder agreement for all three statistics was calculated for 11 variable groups in each of the 13 coding teams (see Table 4 for excerpt and the Supplementary Material for complete intercoder reliability ratings and R code). Before evaluating whether coding agreement reached high (>0.80) or acceptable (0.70–0.79) reliability levels, simply adding the Krippendorff and Fleiss values by variable group and coding team provides a first insight into those variable groups/teams with high/low scores. For the Krippendorff values, Figures 2 and 3 reveal DP1 (clearly defined boundaries) and coding team "AEN" as those with the highest intercoder agreement. In contrast, DP8 (nested governance) and team "ACH" had the lowest intercoder agreement. Fleiss' statistics mirrored those findings (see Supplementary Material). This suggests that determining the evidence of resource and user boundaries within a case study requires less inference from coders than determining whether the

Table 4: Excerpt of intercoder reliability testing results (all statistics).

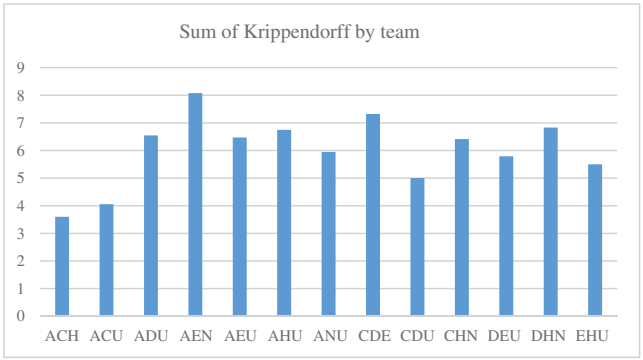
Coding team	Variable group	Krippendorff values	Fleiss values	Percent agreement
ACH	Env	0.603	0.602	80.60
ACH	Soc	0.693	0.692	68.80
ACH	Success	1.000	1.000	100.00
ACH	DP1	0.261	0.256	33.30
ACH	DP2	0.327	0.322	37.50
ACH	DP3	0.387	0.384	64.30
ACH	DP4	0.591	0.590	59.30
ACH	DP5	–0.138	–0.149	50.00
ACH	DP6	–0.241	–0.258	16.70
ACH	DP7	0.389	0.385	50.00
ACH	DP8	–0.274	–0.286	33.30
CHN	Env	0.636	0.634	66.70
CHN	Soc	0.507	0.503	45.80
CHN	Success	–0.063	–0.125	66.70

Column "coding team" identifies the coding team. Column "variable group" identifies the coding variable categories/groups, i.e. "env" = variables 1a–6b; "soc" = variables 7a–13; "success" = variable 14; DP1 = variables 15–18; DP2 = variables 19–22; DP3 = variables 23–26; DP4 = variables 27–32; DP5 = variables 33–34; DP6 = variables 35–36; DP7 = variables 37–38; and DP8 = variables 39–41. Values for Krippendorff's alpha and Fleiss' kappa range between 0 and 1, with 1 demonstrating perfect agreement between coders and 0 indicating agreement that is consistent with chance, i.e. the absence of reliability. Negative alpha values signify coder agreement that is below chance (Krippendorff 2008).



Variable group	Sum of Krippendorff
DP1	8.462
DP2	6.431
DP3	8.043
DP4	7.813
DP5	7.136
DP6	6.416
DP7	7.543
DP8	5.626
Env	6.812
Soc	6.662
Success	7.370

Figure 2: Sum of Krippendorff values by variable group for all coded cases. Results indicate that generally Design Principle 1 (DP1) had the highest overall intercoder agreement and Design Principle 8 (DP8) the lowest.



Team	Sum of Krippendorff
ACH	3.598
ACU	4.055
ADU	6.545
AEN	8.078
AEU	6.473
AHU	6.749
ANU	5.947
CDE	7.324
CDU	5.006
CHN	6.416
DEU	5.792
DHN	6.831
EHU	5.500

Figure 3: Sum of Krippendorff values by coding team/all cases coded. Results reflect highest coder agreement for team AEN and lowest coder agreement for team ACH.

reported institutional structure represents a "nested enterprise." For codebook and coding protocol development purposes, such initial high/low values could be important bellwethers of particularly well or poorly functioning coding questions/teams, identifying weaknesses that may require further investigation in order to strengthen intercoder agreement before commencing with coding the entire dataset.

Despite the aforementioned problems, many of the intercoder agreement ratings were >0.65 for both Krippendorff and Fleiss statistics. This places our data reliability/replicability factor only slightly below the 0.70 score that is generally deemed as acceptable in the literature. Given the subjective nature of some of the variables, the large number of missing values, and the iterative nature of our coding process, such ratings are defensible for the completed project and may easily be improved in the future through the use of a more detailed codebook and coding protocol. More importantly, by disclosing our intercoder reliability ratings,

procedures, preliminary codebook and coding protocol, we have taken additional steps to enhance the ability of others to analyze and replicate our findings as well.

3.8.3. Coder drift

One important reason we found to assess intercoder reliability is known as "coder drift". Coder drift is the process over time, in which coders may become less reliable in their coding due to the adoption of coding biases and the not-rigorous application of coding criteria (Bartholomew et al. 2000). To avoid coder drift, Carey and Gelaude (2008) recommend spot checking of coder agreement throughout the coding process. After coding was completed and intercoder reliability ratings performed, discussion among coders revealed that there may have been some coder drift which could have produced inconsistencies in the way that coders applied information within the text to answer the question of overall success (variable 14). In our study, spot checks of coder agreement throughout the coding process may have mitigated some of the ambiguity with regard to coders' assessment of "success". Subsequent random sampling of the answers given to question 14, as well as purposive sampling of an additional ten cases, revealed notes that indicated several coders may have considered more than the outcome variables in their answer to this question. However, in all but two cases, coders were in agreement with their assessment of the studies overall success or failure, regardless of the potential for coder drift. In the two instances of coder drift where there was no initial coder agreement, the coders were able to resolve the disagreement through discussion. As outlined throughout this paper, a codebook containing detailed coding descriptions that is iteratively updated to include coder questions and coding ambiguities, as well as continuous spot-checking of intercoder agreement, might have resolved these instances of coding bias.

4. Recommended coding protocol

Through analysis of our coding process and review of the literature, we have found that increased transparency, reliability, and replicability are of primary importance in increasing our ability to perform meta-analysis and the synthesis of case studies. While qualitative research often generates complex information that is difficult to process and can lead to judgments based on subjective, or "intuitive heuristics" (Hruschka et al. 2004), the level of agreement can and should be quantified. It is precisely the subjective nature of the evaluations which makes them more susceptible to individual interpretation and the intentional or unintentional introduction of biases, random errors, and other distortions (Hruschka et al. 2004; Krippendorff 2013). The establishment of more rigorous coding protocols including intercoder reliability testing represents an effort to "reduce [such] error and bias" (Hruschka et al. 2004) by ensuring that the data meaning remains consistent across a variety of coders and research teams. In fact, it can be argued that coding is an essential element of classical content analysis because it converts qualitative data into datasets that are supportive of

robust analyses and can be replicated by other scholars (Krippendorff 2013). Replicability creates greater reliability which empirically grounds confidence in the data and, thus, the study findings (Krippendorff 2013). For these reasons, we include here our *Recommended Coding Protocol* (Figure 4). This is based, in retrospect, on the examination of the CBIE meta-analysis project, but we will briefly discuss the considerations which may be affected by project and team type. More detailed information on all steps outlined here can be found in the *Detailed Recommended Coding Protocol* included within the Coding Manual in the Supplementary Material.

4.1. Preliminary considerations

We found a number of preliminary considerations (gray boxes in Figure 4) which should precede the coding process.

4.1.1. Identify dataset

We highly recommend that teams develop a screening process during the identification of the dataset to ensure that cases included in the study have sufficient information to answer the research question. Inclusion/exclusion and text segmentation criteria should be clearly defined and reported. This step is likely to decrease missing and ambiguous data for analysis.

4.1.2. Select qualitative data analysis (QDA) software or other technique for coding

Teams should consider the use of QDA software prior to the commencement of the coding process. Although QDA software will add cost and training considerations to the project, it may facilitate data processing, decrease discrepancies, and potentially reduce the time needed to conduct intercoder reliability testing.

4.1.3. Form coding team

We found that utilizing two or more coders increases data reliability because coding agreement between different people, who have been given the same instructions and have independently coded the same text segments, demonstrates a reduction of subjective biases and increases data reliability (Guest and MacQueen 2008). Coding team dynamics may be a concern, however, such dynamics can be mitigated through the use of more rigorous coding protocols and coder training. Although each additional coder increases the need for iterative intercoder reliability testing and training to achieve intercoder alignment, two coders per text should be a necessary condition for any meta-analyses.

4.2. Coding process

4.2.1. Define coding schema

We recommend that coding schema definition include explicit consideration and documentation of the organization and work processes to be used during the

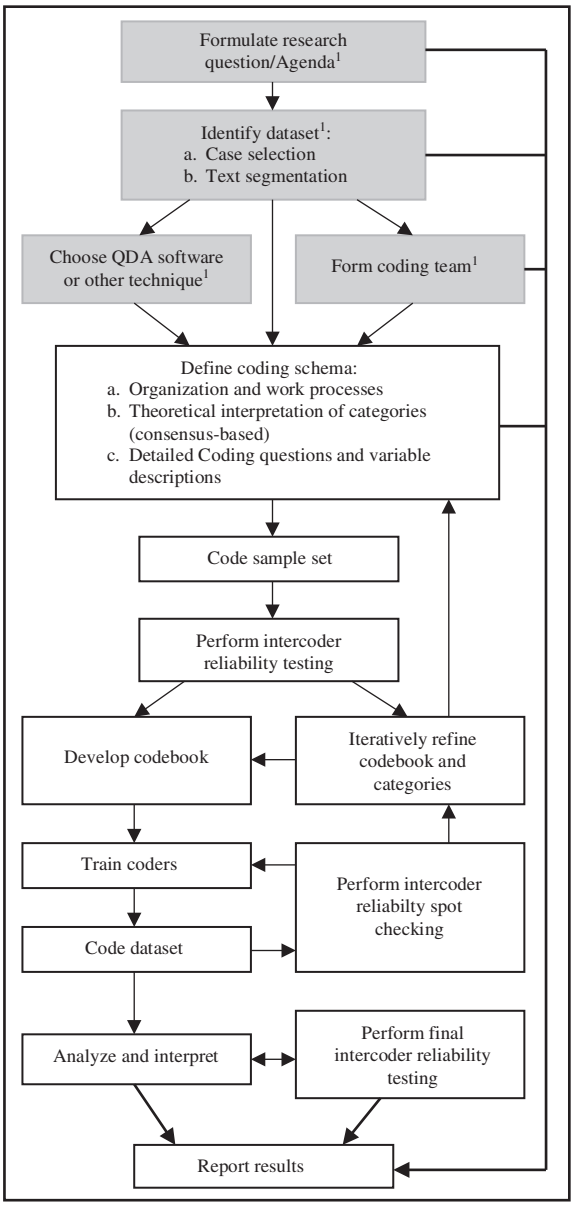


Figure 4: Recommended coding protocol. [1]Boxes shaded in gray represent preliminary considerations, while unshaded boxes are a part of the main coding process.

coding process, the development of detailed coding variable descriptions, and the iterative and consensus-based definition of theoretical categories by the entire coding team.

4.2.2. Sample coding and intercoder reliability testing

We recommend that the principal investigator and all coding team members independently test code a randomly selected subset of the actual dataset, followed by formal intercoder reliability testing of the results until acceptable levels of intercoder reliability ratings have been reached.

4.2.3. Codebook development, iterative refinement and training

We recommend that a consensus-based process of codebook development, based on the previous definition of the coding schema, sample coding, and intercoder reliability testing be included in the coding process. This can be considered part of coder training. Discussions on the development of codes and theoretical categories among the coding team will likely result in increased understanding of key issues and variables to be coded. Training should also include coder instruction in the use of any selected QDA software.

4.2.4. Coding with intercoder reliability spot checks

Once acceptable intercoder reliability ratings have been achieved through sample coding and iterative codebook refinement, the entire dataset can be coded. At least one spot-check should be performed during this process to assess coder drift.

4.2.5. Analyses and interpretation of results with final intercoder reliability testing

The coding process should be assessed, along with final intercoder reliability testing, after coding is complete. The results of these analyses should be reported in the final project outcomes.

4.2.6. Reporting of results

Results should include the analyses of the data produced by the coding process, such as that reported in Baggio et al. (2016) and Barnett et al. (2016), but should also include the explicit disclosure of assumptions made during the preliminary steps of the project, as well as an analyses of the coding process itself and final intercoder reliability testing.

5. Conclusions

Social-ecological systems (SESs) vary across spatial and temporal scales and studying them is critical to understanding governance challenges involving common-pool resources (CPRs). Scholars, like Agrawal (2014) and Araral (2014), see current trajectories within SESs research as fundamental, yet still in their infancy. Araral (2014), in particular, argues that Ostrom's theories may only be applicable to the special case of locally governed, small-scale commons and may not be easily generalized. The body of evidence collected within Ostrom's legacy has not been able to effectively assess natural resource issues at larger scales. We question whether there has been a sufficiently sizable body of data gathered and

analyzed, including information on larger-scale systems, multi-scalar governance structures, temporal dimensions, and other important factors with which to compare the existing studies, or if there are any sufficiently developed methods by which to conduct such comparisons.

It was one of Ostrom's (2005) deep convictions that SESs are composed of a set of universal building blocks which could be tapped to create adaptive and long-enduring governance systems. Work towards creating a methodology that will foster cooperation and cross-comparison of data could allow us to expand our understanding of these systems. By sharing our coding experience and protocols, we hope to stimulate the development of transparency norms within the commons research community which others may build upon as we move further toward the identification of these universal building blocks. It is important to continue pushing social-ecological science towards greater rigor and a greater understanding of the complex interactions that lead to successful outcomes. Towards this goal, we assert that methodology must be tested and refined for more precise measurement of the dependent and independent variables involved in SESs. Furthermore, the commons research community should work to ensure that studies are replicable and that different research teams are able to achieve similar answers. In conclusion, while there may be many challenges and opportunities associated with the coding and synthesis of case studies, increased collaboration and consensus in a few key areas within the research community may lead to new horizons and possibilities in understanding SESs and the commons.

6. Supplementary files

Supplementary File 1: http://doi.org/10.18352/ijc.652.s1 S1 Supplementary material. Dataset repository matrix, coding manual and protocols.

Literature cited

Anderies, J. M., M. A. Janssen, and E. Ostrom. 2004. A Framework to Analyze the Robustness of Social-Ecological Systems from an Institutional Perspective. *Ecology and Society* 9(1):18. [online] http://www.ecologyandsociety.org/vol9/iss1/art18/.

Agrawal, A. 2014. Studying the Commons, Governing Common-Pool Resource Outcomes: Some Concluding Thoughts. *Environmental Science and Policy* 36(2014):86–91. http://dx.doi.org/10.1016/j.envsci.2013.08.012.

Araral, E. 2014. Ostrom, Hardin and the Commons: A Critical Appreciation and a Revisionist View. *Environmental Science and Policy* 36(2014):11–23. http://dx.doi.org/10.1016/j.envsci.2013.07.011.

Baggio, J. A., A. J. Barnett, I. Perez-Ibara, U. Brady, E. Ratajczyk, N. Rollins, C. Rubiños, H. C. Shin, D. J. Yu, R. Aggarwal, J. M. Anderies, and M. A. Janssen. 2016. Explaining Success and Failure in the Commons: The Configural Nature of Ostrom's Institutional Design Principles. *International Journal of the Commons* 10(2):417–439, http://doi.org/10.18352/ijc.634.

Baland, J. M. and J. P. Platteau. 1999. The Ambiguous Impact of Inequality on Local Resource Management. *World Development* 27(5):773–788. http://dx.doi.org/10.1016/S0305-750X(99)00026-1.

Bardhan, P. and D. Mookherjee. 2006. *Decentralization and Local Governance in Developing Countries: A Comparative Perspective (Vol. 1)*. Cambridge, MA: MIT Press.

Barnett, A. J. and J. M. Anderies. 2014. Weak Feedbacks, Governance Mismatches, and the Robustness of Social-Ecological Systems: An Analysis of the Southwest Nova Scotia Lobster Fishery with Comparison to Maine. *Ecology and Society* 19(4):39. http://dx.doi.org/10.5751/ES-06714-190439.

Barnett, A. J., J. A. Baggio, H. C. Shin, D. J. Yu, I. Perez-Ibarra, C. Rubiños, U. Brady, E. Ratajczyk, N. Rollins, R. Aggarwal, J. M. Anderies, and M. A. Janssen. 2016. An Iterative Approach to Case Study Analysis: Insights from Qualitative Analysis of Quantitative Inconsistencies. *International Journal of the Commons* 10(2):467–494, http://doi.org/10.18352/ijc.632.

Bartholomew, K., A. J. Z. Henderson, and J. E. Marcia. 2000. Coded Semi-Structured Interviews in Social Psychological Research. In *Handbook of Research Methods in Social and Personality Psychology*, eds. H. T. Reis and C. M. Judd, 286–312. Cambridge, MA. Cambridge University Press.

Berkes, F. 1989. *Common Property Resources: Ecology and Community-Based Sustainable Development.* London, UK: Belhaven Press.

Bernard, H. R. 2011. *Research Methods in Anthropology: Qualitative and Quantitative Approaches*. Fifth Edition. Plymouth, UK: Altamira Press.

Bernard, H. R. and G. W. Ryan. 2010. *Analyzing Qualitative Data: Systematic Approaches*. Thousand Oaks, CA: Sage Publications, Inc.

Brewer, J. F. 2012. Don't Fence Me In: Boundaries, Policy, and Deliberation in Maine's Lobster Commons. *Annals of the Association of American Geographers* 102(2):383–402. http://dx.doi.org/10.1080/00045608.2011.641889.

Carey, J. W. and D. Gelaude. 2008. Systematic Methods for Collecting and Analyzing Multidisciplinary Team-Based Qualitative Data. In *Handbook for Team-Based Qualitative Research,* eds. G. Guest and K. M. MacQueen, 227–274. Lanham, MD: Altamira Press.

Corti, L. and P. Thompson. 2004. Secondary Analysis of Archived Data. In *Qualitative Research Practice*, eds. C. Seale, G. Gobo, J. F. Gubrium, and D. Silverman, 327–343. London: Sage. http://dx.doi.org/10.4135/9781848608191.d26.

Corti, L., A. Witzel, and L. Bishop. 2005. On the Potentials and Problems of Secondary Analysis: An Introduction to the FQS Special Issue on Secondary Analysis of Qualitative Data. *Forum: Qualitative Social Research (FQS)* 6(1):Art.49.

Cox, M. 2014. Understanding Large Social-Ecological Systems: Introducing the SESMAD Project. *International Journal of the Commons* 8(2):265–276. http://dx.doi.org/10.18352/ijc.406.

Cox, M. 2015. A Basic Guide for Empirical Environmental Social Science. *Ecology and Society* 20(1):63–77. http://dx.doi.org/10.5751/ES-07400-200163.

Cox, M., G. Arnold, and S. Villamayor-Tomas. 2010. A Review of Design Principles for Community-Based Natural Resource Management. *Ecology and Society* 15(4):38. [online] http://www.ecologyandsociety.org/vol15/iss4/art38/.

Denzin, N. K. and Y. S. Lincoln. 2000. *Handbook of qualitative research*. Second Edition. Thousand Oaks, CA: Sage Publications, Inc.

Epstein, G., M. Nenadovic, and A. Boustany. 2014. Into the Deep Blue Sea: Commons Theory and International Governance of Atlantic Bluefin Tuna. *International Journal of the Commons* 8(2):277–303. http://dx.doi.org/10.18352/ijc.410.

Feng, G. C. 2014. Intercoder Reliability Indices: Disuse, Misuse, and Abuse. *Quality and Quantity* 48:1803–1815. http://dx.doi.org/10.1007/s11135-013-9956-8.

Fleischman, F. D., B. Loken, G. A. Garcia-Lopez, and S. Villamajor-Tomas. 2014. Evaluating the Utility of Common-Pool Resource Theory for Understanding Forest Governance and Outcomes in Indonesia Between 1965 and 2012. *International Journal of the Commons* 8(2):304.

Gamer, M., J. Lemon, and P. Singh. 2012. *irr: Various Coefficients of Interrater Reliability and Agreement*. Retrieved from http://cran.r-project.org/web/packages/irr/index.html.

Guest, G. and K. M. MacQueen. 2008. Reevaluating Guidelines in Qualitative Research. In *Handbook for Team-Based Qualitative Research*, eds. G. Guest and K. M. MacQueen, 205–226. Lanham, MD. Altamira Press.

Hinds, P. S., R. J. Vogel, and L. Clarke-Steffen. 1997. The Possibilities and Pitfalls of Doing a Secondary Analysis of a Qualitative Data Set. *Qualitative Health Research* 7:408. http://dx.doi.org/10.1177/104973239700700306.

Hruschka, D. J., D. Schwartz, D. C. St. John, E. Picone-Decaro, R. Jenkins, and J. W. Carey. 2004. Reliability in Coding Open-Ended Data: Lessons Learned from HIV Behavioral Research. *Field Methods* 16(3):307–331. http://dx.doi.org/10.1177/1525822X04266540.

Kelder, J. 2005. Using Someone Else's Data: Problems, Pragmatics and Provisions. *Forum: Qualitative Social Research* 6(1):Art.39.

Krippendorff, K. 2008. Systematic and Random Disagreement and the Reliability of Nominal Data. *Communication Methods and Measure* 2(4):323–338. http://dx.doi.org/10.1080/19312450802467134.

Krippendorff, K. 2013. *Content Analysis: An Introduction to Its Methodology*. Third Edition. Thousand Oaks, CA: Sage Publications, Inc.

MacQueen, K. M., E. McLellan, K. Kay, and B. Milstein. 1998. Codebook Development for Team-Based Qualitative Analysis. *Field Methods* 10(2):31–36. http://dx.doi.org/10.1177/1525822X980100020301.

MacQueen, K. M., E. McLellan-Lemal, K. Bartholow, B. Milstein. 2008. Team-Based Codebook Development: Structure, Process, and Agreement. In *Handbook for Team-Based Qualitative Research*, eds. G. Guest and K. M. MacQueen, 19–135. Lanham, MD: Altamira Press.

Mayring, P. 2000. Qualitative Content Analysis: Research Instrument or Mode of Interpretation. *Forum: Qualitative Social Research* 1(2):Art.20. Retrieved

October 1, 2015 from http://www.qualitative-research.net/index.php/fqs/article/view/1089/2385.

McKean, M. 1992. Success on the Commons: A Comparative Examination of Institutions for Common Property Resource Management. *Journal of theoretical politics* 4(3):247–281. http://dx.doi.org/10.1177/0951692892004003002.

Medjedović, I. and A. Witzel. 2005. Secondary Analysis of Interviews: Using Codes and Theoretical Concepts from the Primary Study. *Forum: Qualitative Social Research* 6(1):Art.46.

Namey, E., G. Guest, L. Thairu, and L. Johnson. 2008. Data Reduction Techniques for Large Qualitative Data Sets. In *Handbook for Team-Based Qualitative Research*, eds. G. Guest and K. M. MacQueen, 137–161. Lanham, MD: Altamira Press.

Netting, R. McC. 1976. What Alpine Peasants have in Common: Observations on Communal Tenure in a Swiss Village. *Human Ecology* 4:135–146. http://dx.doi.org/10.1007/BF01531217.

Ostrom, E. 1990. *Governing the Commons: the Evolution of Institutions for Collective Action.* Cambridge, MA: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511807763.

Ostrom, E. 2005. *Understanding Institutional Diversity.* Princeton, N.J. Princeton University Press.

Ostrom, E. 2012. Why do We Need to Protect Institutional Diversity? [Keynote Lecture]. *European Political Science* 11(2012):128–147. http://dx.doi.org/10.1057/eps.2011.37.

Ostrom, E., A. Agrawal, W. Blomquist, E. Schlager, and S. Y. Tang. 1989. CPR Coding Manual. Bloomington, In: Indiana University, Workshop in Political Theory and Policy Analysis. https://seslibrary.asu.edu/sites/default/files/cprcodingmanual-fullwcovercopytoc.pdf.

Ostrom, E., P. Benjamin, and G. P. Shivakoti. 1993. Use of Case Studies and Structural Coding in a Relational Database for Storage and Analysis of Irrigation Institutions and Systems. Proceedings of an International Workshop held at the Institute of Agriculture and Animal Science (IAAS). Retrieved from http://dlc.dlib.indiana.edu/dlc/handle/10535/7996.

Pitt, J., J. Schaumeier, and A. Artikis. 2012. Axiomatization of Socio-Economic Principles for Self-Organizing Institutions: Concepts, Experiments and Challenges. *ACM Transactions on Autonomous and Adaptive Systems* 7(4):Art.39. http://dx.doi.org/10.1145/2382570.2382575.

Poteete, A. R., M. A. Janssen, and E. Ostrom. 2010. *Working Together: Collective Action, the Commons, and Multiple Methods in Practice*. Princeton, NJ: Princeton University Press. http://dx.doi.org/10.1515/9781400835157.

Rambaree, K. 2007. Bringing Rigour in Qualitative Social Research: The Use of CAQDAS. *University of Mauritius Research Journal* 13(A):1–16. Retrieved from http://www.gfmer.ch/GFMER_members/pdf/CAQDAS_Rambaree_2007.pdf.

Roberts, M. 2005. Definition: schema. Whatis.com. http://searchsqlserver.techtarget.com/definition/schema.

Rourke, L. and T. Anderson. 2004. Validity in Quantitative Content Analysis. *Educational Technology Research and Development* 52(1):5–18. http://dx.doi.org/10.1007/BF02504769.

Rudel, T. K. 2008. Meta-Analyses of Case Studies: A Method for Studying Regional and Global Environmental Change. *Global Environmental Change* 18(1):18–25. http://dx.doi.org/10.1016/j.gloenvcha.2007.06.001.

Savage, M. 2005. Revisiting Classic Qualitative Studies. *Forum: Qualitative Social Research* 6(1):Art.31.

Schlager, E. 2004. Common-Pool Resource Theory. In *Environmental Governance Reconsidered: Challenges, Choices, and Opportunities*, eds. R. F. Durant, D. J. Fiorino, and R. O'Leary, 145–175. Cambridge, MA: MIT Press.

Stemler, S. 2001. An Overview of Content Analysis. *Practical Assessment, Research and Evaluation* 7(17):137–146.

Villamayor-Tomas, S., F. D. Fleischman, I. Perez Ibarra, A. Thiel, and F. van Laerhoven. 2014. From Sandoz to Salmon: Conceptualizing Resource and Institutional Dynamics in the Rhine Watershed Through the SES Framework. *International Journal of the Commons* 8(2):361. http://dx.doi.org/10.18352/ijc.411.

Wade, R. 1984. Irrigation Reform in Conditions of Populist Anarchy: An Indian Case. *Journal of Development Economics* 14(3):285–303. http://dx.doi.org/10.1016/0304-3878(84)90060-9.

Weed, M. 2005. "Meta Interpretation": A Method for the Interpretive Synthesis of Qualitative Research. *Forum: Qualitative Social Research* 6(1):Art.37.

Wollenberg, E., L. Merino, A. Agrawal, and E. Ostrom. 2007. Fourteen Years of Monitoring Community-Managed Forests: Learning from IFRI's Experience. *International Forestry Review* 9(2):670–684. http://dx.doi.org/10.1505/ifor.9.2.670.