IDA

INSTITUTE FOR DEFENSE ANALYSES

A Qualitative Data Collection Strategy for Africa

Ashley N. Bybee, Project Leader
Dominick E. Wright

INSTITUTE FOR DEFENSE ANALYSES

A Qualitative Data Collection Strategy for Africa

Ashley N. Bybee, Project Leader
Dominick E. Wright

Executive Summary

Sponsor and Objectives

In this study, sponsored by the Rapid Reaction Technology Office (RRTO) in the Office of the Deputy Assistant Secretary of Defense for Rapid Fielding (DASD/RF), for the Assistant Secretary of Defense for Research and Engineering (ASD/R&E), IDA examined qualitative data available for use in models, simulations, and other computational tools (MS&T) for analysis of Africa. Through the identification of gaps in qualitative data, IDA developed a Qualitative Data Collection Strategy (QDCS) to address the most significant qualitative data gaps pertaining to Africa. In doing so, we sought to:

- Improve the USG's qualitative data collection efforts by increasing availability and awareness of accurate and valid data available to analysts.

- Coordinate among the communities of interest to avoid duplication of data collection efforts.

- Ensure the most efficient allocation of resources to fill in gaps in qualitative data.

By most definitions, qualitative data comprise any non-numeric description of a person, place, thing, event, activity, or concept. A qualitative factor is one that typically represents structural assumptions that are not naturally quantified. This study combines these two key features to produce a definition that includes all descriptions of persons, places, things, events, activities, or concepts that are not numerical or not naturally numerical. This amendment recognizes that many quantified data are inherently qualitative in nature, requiring some subjective interpretation when coding into an ordered (or ordinal) scale. By this definition, IDA includes unstructured and entirely textual data (e.g., focus group data collected through open discussions or anthropological methods) as well as structured (coded) data (e.g., quantized public opinion data collected through various sampling methods used in polling and surveys). This definition includes what is commonly referred to as socio-cultural data (e.g., descriptions of ethnicity, culture, beliefs) but may also include other types of qualitative data such as geographic (e.g., qualitative designations of soil and terrain types along with geo-located socio-cultural data), humanitarian (e.g., reports describing wellbeing and needs), and health-related data (e.g., disease risk propensities for locales).

The long-term goal of this line of inquiry is to facilitate more accurate social science modeling, which this study contributes to by identifying existing qualitative data sources

that may be unknown to members of the MS&T and broader analytic communities, describing possible methodologies to address or fill identified gaps, and identifying synergies with existing efforts where collaboration can occur to support the development of a community standard.

Improving the performance of MS&T through the incorporation of better data inputs not only increases their value to consumers of these analytic products, but ultimately improves the ability of policy-makers to make well-informed decisions. There are high stakes involved when considering whether to enter a foreign country, either in a civilian capacity or a military one. Either way, the decision to do so might come down to insights and considerations offered using qualitative data. Improving the analyses that inform these decisions by using better data in MS&T translate into an enriched understanding of complex environments.

Findings and Recommendations

IDA's findings stem from two phases of preliminary research focused on a) the analysis of existing MS&T used by analysts of Africa, and b) engagement with African scholars and other Africa-based researchers to identify the types of data that have the greatest explanatory power in the African context. The associated recommendations fall into three broad categories, which reflect immediate process improvements, some near-term actions to address the most pressing data gaps, and a long-term plan to ensure a sustainable flow of needed data.

Immediate Process Improvements

Finding 1: There are a number of qualitative data sources that are available but unknown to many analysts.

IDA has compiled a list of the sources encountered during the course of this research, which can be incorporated into existing data portals. This list has been compared with the datasets contained in DataCards (a catalog of indexed information on available quantitative and qualitative datasets, as well as portals to general information) and the Cultural Knowledge Consortium (CKC, a Socio-cultural Knowledge Infrastructure (SKI) to facilitate access among multi-disciplinary, worldwide, social science knowledge holders that fosters collaborative engagement in support of socio-cultural analysis needs,) so that it avoids duplication with sources that have already been captured through those portals.

Recommendation 1: Disseminate list of qualitative data sources.

IDA recommends this list be provided to the DataCards Program Manager and disseminated as widely as possible among the community of interest.

Finding 2: There are several existing portals for socio-cultural data, two of which are DataCards and the CKC.

Of the two (described in finding 1), DataCards is currently best-suited to serving immediate data needs for analysts seeking information on topics and issues throughout Africa.

Recommendation 2: Raise awareness and increase use of DataCards.

Because DataCards is accessible to users outside of DoD (including academia, the intelligence communities, and even some international entities, with wide usage encouraged) it has the potential to become the central portal for all socio-cultural datasets. As such, IDA recommends raising awareness of the tool to attract more contributors.

Finding 3: Some data producers have holdings that are only partially observable by USG consumers.

Some of the data available through DataCards are comparatively less discoverable relative to other records because their entries contain less descriptive information. This impedes the discovery and use of valuable data that might suit specific needs.

Recommendation 3: Improve methods to connect stakeholders to rigorous collection centers.

Just as some private survey organizations have made their survey questions available to the public (excluding the raw data that is available for a fee), the USG can similarly develop a technological solution that maintains a searchable catalog of all data available to data consumers. Once identified, these data could then be obtained upon request.

Finding 4: There are some qualitative data gaps that will persist regardless of the resource levels invested to fill them.

Regardless of the time and resources spent collecting data, there will always be some data gaps that persist, either because they are too sensitive or they might be obsolete by the time the data are collected. Rather than excluding unknown variables or substituting them with potentially inaccurate data, analysts need to apply methodologies to control or adjust the model or simulation accordingly.

Recommendation 4: Survey the M&S community for "best practices" when imputing unknown data.

OSD should facilitate the collection of ideas and perspectives on how end-users account for or otherwise impute unknown parameter values into M&S and what

techniques they have determined to be "best" for their purposes. These findings should be captured in a "living document" and continually updated based on technological changes and the constant stream of new insights within the M&S community

Near-term Actions to Address the Most Pressing Data Gaps

Finding 5: There are a large number of qualitative data needs and only limited resources to fill them.

Given the limited resources available for collection amid a clear demand signal for specific data points, a prioritization of these data needs would assist the government in allocating the appropriate level of resources to the collection of the highest priority data needs. More fundamentally, IDA notes the absence of an officially vetted and approved list of socio-cultural data requirements. To date, there has not been a systematic approach to documenting what socio-cultural information is relevant for the military.

Recommendation 5: Establish qualitative data requirements and prioritize them. In the interim, prioritize qualitative data needs.

IDA recommends OSD lead the process for formally vetting, approving, and prioritizing qualitative data requirements for MS&T. In the interim, OSD could lead a practical prioritization of the qualitative data needed by DoD organizations tasked with analysis of Africa. Ascertaining the most frequently cited data gaps would help OSD to determine those data points that would have the maximum utility for all. Discussions should address geographic priorities, thematic priorities, and any other relevant characterization of data as identified by stakeholders.

Finding 6: There is potential for proven methodologies to yield previously unavailable qualitative data from Africa.

Various unconventional qualitative data collection techniques have been applied in other regions that have the potential to yield valuable data from Africa. It is worth testing these promising qualitative data collection methodologies in Africa, specifically those that have proven to be fruitful in other regions or contexts.

Recommendation 6a: Cultivate collaborative research networks for improved access to local data.

IDA recommends that DoD test promising qualitative data collection methodologies in Africa, specifically those that have proven to be fruitful in other regions or contexts. One such example that has been an effective approach in southeast Asia are collaborative research networks convening non-official partners, e.g., traditional authorities, youth groups, religious institutions, the private sector, academia, or NGOs, to discuss topics of

mutual concern. As a result of the research-based engagement in this region, the USG has augmented its knowledge base and now has a deeper understanding of the nuances and complexities that characterize the region.

Recommendation 6b: Engage diaspora communities residing in CONUS.

The African diaspora that resides in the U.S. is a largely untapped source of qualitative data. Not only are the diaspora a resource geographically convenient to U.S.-based researchers (negating the need to make costly data collection trips to Africa), but engaging them in the U.S. overcomes many of the challenges of bureaucracy and corruption often associated with data collection in Africa. Moreover, this is a valuable method to elicit critical data from population samples that may serve as useful proxies for otherwise inaccessible populations in Africa.

Finding 7: There are several new DoD initiatives and methodologies under way for collecting socio-cultural data in Africa.

Special Operations Command (SOCOM), the Defense Intelligence Information Enterprise (D2IE), the Joint Staff J-7, among others, are all investigating new methodologies to collect socio-cultural data. Each has its own, subjective need but alternate methods for collection should prove universally-applicable across a number of them.

Recommendation 7: Work with the interagency to support experimentation and deployment of new methodologies for socio-cultural data collection.

Where opportunities exist to collaborate with DoD and other interagency partners to experiment with the deployment of new methodologies for socio-cultural data collection, IDA recommends RRTO participate in these activities to facilitate the development of technologies to support these methodologies.

Finding 8: There is a need to facilitate personal contacts and raise awareness of qualitative data sources among the community of interest.

The use of one data portal, such as DataCards, will increase awareness and access of all data across the community. Knowledge of these data sources, however, will always be contingent on the degree of use of such a portal. Moreover, there will always be new data sources coming online that will not be captured in such a portal. A secondary mechanism to track and raise awareness of such data collection efforts would be beneficial to ensure maximum exposure across the community of analysts and other stakeholders.

Recommendation 8: Partner with NDU to hold regular conferences convening data collectors and owners of qualitative data.

Given the overlapping interests among data consumers and data producers, there is a potential synergy to be gained by convening these communities in a mutually beneficial forum. The regular Socio-Cultural Data Evaluation Summits (otherwise referred to as "Data Summit") organized by NDU are an ideal opportunity to assemble these different communities that have very similar interests. This forum would be a prime opportunity to convene data producers alongside data consumers to showcase collection efforts currently under way that will yield new data sources in the near future.

Long-term Plan to Ensure a Sustainable Flow of Needed Data

Finding 9: Local capacity for qualitative data collection is low.

In Africa, scarce resources are focused on high priority activities, with less vital activities (such as data collection) often left by the wayside. Moreover, the requisite skills to administer surveys, conduct interviews, and other qualitative methodologies are severely lacking. As a result, qualitative data collection is typically performed by external actors on an *ad hoc* basis to serve immediate data needs. Data are not collected in a sustainable fashion or using a consistent methodology at regular intervals over time, which contributes to the problem of poor time series data. So that the USG (and others) can leverage the data collected by Africans without continuing to invest massive resources indefinitely, it should consider building local capacity for this collection. Such investments could have a strategic payoff (access to data), while contributing to DoD's partnership capacity building mission. In areas in which DoD could benefit from more and improved qualitative data (socio-cultural data that could assist with counter-terrorism operations), it would be appropriate for OSD to support the growth of local capacity for qualitative data collection.

Recommendation 9: Increase technical training for local qualitative data collection, especially capacity to execute national censuses.

There are several avenues through which OSD can contribute to this endeavor. IDA recommends that DoD partner with local data collection organizations that have the capacity themselves to run technical training programs for local Africans. Independent, African-based survey firms such as Afrobarometer; academic institutes such as the Centre for Social Science Research at the University of Cape Town; or multinational organizations such as the UN Office on Drugs and Crime (UNODC), the UN Institute for Training and Research (UNITAR) or the UN's Economic and Social Council (ECOSOC) are among some reputable organizations and potential partners for this activity. Building upon the work and achievements of African data collection institutions in the region is

not only efficient, but it prevents the problem often encountered by Western institutions whose prescriptive approach is resented by Africans. IDA also recommends facilitating collaboration among such institutions to leverage each other's capabilities and share lessons learned.

Training should be focused in the areas of survey administration, including mobile phone surveys, and the administration of national censuses. The International Programs Center for Technical Assistance at the U.S. Census Bureau, has already worked with some African partners to improve census processes.

Contents

1. Introduction

In this study, sponsored by the Rapid Reaction Technology Office (RRTO) in the Office of the Deputy Assistant Secretary of Defense for Rapid Fielding (DASD/RF), for the Assistant Secretary of Defense for Research and Engineering (ASD/R&E), IDA examined qualitative data available for use in models, simulations, and other computational tools (MS&T) for analysis of Africa. Since discovery of gaps or instances where available data could not satisfy informational needs for MS&T was inevitable, RRTO asked IDA to propose a strategy for filling at least some of the gaps discovered during its survey.[1] OSD recognizes the value MS&T have to support decision-making and practice at the strategic, operational, and tactical levels, while acknowledging the limitations inherent in any social science modeling that depends heavily on indeterminate and variable human behavior. The value of MS&T, therefore, is contingent on the availability of high-quality data input to the system.[2] Use of MS&T applied toward the study and analysis of Africa is likely to prove unsatisfactory because of the comparatively low level of data from the continent relative to other regions. One concern of the U.S. Government (USG) is that many of the problem sets and possible contingencies that Africa presents involve more socio-cultural dynamics than physics-based ones. A paucity of relevant data therefore impairs the utility of MS&T to answer questions concerning current and future trends or events on the continent. High-quality, validated data used in computational modeling will greatly improve utility of MS&T analyses by providing a more accurate portrayal of the social, cultural, political, and even economic landscapes in various African regions.

Because this study focuses exclusively on the African continent, IDA sees the Africa Command (AFRICOM), its components, and other USG organizations with

---

[1] Many of the findings and recommendations in this paper refer to data *needs*, which are fundamentally different from data *requirements*. The former include information that is needed for the successful application of MS&T. These may include developer-stated needs (which they may refer to as requirements but are different from official DoD requirements.) The latter are formally vetted needs that have been approved by both technical and governance bodies. This is a critical differentiation and it is worth noting the absence of socio-cultural data *requirements* in DoD, despite a litany of data *needs*.

[2] There is some debate over the necessity of *high-quality* data to achieve the results that will suffice for DoD's needs, particularly in light of the difficulty and high cost in attaining such high-quality data. Many analysts argue that mediocre and even low-quality data may be sufficient to provide an acceptable result, i.e., within a certain confidence level or within acceptable parameters given the inherent accuracy of the program itself. It is therefore important to note that data needs are subjective, and the highest quality data are not always necessary or even appropriate for *all* MS&T, but the findings and recommendations in this report assume that, given the subtle nuances contained within much qualitative data, the use of high-quality qualitative data in MS&T will significantly improve the performance of those programs.

activities in Africa as primary beneficiaries of this research. This paper presents a Qualitative Data Collection Strategy (QDCS) for Africa based on IDA's previous findings and subsequent analysis of the most significant qualitative data gaps pertaining to Africa.[3]

A. Background

Data collection in Africa is a painstaking and expensive process, but there appears to be universal appreciation for its importance and for maintaining the integrity of the process. There are many reasons for the dearth of both qualitative and quantitative data on Africa and much of the developing world. The capacity of the public sector to conduct data collection in Africa is noticeably less robust than in much of the developed world. This is particularly true when comparing data collected from different strata or levels of society. Whereas developed countries routinely collect data at the national, state/provincial, city, neighborhood, household, and individual levels, developing countries (including most in Africa) collect relatively lower volumes of data at each of these levels. National-level data are perhaps the best represented, with decreasing numbers of data available as one progresses toward higher levels of fidelity, i.e., the individual. Moreover, the limited resources that any government is willing to invest in such a massive, complex continent have also precluded extensive data collection. Since there is very limited data collection by local (African) research organizations, most academic and policy-related research relies heavily on qualitative data produced by Western-educated social scientists, anthropologists, and political scientists. These researchers tend to lack an African perspective on their research agenda and data collection, which can militate against cultural bias inherent in a Western approach. As a result, it is difficult to gauge the relevance of many qualitative data or to know whether the data have sufficient explanatory power in the African context. Even where the data exist, they are often not captured in written texts or are available only in outdated colonial texts. Rather most relevant data are held within the minds of African people where the only means of access is through oral discussions, which is a very time consuming and costly to collect.

Moreover, the sensitive nature of many "taboo" research areas such as sexual behavior, religious practices, lifestyle habits, and drug consumption, as well as important security issues such as illicit trafficking (e.g., small arms/light weapons, weapons of mass destruction, drugs, humans), and the nexus of each with terrorism have precluded meaningful discussion in these areas and the elicitation of native perceptions of these

---

[3] IDA's findings are documented in: Ashley Bybee and Dominick Wright, IDA Document D-4629, *Designing a Qualitative Data Collection Strategy (QDCS) for Africa – Phase I: A Gap Analysis of Existing Models, Simulations, and Tools Relating to Africa*, June 2012, and an informal report delivered to the sponsor titled "Phase II: Qualitative Data Gaps from the African Perspective," August 29, 2012.

issues. As a result, these qualitative data points are currently, and understandably, absent for much of Africa.

From a data user's perspective, there are additional concerns relating to the supply, format, and quality of qualitative data that affect their ability to be used in MS&T. A common concern voiced by researchers is the difficulty in identifying subject matter experts (SMEs) who can provide reliable, high quality data. Another issue is consistent access, which many data users lack. For example, they may receive one or two qualitative data sources that represent valuable "snapshots" at a given point in time, but they do not have access to those same data over time. This absence of sufficient "time series" data makes it difficult or even impossible to understand overarching trends that are so critical for researchers to determine causal relationships and associations with other examined variables. Moreover, without some indigenous insight into local geo-political and environmental conditions, it is inordinately more difficult for outsiders to know how relevant and valid data are over time, or how applicable the data might be across a span of operational contexts. Finally, although there is a DoD data standard within the context of M&S (DoD Directive 5000.59, 2007), it contains no specification for socio-cultural data and adherence to its prescriptions is variable. Combine this with the fact that data used for alternate purposes have alternate standards as well, and the problem of regularity in critical data features, such as format, becomes clear. As a result, qualitative (and quantitative) data lack consistency in critical features, such as format and unit of measurement, which makes it difficult to achieve seamless inputs into MS&T.

To address these challenges (among numerous others) is not just important on academic or philosophical grounds. There are actual implications for all of the MS&T IDA encountered during its initial survey. (Appendix A provides a full list and description of these MS&T.)[4] The following are just a few of the MS&T used for the analysis of the African continent whose designers have stated would benefit from some improved qualitative data:

- Competitive Influence Game (U.S. Army)
- Cultural Geography (U.S. Training and Doctrine Command)
- Geospatial Information Awareness/Infection Disease (GIA/ID) (Naval Research Laboratory)
- HOA-Viewer (Department of State's Humanitarian Information Unit)
- Composite Vulnerability Map (University of Texas)

---

[4] For the associated gap analysis, see Ashley Bybee and Dominick Wright, IDA Document D-4629, Designing a Qualitative Data Collection Strategy (QDCS) for Africa – Phase I: A Gap Analysis of Existing Models, Simulations, and Tools Relating to Africa, June 2012.

- RiftLand (George Mason University's Center for Social Complexity).

B. Objective

The immediate objective of this study is to design a QDCS that will:

- Improve the USG's qualitative data collection efforts by increasing availability and awareness of accurate and valid data available to analysts.

- Coordinate among the communities of interest to avoid duplication of data collection efforts.

- Ensure the most efficient allocation of resources to fill in gaps in qualitative data.

Achieving these objectives will unquestionably improve the value of MS&T to support USG decision-making by presenting a more accurate socio-cultural landscape to inform policy-makers. One must recognize, however, the limitations inherent in any social science modeling that depends heavily on indeterminate and variable human behavior. Because it is highly unlikely that qualitative MS&T will ever perform accurately enough to reliably and consistently predict reality, users must treat them as one tool within a larger toolkit. At this point, therefore, MS&T should be used to motivate thought and discussions, rather than serving as a prediction or forecasting tool. The MS&T IDA surveyed are all intended to characterize complex socio-cultural-economic-political dynamics rather than identify the result of a given scenario. In doing so, they reveal possible interactions, inform the decision-making process, and provide a point of departure for further discussion, research, and analysis. Moreover, unlike hard scientific models that will generate a clear answer, social science models require some expertise for interpretation.[5] These are all critical functions that warrant continued expenditures in and improvements to qualitative MS&T.

The long-term goal of this line of inquiry is therefore to facilitate more accurate social science modeling, which this study contributes to by identifying existing qualitative data sources that might be unknown to members of the community, describing possible methodologies to address or fill identified gaps, and identifying synergies with existing efforts where collaboration can occur to support the development of a community standard.

C. Study Approach and Methodology

IDA approached this project in three distinct phases:

---

[5] Lisa Costa, "Sudan Strategic Assessment: Understanding the Dynamics of Complex Socio-Cultural Environments" October 26, 2007, and Joshua Busby and Jennifer Hazen, "Mapping and Modeling Climate Security Vulnerability: Workshop Report," Robert Strauss Center for International Security and Law, October 2011.

The first phase entailed a survey of the existing MS&T used to analyze the African continent to develop a keen sense of what capabilities are most used and most desired by the community of Africa analysts. The methodology employed for this first phase of research included an extensive literature review of known reports and articles on the subject of M&S in an effort to identify those applicable to Africa and suitable for further study. IDA reviewed DoD's M&S Catalog[6] and, with guidance from OSD, contacted several individuals leading projects relevant to the study. IDA also surveyed M&S currently used by USG organizations and some in development in academia. Additional recommendations and points of contacts were provided, which enlarged IDA's pool of interviewees to a sufficient sample size. For each of the projects, IDA interviewed the owners/administrators of the MS&T and sought the following information: types/sources of qualitative data used, data collection/validation methodologies, assessment of data quality, format of data, challenges to collection/analysis, and gaps in qualitative data. During this phase of data collection, IDA observed the execution of the U.S. Army's Asymmetric Warfare Group's Competitive Influence Game (CIG) in Vicenza, Italy, in January 2012. During this three-day simulation, which focused on violent extremist organizations, radicalization, and piracy in the Horn of Africa, IDA conducted interviews with software designers as well as the SMEs to understand how this particular simulation used qualitative data and how it addressed gaps in those data.

The second phase of the task involved engagement with African scholars and other Africa-based researchers with whom IDA's team of Africa experts have existing academic contacts.[7] IDA views this step as a critical feature of a data collection strategy, because it takes into account indigenous insights into African issues. While U.S.-based researchers and designers of M&S have clear data needs for their systems, such data might not have the most explanatory power in the African context. As a result, researchers might find that they are analyzing data that do not reveal new insights or shed new light on emerging trends. Soliciting input from Africans on what they perceive to be the most salient information to capture to explain certain phenomena, while identifying emerging trends that might not be on Americans' radar, represents a strategic investment with immediate and long-term returns. Moreover, including Africans as active participants in this phase of strategizing will better position the USG and research institutions to engage Africans on issues of mutual concern in the future and cultivate long-term partnerships that yield new data (including real-time data) for both the U.S. and its African partners.

---

[6] https://mscatalog.osd.mil/intro/index.aspx

[7] The findings from this phase of research were delivered to the sponsor in an informal report titled "Qualitative Data Gaps from the African Perspective" and are available upon request.

The research and analyses conducted in phases 1 and 2 culminated in the development of QDCS, which is presented in this paper. This strategy takes into consideration the needs of MS&T pertaining to Africa while ensuring that data needs are attuned to the interests of African partners.

D. Scope

1. Models, Simulations, and Tools (MS&T)

The computer applications included for analysis in this study consisted of models, simulations, and some relevant tools. Models are "physical, mathematical, or otherwise logical representations of a system, entity, phenomenon, or process."[8] They are simplified representations of a system for which designers have implicitly or explicitly specified the conditions (e.g., time and space) under which it might appropriately be used to understand a "real" (i.e., empirically observable) system. Simulations are "methods for implementing models over time."[9] Whereas application of a model might provide an answer for a specific set of temporal and spatial conditions, a simulation extends these results over a time or space continuum. Adjunct tools (hereafter referred to as tools) are "software and/or hardware used to provide part of a simulation environment or to transform and manage data used by or produced by a simulation environment."[10] They differ from models and simulations in that they are not, and do not mean to be, logical representations of systems. Instead, they are used to manage, store, and represent information produced from models and simulations along with other origins (including other tools). The main MS&T targeted by IDA were those currently being used by the USG for the analysis of Africa. Because the actual number of MS&T used by the USG was not as high as expected, IDA broadened the scope of the study to include some MS&T used in academia, since their qualitative data gaps are also helpful data points.

2. Qualitative Data

Qualitative data comprise any "non-numeric description of a person, place, thing, event, activity, or concept."[11] A qualitative factor is one "that typically represents structural assumptions that are not naturally quantified."[12] This study combines these two key features to produce a definition that includes all descriptions of persons, places, things, events, activities, or concepts that are not numerical or not naturally numerical. This amendment recognizes that many quantified data are inherently qualitative in nature,

---

[8] M&S Glossary, Retrieved on October 8, 2012. Available at: http://www.msco.mil/MSGlossary.html.
[9] Ibid.
[10] Ibid.
[11] Ibid.
[12] Ibid.

requiring some subjective interpretation when coding into an ordered (or ordinal) scale. By this definition, IDA includes unstructured and entirely textual data (e.g., focus group data collected through open discussions or anthropological methods) as well as structured (coded) data (e.g., public opinion data collected through quantitative methods such as polling of respondents selected by multi-stage, random and stratified sampling). Conversely, quantitative data are inherently "numerical expressions that use numbers, upon which mathematical operations can be performed."[13] In terms of collection methodologies, qualitative data are typically collected directly (e.g., field research and observation studies, focus group discussions, surveys) while naturally occurring quantitative data are usually collected indirectly via instruments (e.g., census survey instruments producing counts of households in housing tracts as well as imagery data producing satellites).

Consider, for example, a consumer who must decide between two brands of a product. Quantitative data distinguishing between the two might include price and quantitative descriptions of their alternate compositions (e.g., chemical make-up, mechanical configuration). Price data offer a straightforward way to compare the two brands. What is difficult to determine is the degree to which a consumer prefers one brand to the other. Even more difficult to determine is the reason that preference exists. Does the consumer prefer Brand A to Brand B because it is cheaper (easily quantified) or somehow "better" (difficult to quantify even when the degree of preference rests on a survey instrument, such as a Likert scale).[14] Examples such as this one illustrate that, by comparison, quantitative data are conceptually easier to grasp and measure. For these same reasons, quantitative data are also easier to collect, irrespective of collection conditions. As a result, the majority of data usable in MS&T are mostly quantitative. There is relatively less qualitative information available to characterize elements of systems that are difficult to represent but no less important to understand.

This is especially true in the developing world, where fewer resources and assets are available for data collection.[15] Nonetheless, the security challenges associated with the post-9/11 geo-political environment have only increased calls for more African data. IDA's casual observations suggest that a number of data needs are associated with the perceived plights of Africans, which over time could translate into major security

---

[13] Ibid.

[14] The Likert scale is the most widely used approach to survey research where responses are chosen among a ranking of multiple categories

[15] Baisch Jürgen, "Data Shortage in Africa." 2008. Retrieved on October 8, 2012. Available at: http://www.water-for-africa.org/tl_files/content/download_public/IWFA-Data_Shortage_in_Africa.pdf. and Eileen Hoal, "Famine in the Presence of the Genomic Data Feast," Science. February 18, 2011. Vol, 331 (6019): 874.

concerns where basic human needs are not met.[16] This appears to be the impetus for many climate and water data collection efforts.[17] Analyses of these data are focused on how climate change and water levels affect migratory flows that could in turn indicate potential human insecurity or outright conflict. Efforts to collect quantitative data (such as water levels, GPS coordinates, and even the number of news events associated with these observations) are all critically important. These data can answer questions such as:

- Who is affected by climate change?

- What is the impact of decreasing water levels on a given population?

- Where are populations moving in order to access more water?"

Most analysts agree, however, that only qualitative data can answer questions such as: "Why are populations choosing to move to certain locations, over others?" and "how are they choosing to move?"[18] Without qualitative data, analysts run the risk of over-assigning importance to quantitative data simply because they are available. The focus for IDA, therefore, is to develop a strategy for complementing the quantitative data that answer the "who?" "what?" and "where?" questions with qualitative data that can relate the "why?" and "how?" also associated with these topics.

E.   Document Outline

This document is organized as follows: After this Introduction chapter, Chapter 2 presents the conceptual framework that was used to identify findings and assess recommendations. Chapters 3, 4, and 5 present IDA's findings and recommendations and describe some possible initiatives that might be adopted to fill or address the qualitative data gaps identified in IDA's initial report. These recommendations and initiatives fall into three categories that reflect immediate process improvements (Chapter 3), some near-term solutions for the most pressing data gaps (Chapter 4), and a long-term plan to ensure a sustainable stream of needed data (Chapter 5).

---

[16] The "January 2012: Special Issue on Climate Change and Conflict," published by the Journal of Peace Research is but one example of the emphasis on physical data. The issue is available at: http://jpr.sagepub.com/cgi/collection/special_issue_on_climate_change_and_conflict.

[17] Ochieng' Ogodo, "Africa Facing Climate Data Shortage." November 11, 2009. Retrieved on October 8, 2012. Available at: http://www.scidev.net/en/news/africa-facing-climate-data-shortage.html. See also the Institute Water for Africa (IWAF) website at: http://www.water-for-africa.org/en/home_articles/articles/africa-isnt-only-suffering-from-water-shortage.html.

[18] Answering the "why?" and "how?" questions are qualitative in that they are assessments made by individuals and asserted in the form of attitudes. The fact that the data are qualitative does not preclude the use of quantitative methods for analyzing it. For example, a representative sample of attitudes describing why a population would choose to move to a certain location over others, coupled with the qualitative attributes of respondents would lend itself to a variety of statistical methods used to understand the correlation between individual characteristics and predilections for moving.

2. Conceptual Framework

A. The MS&T Marketplace

The findings and recommendations contained in this strategy are presented in terms of a conceptual framework that describes an "MS&T marketplace." DoD data (qualitative and quantitative) and the MS&T Marketplace (hereafter referenced as the *marketplace*) contain both a supply of data and MS&T as well as a demand for these two items. Since the focus of this paper is on data, specifically qualitative data, the conceptual framework here treats supply of MS&T as outside current scope and therefore determined externally.

This conceptual framework is useful for establishing the difference between the findings (the status quo) and recommendations (to achieve an ideal end state). As outlined, the conceptual framework will serve as an instrument for assessing data-related aspects of the DoD data and MS&T marketplace with respect to analyses of African topics. Although this study is explicitly focused on MS&T used for the analysis of Africa, the marketplace is not unique to Africa. It is region-neutral, which facilitates the ability of analysts to apply findings and recommendations within a context of identified gaps in available products, functions, and so forth. Situating findings and recommendations within alternate aspects of the marketplace will facilitate decision-making capabilities regarding where and how to proceed in the effort to make it a robust center for exchange and creation of high-quality, analytic products.

B. Actors and Elements

The DoD marketplace for data and MS&T includes actors performing any one of four, non-mutually exclusive roles:

- Data producer: Data producers are those who generate factual information (i.e., empirical observations) through various means of collection and recording.

- MS&T producer: MS&T producers are individuals or organizations who develop, both conceptually and technically, the models, simulations, or tools that are used for analysis of a specified issue.

- Data consumer: Data consumers are the analysts who require qualitative and quantitative data in order to carry out their daily duties. They may use data for an array of purposes, such as conducting assessments or informing strategic planning.

- MS&T consumer: MS&T consumers are also analysts who use these automated programs to assist in their analyses.

The marketplace also includes the products generated by these actors, most notably Data (D), Models (M), Simulations (S), and Tools (T).

C.   The Data-MS&T Continuum

Within the marketplace, there exists a continuum that can be used to describe the way in which consumers use data and MS&T. Depending on the duties and assignments of a given analyst, he or she will make use of these elements in different ways. The decision is made once an analyst is presented with a task that requires an answer or "output" that can be derived from some use or manipulation of data with MS&T. Alternate pairings of data with MS&T create a continuum of use describing the combination of means used to complete an assigned task. Table 1 describes how data and MS&T might be used along this continuum.

Table 1. The MS&T Continuum

Marketplace Elements	Generic Task Description	Example Question
D	References raw data only.	How many ethnic groups and people identifying with each are present in region X of country Y?
D+T	Manipulating data in a form of organization (e.g., charts) without extracting additional meaning.	What is the geographical distribution of people identifying with alternate ethnicities in region X of country Y?
D+MS	Manipulating data into a secondary output by applying logic to extract additional meaning.	What are the rates of change associated with distributions of people identifying with alternate ethnicities in region X of country Y from period 1 – period 3? What is the forecasted distribution of period 5?
D+MST	Manipulating data into a secondary output by applying logic and organizing the extracted meaning (e.g., charts).	What is the geographical distribution of observed and forecasted transition rates for people identifying with alternate ethnicities in region X of country Y?

Data (D), Tools (T), Models (M), and Simulations (S)

The use of these elements requires consumers to be familiar with the availability of each or know where to look for the needed information. Knowledge portals and repositories come in numerous forms but universally share one quality: they do not capture the entire supply of available data. Data supply is constantly increasing as a result new needs and injections of resources to collect those data priorities. Knowledge portals and repositories capture what is available and known to their administrators. To the extent that data producers generate products that are available yet unknown, there will always be a "gap" to close by making them available.

D. Data Attributes

Across the Data-MS&T continuum, the need for certain data attributes will remain constant:

- Phenomenon: the social (human-driven) and/or physical (nature-driven) topics of interest such as:

    – Water availability

    – Ethnic group maps

    – Occurrence of violent conflict

    – Patterns of electoral support

    – Public goods provision or governance

- Time: the period of interest

- Area: the geographical area or areas of interest

- Format: requirements describing form of the data such as:

    – Structured or unstructured

    – Quantified or textual.

These four attributes are not an exhaustive list of data characteristics, but they do represent a high-level abstraction at which all data are comparable. In an ideal marketplace, a consumer would be able to articulate a data need along these dimensions, the supply of data would contain one or more products satisfying that need, and a mechanism for exchange would quickly identify and link the consumer with the appropriate set of products. Thus the critical elements of the ideal marketplace would include a supply of data products that satisfies all data needs (or data demand) and a mechanism for exchange that is fully aware of all available data products.

E. Application to Findings and Recommendations

Where appropriate, the findings and recommendations presented in the next chapter are characterized in terms consistent with the MS&T marketplace described above. In addition to outlining the time horizon required until a benefit is realized and the expected cost, the QDCS also describes where within the marketplace IDA has identified a shortcoming and how the recommendation addresses it.

3. Immediate Process Improvements

OSD and interested stakeholders can adopt several simple process improvements to enhance their access to qualitative data on Africa. The following findings and associated recommendations do not require additional resources to be expended by the USG and could be adopted immediately.

Finding 1: There are a number of qualitative data sources that are available but unknown to many analysts.

The process of interviewing individuals and organizations has revealed a number of qualitative data sources and collection efforts currently under way that might not be well known among the community of analysts and other stakeholders. Many contain survey data while others present findings of anthropological research in certain specific communities that help analysts to understand unfolding developments in a given country. For example, the "Master Narratives" produced by the Open Source Center present historically grounded stories that reflect a community's identity and experiences, and explain their hopes, aspirations, and concerns. Vital data sources like these represent valuable additions to the pool of available and readily discoverable data.

Appendix B (an Excel spreadsheet) lists all such qualitative data sources encountered by IDA, including several descriptive details about the datasets themselves. This list has been compared with the datasets contained in DataCards and the data provided by the Cultural Knowledge Consortium (CKC) to avoid duplication with data sources that have already been captured through those portals. The spreadsheet is formatted in such a way and contains the appropriate fields to facilitate entry into the DataCards portal. (See Finding 2.)

Recommendation 1: Disseminate List of Qualitative Data Sources.

IDA recommends this list be provided to the DataCards Program Manager and disseminated as widely as possible among the community of interest, so all parties may benefit from datasets they might otherwise not know of.

Finding 2: There are several existing portals for socio-cultural data, two of which are DataCards and the Cultural Knowledge Consortium. Of the two, DataCards is currently best-suited to serving immediate data needs for analysts seeking information on topics and issues throughout Africa.

The Center for Technology and National Security Policy (CTNSP) at National Defense University (NDU) is currently coordinating a project called "DataCards," which

serves as a repository of information on available quantitative and qualitative datasets, as well as portals to general information.[19] According to the DataCards website:

> DataCards is a structured collection tool that indexes data sources that relate to irregular warfare, assessment, or can be used for socio-cultural modeling... These cards provide a summary description and evaluation of the content, quality, intended purposes, and potentially appropriate uses of each source.[20]

Although the program initially focused on Afghanistan, it subsequently widened its collection strategy to include information from across the globe with a large concentration effort geared toward Africa. As a result, the DataCards website (www.datacards.org) now contains more than 367 cards for Africa (approximately 22 percent of the 1,655 currently available). Each data card consists of a profile describing the data and links to the original source by listing a URL or a POC. An ongoing effort pursued by the DataCards team includes appending each record with the actual data. Currently, it is unclear how many records possess these attachments, but it is reasonable to think a plurality of them will be of the quantitative and quantified sort.

Datacards is somewhat unique in that it encourages users from beyond DoD, e.g., academia, the intelligence communities, and even some international users. Since DoD is not the obvious source of socio-cultural data for non-DoD researchers (and probably never will be), there is wisdom in granting access of unclassified data to other government agencies, the NGO community, academia, and even foreign users. This encourages reciprocation and increases the likelihood these communities will provide additional socio-cultural data for DoD's benefit.

The CKC, administered by the U.S. Army Training and Doctrine Command (TRADOC) Analysis Center (TRAC) at Fort Leavenworth, Kansas, is another DoD effort to serve the socio-cultural needs of the combatant commands. According to the CKC website:

> The Cultural Knowledge Consortium (CKC) provides a Socio-cultural Knowledge Infrastructure (SKI) to facilitate access among multi-disciplinary, worldwide, social science knowledge holders that fosters collaborative engagement in support of socio-cultural analysis

---

[19] Several findings and recommendations in this report refer to a *portal*, which is fundamentally different from a *repository*. A repository refers to the actual source of the data, while a portal (also known as a catalog or brokering system) refers to a list of references or a system for easily accessing remote data provided by the data owner. The latter, i.e., a portal with links to datasets, is preferable as it allows data owners to update their data as appropriate while the portal will automatically capture the most recent revisions without the expense of maintaining yet another database. Datacards is currently a portal for socio-cultural data, though it is seeking to acquire raw data, thus positioning itself as a repository as well.

[20] https://www.datacards.org/, Accessed 20120927.

requirements. The CKC supports U.S. government and military decision-makers, while supporting collaboration and knowledge sharing throughout the socio-cultural community.[21]

As the CKC mission statement clearly articulates, the general goal of the effort is to enhance awareness and by extension usage of socio-cultural data throughout the DoD community. In spirit, it does not differ significantly from the DataCards effort. In form, however, the two differ considerably.[22] CKC operates more as a portal, where interested parties can connect with CKC-vetted SMEs (i.e., regional and functional scholars), read blogs, gain awareness of upcoming events, and so forth. A capability the organization aspires to offer in the future includes access to a sanitized version of the Distributed Common Ground System - Army (DCGS-A),[23] which will extend its unclassified offerings to include discoverable, searchable, and exploitable databases.[24]

DataCards, on the other hand, is a meta-database (an alternate term for a catalog serving as a database of databases). For example, it does not currently purport to include listings of SMEs, as CKC does. Once DCGS-A capabilities become part of the CKC UNCLASSIFED holdings, there will be some overlap between its offerings and those of DataCards. Until that time, IDA views DataCards as the better option within the UNCLASSIFIED arena for servicing the preliminary qualitative data needs of analysts focused on topics throughout Africa.

Recommendation 2: Raise awareness and Increase Use of DataCards.

Because DataCards is accessible to users outside of DoD (including academia, the intelligence communities, and even some international entities, with wide usage encouraged) it has the potential to become the central portal for all socio-cultural datasets, IDA recommends raising awareness of the tool to attract more contributors. IDA can do its part by formatting the list of data sources that it has compiled throughout this project for easy entry into DataCards and advertise those additions to all those interviewed. Incorporating IDA's list of qualitative data sources with this existing portal will ensure maximum benefit of OSD investments.

---

[21] https://culturalknowledge.org/, Accessed 20120927.

[22] Comparison of the two information portals will focus on their offerings presented in the unclassified arena, which IDA recognizes as only a partial representation of the overall capabilities possessed.

[23] https://secureweb2.hqda.pentagon.mil/VDAS_ArmyPostureStatement/2011/information_papers/PostedDocument.asp?id=151 Accessed on 20120927.

[24] https://www.culturalknowledge.org/data-brokering.aspx Accessed on 20120927.

Finding 3: Some data producers have holdings that are only partially observable by USG consumers.

Some data providers have records in DataCards, but their data are comparatively less discoverable relative to other records because their entries contain less descriptive information. Currently, this is not an efficient system because these records are not searchable by, for example, country, time period, or subject matter.

In interviews, IDA learned of several data collection centers that tend to serve a narrow consumer base. These data purveyors do not deliberately limit their distribution, but due to the absence of a system in which their data may be discovered and searched by the broader community, its existence is unknown to many. For example, the State Department's Office of Opinion Research (OOR) collects opinion data worldwide and has three analysts dedicated to sub-Saharan Africa. OOR's staff comprises methodologists trained in survey methods and statistical analysis, who have some regional expertise. OOR has been collecting data as requested and providing them to the Strategic Communication Division at AFRICOM, yet there is currently no mechanism for other offices and directorates within the Command to search or discover these data.

IDA learned that some data collectors are at times reluctant to make their data available publicly out of concern the data might be misused. They prefer for those interested data consumers to contact them directly with a request for data on a certain topic. Through discussions with potential consumers, OOR can determine which survey questions might be of interest, and then devote resources toward conducting analyses that service the expressed interest. This process hinges on questions survey collectors have at their disposal and internal resources available to conduct associated analyses.

Recommendation 3: Improve methods to connect stakeholders to rigorous collection centers.

Unlike Gallup, Pew, and other private industry survey firms, OOR and similar organizations do not treat their questions as proprietary and part of a business model. Nonetheless, IDA suggests that (similar to these organizations) it is possible to develop a technological solution that maintains a searchable catalog of survey questions, countries surveyed, the appropriate level of analysis (e.g., country, county or state, city), and the period covered.[25] Such a solution satisfies the survey collector concern over misuse while making the community more fully aware of polling questions available for analysis.

---

[25] Survey collectors can determine whether they want to post aggregate, descriptive statistics for others in the community to review.

Finding 4: There are some qualitative data gaps that will persist regardless of the resource levels invested to fill them.

One fundamental truth that must be acknowledged when analyzing African phenomena is that regardless of the time and resources spent collecting data, there will always be some data gaps that remain. These data may be unavailable because they are too sensitive and therefore difficult to collect by "outsiders" such as Western researchers. More likely, the data could be obsolete by the time it is collected. In these cases, it might be counterproductive to use one's "best guess" or "next best thing," lest the data be inaccurate and possibly affect the results of the model or simulation.

Rather than excluding unknown variables or substituting them with poor quality or potentially inaccurate data, analysts need to acknowledge that there will be data gaps that remain and apply methodologies to control or adjust the model or simulation accordingly. This sometimes involves making educated guesses where data are unavailable, although there are several techniques that M&S designers and users can use to do this. For example:

- Establish the initial conditions (e.g., population distributions, inter-group relations, socio-economic indicators) used as empirical anchors for synthetic populations.[26] This technique still involves making assumptions regarding the relationships between variables, but at least their starting values have empirical origins.

- Use proxy data elicited from SMEs combined with crowdsourcing techniques to derive values for empirically unavailable quantities of interest, such as the likely response of groups to kinetic and non-kinetic courses of action.[27] This technique amounts to relying upon experts to provide proxy data describing everything from "initial conditions" values to relationships between variables (e.g., conditional behavioral response, marginal elasticities – or regression slopes).

- Apply multiple runs, i.e. "Monte Carlo sampling" to determine how sensitive results are to changes in the unknown variable.[28] Rather than impute values for missing data, this technique requires analysts to conduct multiple runs of a model or simulation. This method enables researchers to determine the robustness of

---

[26] This is the technique used by the U.S. Army's Asymmetric Warfare Group (AWG) in cycle 6 of the Competitive Influence Game (CIG), which IDA observed.

[27] This is the technique used in Irregular Warfare simulations by TRADOC-TRAC and the Marine Corps Combat Development Command (MCDC) in conjunction with the Cost Assessment and Program Evaluation (CAPE).

[28] This technique may also be used to address the debate over the necessity of high-quality data by determining the quality of data needed to produce a result of acceptable credibility. For example, if a range of input values produce the same output value, expending resources to pinpoint the exact input value does not provide the return one would expect for that investment.

analytic estimates conditioned on alternate values of the missing data.[29] One output from these techniques includes an assessment of analytic result stability along with the estimated importance of the missing data. For example, if estimated results are relatively consistent across large subsets of missing variable values, then there is support for the inference that the missing variable is not substantively important. Alternatively, if estimated results depend heavily on certain values of the missing variable, then it is substantively significant and the situation warrants some additional effort toward collecting empirical information. Researchers can systematically apply these methods for all missing variables, interactions between missing and known variables, as well as interactions between jointly missing variables; however, doing so involves increasing levels of analytic complexity.

- Establish a method for monitoring the quality of data input into the system and assessing levels of confidence associated with the aggregate outputs. Such a method would likely begin with incorporating values for measurement error associated with variables possessing values, an attribute potentially extracted from ratings DataCards plans to incorporate within its holdings. This system would allow researchers to identify what data are of the highest quality and therefore apply datasets appropriately to the analysis at hand.

Recommendation 4: Survey the M&S Community for "Best Practices" when Imputing Unknown Data

IDA recommends a broad USG M&S community survey (regardless of regional application) to learn:

- How end-users account for or otherwise impute unknown parameter values

- Which types of M&S they are using and how data imputation of various types combines with variables having known values in their analyses

---

[29] The technical process for Monte Carlo sampling requires researchers to take the following steps: a) establish boundaries for a missing data point, which should reflect some empirically determined values to ensure results are reasonable and realistic; this boundary-setting is synonymous with establishing "initial conditions" values, as described previously; b) select the appropriate mathematical distributions (e.g., normal, Poisson, beta, and so forth) characterizing relative density of values throughout the population; this is the "distribution space"; c) establish the "parameter space," which is the range of values associated with each distribution (e.g., the mean and standard deviation of a normal distribution or the combined mean and variance shaping a Poisson distribution). Treating the boundary values as fixed or constant, while iteratively and systematically sampling distribution and parameter spaces provides input values for use in constructive and statistical models and a means for researchers to apply appropriate methods (e.g., topology) for reviewing results according to "pooling" and "separating" tendencies. Further analysis of the estimated results contributes toward assessing the substantive importance and impact of the variable in question.

- What techniques, such as those described above, they have determined to be "best" for their purposes.

Upon completion of such a survey, IDA recommends capturing findings in a document detailing the best practices currently used by the USG for imputing unknown data into M&S. This would enable OSD to normalize currently disparate data imputation efforts while improving the quality of analysis throughout the M&S community. Given rapid technological changes and the constant stream of exceptional insights within the M&S community, IDA recommends this be an ongoing process to ensure the USG is always aware of new techniques to address data gaps.

Once the USG's MS&T community has a sound understanding of these methods, it might consider expanding this survey to include techniques employed by academia and private industry. Because they have large financial interests in potentially volatile regions such as Africa, many private sector companies use models that help them to assess stability and other local dynamics. For example, oil and gas companies need to identify risks that could potentially affect their planned or ongoing operations in resource-rich host countries. Their assessments are based entirely on real-world scenarios so one must presume they have developed techniques to overcome or address gaps in qualitative data. Similarly, academics have likely developed their own techniques for overcoming these issues. By leveraging the insights of these two research communities, the USG would be better positioned to refine and improve its own best practices.

4. Near-Term "Surge" in the Collection of the Most Critical Data

With modest funding, OSD can initiate a near-term "surge" in the collection of a targeted set of frequently cited or "low-hanging" data gaps. A prioritization of these data needs needs to occur first, after which OSD can test one or more of the methodologies described below to collect the needed data.

Finding 5: There are a large number of qualitative data needs and only limited resources to fill them.

Throughout this process, IDA found many aspects of MS&T that could benefit from the infusion of new or improved qualitative data. These gaps are documented in the first phase of the research.[30] Many gaps are not unique to the specific MS&T, i.e., they represent similar data needs among many within the community of modelers and Africanists.[31] Given the limited resources available for collection amid a clear demand signal for specific data points, a prioritization of these data needs would assist the government in allocating the appropriate level of resources to the collection of the highest priority data needs.

More fundamentally, IDA notes the absence of an officially-vetted and approved list of socio-cultural data requirements. To date, there has not been a systematic approach to documenting what socio-cultural information is relevant for the military. Although the DoD community has been operating with unofficial data needs (versus official data requirements), establishing a list of formal socio-cultural data requirements would be a logical first step in the prioritization process. The figure below depicts the hierarchy of general data needs, software requirements, and the specific DoD-vetted data requirements for M&S.

---

[30] Ashley Bybee and Dominick Wright, IDA Document D-4629, Designing a Qualitative Data Collection Strategy (QDCS) for Africa – Phase I: A Gap Analysis of Existing Models, Simulations, and Tools Relating to Africa, June 2012.

[31] Formal discussions with AFRICOM were not held to ascertain the Command's most pressing data gaps; however, IDA did infer through informal discussions with other analysts and researchers, that the Command's research priorities are highly diverse in total yet extremely narrow in focus and therefor require an array of very specific data.
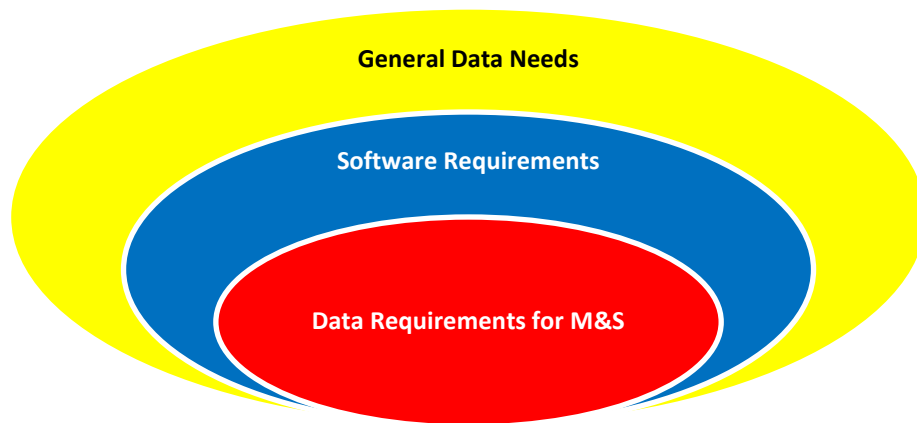
Hierarchy of Data Needs and Requirements

Recommendation 5: Establish qualitative data requirements and prioritize them. In the interim, prioritize qualitative data needs.

IDA recommends OSD lead the process for formally vetting, approving, and prioritizing qualitative data requirements for MS&T. In the interim, OSD may lead a practical prioritization of the qualitative data needed by DoD organizations tasked with analysis of Africa. This would require in-depth discussions with relevant stakeholders to improve understanding of their respective qualitative data needs. In some cases, detailed qualitative data requirements might not be known or clear to analysts. In such cases, ascertaining their broad operational priorities in conjunction with the processes and means used to achieve them might help isolate specific data needs. This process would also afford DoD an opportunity to refine the definition of these data needs to a level of detail that would facilitate collection.

Prioritization can occur in different ways, depending on the consumer of the data. The individual M&S could be prioritized themselves based on the USG's assessment of their utility. Because M&S are idiosyncratic, data needs would be derived from an identification of each M&S's missing inputs. This would obviate the need for a larger discussion among the broader community of interest, yet would only serve the needs of that particular model or simulation. Thus such a case-by-case approach would ensure the exact specifications of data needs are known (such as formatting requirements) but this approach could become quite costly without considering efficiencies with other M&S.

A more cost-effective approach would be further collaboration among DoD organizations, incorporating the inputs of as many stakeholders as possible including but not limited to AFRICOM, U.S. Special Operations Command (SOCOM), TRAC at Fort Leavenworth (TRAC-FLVN), Undersecretary of Defense for Intelligence (USDI), and other relevant stakeholders. Ascertaining the most frequently cited data gaps would help OSD to determine those data points that would have the maximum utility for all. Once collected, these could satisfy the needs of numerous stakeholders.

Discussions should address geographic priorities (e.g., Horn of Africa (HOA), Sahel, west, central, south), thematic priorities (e.g., relationship-mapping, public opinion/attitude, demographic), and any other relevant characterization of data as identified by stakeholders. Following such an analysis, OSD can officially prioritize the community's qualitative data requirements

Finding 6: There is potential for proven methodologies to yield previously unavailable qualitative data from Africa.

Through discussions with historians, anthropologists, and social scientists, IDA learned of various unconventional qualitative data collection techniques that have the potential to yield valuable data from Africa. For example, a group of analysts within IDA have a proven track record using one such methodology in Southeast Asia where it has successfully formed collaborative research networks that facilitate access to local data. As a result of IDA's engagement in this region, we have been able to augment the USG's existing knowledge base for Southeast Asia and contribute to a deeper understanding of the nuances and complexities that characterize the region.

Recommendation 6a: Cultivate collaborative research networks for improved access to local data.

IDA recommends that DoD test promising qualitative data collection methodologies in Africa, specifically those that have proven to be fruitful in other regions or contexts. The features of one such network that has facilitated the effective elicitation of sensitive information (otherwise inaccessible to the USG) include:

- A focus on "track 2 engagement" with participation from non-official counterparts, e.g., traditional authorities, civil society, youth groups, religious institutions, the private sector, or academia. Individuals from such backgrounds represent purposive samples, i.e., samples that are "information rich," which provide the greatest insight into the data being sought.[32]

- A sustained engagement that cultivates trust, strengthens relationships, and encourages collaboration – all in support of shared interests. Almost all qualitative research methodologies require the development and maintenance of relationships with research subjects, which is important for effective sampling and for the credibility of the research.[33]

---

[32] Kelly Devers and Richard Frankel, "Study Design in Qualitative Research—2: Sampling and Data Collection Strategies" *Education for Health*, Vol. 13, No. 2, 2000, 263–271. And Miles & Huberman (1994, p. 34)

[33] Frankel, R.M. and Devers, K.J., (2000a). Qualitative research: a consumer's guide. *Education for Health*, 13, 113–123.

- A "strategic listening" approach, whereby U.S. participation is minimal, encouraging a comfortable venue for frank and open dialogue among local participants. This approach allows U.S. observers/listeners to absorb insights relating to the nature of the environment through local lens.

- A discussion on topics of mutual concern. The most detailed and insightful data can be drawn from discussions where all participants have a vested interest. Often, Western-led discussions in Africa revolve around Western notions of security, democratization, institution-building, or other topics that are the focus of Western research in Africa. As phase 2 of this task revealed, African perspectives on such issues often do not correspond with Western perspectives. The value of the forum described in this methodology is that it elicits African perspectives without the overlay of U.S. concerns.

Recommendation 6b: Engage diaspora communities residing in CONUS.

The African diaspora that resides in the U.S. is a largely untapped source of qualitative data. Preliminary IDA research suggests that African communities, such as a large Somali diaspora residing in Minneapolis or Washington DC, can provide invaluable qualitative data for U.S.-based researchers.[34] Not only are the diaspora a resource geographically convenient to U.S.-based researchers (negating the need to make costly data collection trips to Africa), but engaging them in the U.S. overcomes many of the challenges of bureaucracy and corruption often associated with data collection in Africa.[35]

IDA recommends that the DoD test qualitative data collection methodologies that target African diaspora communities residing in the U.S. This is a cost-efficient way to elicit critical data points from population samples that might serve as useful proxies for otherwise inaccessible populations in Africa.

There will, however, be some challenges when engaging African diaspora in the U.S. First, as with the previous recommendation, members of the African diaspora (as well as Africans who reside in their home countries) are likely to be reluctant to engage with the USG. Based on IDA's previous experiences, foreign nationals residing in the U.S., particularly those from the Middle East and Africa, often sense suspicion from the USG who they feel might be monitoring actions that could be construed as extremist activity. Conversely, academia and independent research institutes have enjoyed success

---

[34] Janette Yarwood, "A New Threat: Radicalized Somali-American Youth," IDA Research Notes, Summer 2012. Available at: https://www.ida.org/upload/research%20notes/researchnotessummer2012.pdf

[35] Ashley Bybee and Dominick Wright, IDA Document D-4629, Designing a Qualitative Data Collection Strategy (QDCS) for Africa – Phase I: A Gap Analysis of Existing Models, Simulations, and Tools Relating to Africa, June 2012.

in engaging foreign populations. This is one key to eliciting sensitive information that otherwise remains unavailable to the USG. Secondly, diaspora populations tend to be inherently biased in many ways based on their decisions to leave their native countries and reside in the U.S. Their loyalty to, allegiance to, or perceptions of their native countries will likely differ from citizens who continue to reside in Africa based on experiences that may have prompted them to emigrate (e.g., persecution, oppression, desire for greater opportunities). Therefore the data collected from these individuals must be carefully assessed for potential bias and applied appropriately. This requires the objective assessment of a trained researcher familiar with the region from which they come who can differentiate subtle differences in attitudes.

Finding 7: There are several new DoD initiatives and methodologies under way for collecting socio-cultural data in Africa.

IDA discovered several recent initiatives and possible new methodologies currently deployed for collecting socio-cultural data. Special Operations Command (SOCOM), for example, is currently experimenting with remote sensing (i.e., sensor-based methodologies for data collection in the Sahel and Uganda, for tracking AQIM and the Lord's Resistance Army (LRA), respectively). The National Reconnaissance Office (NRO) is looking to deploy an instantiation of Savanna (the primary analytic interface currently used by AFRICOM to display Serengeti data) with more than a thousand licenses for use throughout the community as part of the Defense Intelligence Information Enterprise (D2IE). The Joint Staff J-7 is working with the geographical combatant commands (GCC) to draft a CONOPS for civil information fusion centers (CIFCs). Each GCC has its own subjective data needs, but alternate methods for collection should prove universally-applicable across a number of them.

Recommendation 7: Work with the interagency to support experimentation and deployment of new methodologies for socio-cultural data collection.

IDA recommends that OSD consider working with these nascent initiatives and existing organizations to develop an overall strategy for data collection and storage from which the entire community can benefit. Where opportunities exist to collaborate with DoD and other interagency partners to experiment with the deployment of new methodologies for socio-cultural data collection, IDA recommends OSD partake in these activities to maximize efficiencies across the government.

Finding 8: Need to facilitate personal contacts and raise awareness of qualitative data sources among the community of interest.

As stated in finding one, IDA finds that the community of qualitative data consumers has varying levels of awareness and access to currently available data. We

attribute these differences to the client-driven and therefore *ad hoc* nature of many data collection efforts. The use of one data portal, such as DataCards, will make great strides toward increasing awareness and access of all data across the community. Knowledge of these data sources, however, will always be contingent on the degree of use of such a portal. Moreover, there will always be new data sources coming online that will not be captured in such a portal. A secondary mechanism to track and raise awareness of such data collection efforts would be beneficial to ensure maximum exposure across the community of analysts and other stakeholders.

Recommendation 8. Partner with NDU to hold regular conferences convening data collectors and owners of qualitative data.

Given the overlapping interests among data consumers and data producers, there is a potential synergy to be gained by convening these communities in a mutually beneficial forum. The regular Socio-Cultural Data Evaluation Summits (otherwise referred to as "Data Summit") organized by NDU are ideal opportunities to assemble these different communities that have very similar interests. These summits have, to date, discussed the current state of the DataCards database as well as larger substantive issues pertaining to socio-cultural data (e.g., definitions, applications, integration into platforms). Given the heavy focus of DataCards on AFRICOM's Area of Responsibility (AOR) due to a higher level of sponsorship from that Command than other partners, IDA believes there are major benefits to be gained by those interviewed for this task by including them in these events.

This forum would also be a prime opportunity to convene data producers alongside data consumers to showcase collection efforts currently under way that will yield new data sources in the near future.

5. Long-Term Plan to Grow Local Capacity

A near-term "surge" in the collection of the most pressing data gaps should be accompanied by a long-term plan that ensures a sustainable flow of those and other desired data. Given that one of the most frequently cited data gaps is the time series data, which enables analysts to track trends (trends that often signal imminent instability), ensuring a constant and reliable stream of these data should be a top policy priority for the USG. This aspect of the QDCS will require the greatest investment of resources, but the return on that investment is invaluable not only for MS&T but for all analysis of African trends.

Finding 9: Local capacity for qualitative data collection is low.

Broadly speaking, and with few exceptions, the local technical capacity in Africa for data collection is extremely low. As with most developing regions, scarce resources (from national governments and donors) are focused on high priority activities, with less vital activities (such as data collection) often left by the wayside. Moreover, the requisite skills to administer surveys, conduct interviews, and other qualitative methodologies are severely lacking. In other words, very few Africans know how to collect qualitative data, even if they had the financial resources to do so. (It should be noted, however, that although local capacity for data collection is low, there are a handful of professional firms in Africa with trained, competent staff. These firms are typically sub-contracted by western or international organizations to collect data, and do so efficiently and effectively.)

As a result, targeted qualitative data collection is typically performed by external actors (mostly Western) on an *ad hoc* basis to serve immediate data needs, usually for academic purposes. Data are not collected in a sustainable fashion or using a consistent methodology at regular intervals over time, which contributes to the aforementioned problem of poor time series data. As a result, data consumers, including but not limited to DoD and other U.S. government agencies, do not have the qualitative data they require to populate their MS&T and perform various other types of analyses.

The solution is to grow the technical capacity of local data collectors, so that the USG (and others) can eventually leverage the data collected by Africans without continuing to invest massive resources indefinitely. Such investments could have a strategic payoff (access to data), while contributing to DoD's partnership capacity building mission. In areas where DoD could benefit from more and improved qualitative data (socio-cultural data that could assist with counter-terrorism operations), it would be

appropriate for OSD to support the growth of local capacity for qualitative data collection. There are several avenues through which OSD can contribute to this endeavor, which are described next.

Recommendation 9. Increase technical training for local qualitative data collection, especially capacity to execute national censuses.

IDA recommends DoD partner with local data collection organizations that have the capacity themselves to run technical training programs for local Africans. These may be government institutions, private African organizations, or multinational organizations operating in Africa. All can benefit from improved training, but building upon the work and achievements of African data collection institutions in the region is not only efficient, but it prevents the problem often encountered by Western institutions whose prescriptive approach is resented by Africans. Recognizing the initiative and progress of local institutions ensures an "African solution to an African problem." IDA also recommends facilitating collaboration among such institutions to leverage each other's capabilities and share lessons learned.

Independent, African-based survey firms such as Afrobarometer, academic institutes such as the Centre for Social Science Research at the University of Cape Town, or multinational organizations such as the UN Office on Drugs and Crime (UNODC),the UN Institute for Training and Research (UNITAR) or the UN's Economic and Social Council (ECOSOC) are among some reputable organizations and potential partners for this activity. With guidance from DoD and its African or international counterparts, these training programs can be tailored to ensure both parties benefit by targeting mutually agreed-upon data requirements. Mutual benefit is essential to ensure the sustained flow of data after training is complete or funding ceases.

There are numerous substantive areas in which technical training can be concentrated. Given the rapid expansion of cell phone usage in Africa in recent years, training for mobile phone surveys would be one important area of focus.[36] Yet the most critical area to build local capacity for data collection should support the administration of national censuses. Currently census data are available for some African countries that have achieved a certain level of technical capacity for data collection. A common complaint among the M&S community, however, is that in many countries, censuses may not be publicly unavailable, they may be unreliable (due to poor collection techniques or official manipulation), or they may be absent altogether. The absence of critical socio-economic and demographic data, which are essential for public policy analyses, also

---

[36] Benjamin Loevinsohn, "Collecting Household Level Data in South Sudan Through the Use of Mobil Phone Surveys Cluster Leader," The World Bank, Power Point Presentation delivered at the World-Wide Human Geography Data Working Group, U.S. Geological Survey, Reston, VA, 27-28 March 2012.

impedes the effective execution of established survey methodologies (i.e., Primary Sampling Units stratified by population size).

IDA recommends DoD work with the International Programs Center for Technical Assistance, Population Division, U.S. Census Bureau, U.S. Department of Commerce, to coordinate training for African government census takers. The U.S. Census Bureau has already worked with the National Bureau of Statistics (NBS) of the Republic of South Sudan to enhance its statistical capacity.[37]

---

[37] Oliver P. Fischer, "U.S. Census Bureau in Sudan," International Programs Center for Technical Assistance, Population Division, U.S. Census Bureau, Power Point Presentation delivered at the World-Wide Human Geography Data Working Group, U.S. Geological Survey, Reston, VA, 27-28 March 2012.

Appendix A: MS&T Survey

Competitive Influence Game (CIG)

Producer: Johns Hopkins University, Applied Physics Lab (APL)

Type: Simulation (Independent & Federated) – CIG is an "independent" simulation because it can run entirely on its own when provided sufficient amounts of data inputs. It also has the ability to federate (i.e., the ability to combine with multiple model or simulation inputs), as it is equipped with a Federation Object Model (FOM), which describes the shared object, attributes and interactions for the whole federation. It is unclear at this time whether the CIG FOM satisfies governmental, High Level Architecture (HLA) standards.

Purpose: Currently used to support exercises and high-level wargaming (e.g., the AOWG/AWG Cycles), its developers at APL originally conceived of it as an attempt to provide a generalized behavioral model characterized after the fictional Seldon equations (the one elaborated upon by Isaac Asimov in the 1951 novel, *The Foundation*). Asimov described the Seldon equations as essentially statistical models with historical data of a sufficient size and variability that they are collectively representative of the population under consideration. The intent is not to provide point predictions that accurately capture the behavior of an individual but instead to generate accurate forecasts of how populations will behave in the aggregate. CIG adheres to the spirit of Seldon equations in structure but variation in the number, quality, and empirical anchoring of inputs causes it to differ in form.

Inputs: Generation of behavioral outcomes in CIG is similar to that of tabletop board games, such as Risk and others that model probabilistic outcomes using die rolls. Although probability distributions are always normal or "bell curves," their shape (i.e., location of mean values and population variance) results from the conditional mapping of behavioral outcomes within the game. Currently, the setting of "initial conditions" or starting values for data in the simulation along with the properties governing values for the conditional mappings occurs primarily according to subjective inputs from SMEs. While all of the SME-elicited relational estimates are qualitative, the nature of "initial conditions" inputs describing existing conditions varies between quantitative and qualitative.

Composite Vulnerability Map

Producer: University of Texas, Climate Change and African Political Stability Program (CCAPS)

Type: Web-Based Tool

Purpose: The Composite Vulnerability Map models which parts of Africa are most vulnerable to climate change in the mid-21st century range. It provides scholars, policymakers, analysts, and those supporting them with the ability to visualize imagery,

events (from human behavior), and other types of related data in the effort to characterize the relationship between various physical and social environmental variables and human conflict. Its most mature capability is the ability to generate layered visualizations containing imagery data, such as precipitation, and a large variety of violent events (i.e., sub-nationals against sub-nationals, states against sub-nationals, and sub-nationals against the state). Data on governance characteristics will eventually extend beyond that available in other datasets (e.g., PolityIV) to describe state features of constitutional processes and other manner of non-quantitative and subjective information. Besides making visualization tools accessible by the public, the project also provides links for downloading the represented data.

Inputs: Imagery data (e.g., drawn from NASA, NGA, and other similar sources) and originally collected, spatio-temporal (i.e., geo-located and time-coded) event data from systematically coded news events. Maintaining updated imagery information is an external matter for the project and characterizing historical processes leading to socio-political events, such as referenda and drafting, are fixed, historical features of countries requiring only one pass to provide information (unless the feature in question changes). On the other hand, event data in the tool suffers from a lag between social processes generating events on a daily/weekly/monthly/quarterly rate (depending on the nature of conflict in the specific locale) and the ability to code them into datasets.

Cultural Geography (CG)[38]

Producer: United States Training and Doctrine Command (TRADOC), Analysis Center (TRAC)-Monterey

Type: Pseudo-Agent Based Model (ABM)

Purpose: The purpose of CG is to provide a platform for considering the consequences of kinetic and non-kinetic actions taken by military actors within simulated socio-cultural environments. It is part of the Social Impact Model (SIM) system, which is a type of model federation described as "a tool for irregular warfare adjudication, analysis, and validation." Given that the capability hails from TRADOC, its primary purpose is to support training in areas such as the selection and prioritization of courses of action (COAs) within the context of a COIN socio-cultural environment.

Inputs: CG possesses the ability to model micro-level agents, but the complexity of its architecture and vastness of its parameters has in practice led to the modeling of "representative agents." Examples of such actor agents include a community, a government, an ethnic group, an insurgent, and so forth. Individually, requisite inputs include data on the preferences these actors hold over a variety of outcomes, prior beliefs about the preferences of other actors, relational mappings for actions and changes to the environments as indirect influences on outcome evaluations, and so forth. Social network

---

[38] The IDA team received access to the code and other documentation for CG. Additionally, IDA coordinated with National Defense University which is overseeing a validation project for CG and ATHENA. Working through the complex architecture and processes of CG is an extensive effort extending beyond the scope of IDA's tasking, so the team has relied upon available documentation as well as interviews with TRAC-Monterey and NDU to complete this entry in the report.

components in the model require data on the relationships between groups, i.e., who shares a connection with whom and the relative value of this relationship. These are just some of the numerous data inputs for calibrating parameters in the model.

Geospatial Information Awareness/Infection Disease (GIA/ID)

Producer: Naval Research Lab (NRL)

Type: Computational Analytic Model

Purpose: Africa is a continent where the emergence and spread of disease are persistent threats. Enhancing geospatial information for the purpose of situational awareness has gained traction and considerable development throughout the West. GIA/ID is an initiative led by NRL to expand the community of interest and practice throughout Africa. As an initial step, GIA/ID is a "proof-of-concept" attempt to demonstrate the ability to identify the emergent flash point of a disease (geo-referenced), to track its spread (geographically and temporally), and to identify factors—including social and environmental – associated with these empirical trends. The hope is that if conducted successfully, analysis of these three components will provide indicators and warnings for American and partnering forces. Additionally, outputs from GIA/ID should identify interventions tailored to the specific socio-environmental conditions responsible for identified pandemics, limiting the need to rely upon "cookie cutter" solutions commonly applied under conditions characterized by low information.

Inputs: Current inputs to GIA/ID include an extensive surveying of the population in the Sierra Leoneon town of Bo, used to establish what NRL analysts described as the denominator. Specifically, the denominator is a geo-referenced count of the population on a grid-by-grid basis across the territory. This required extensive resources to collect. Another input is the counting of diseased individuals, which constitutes the numerator. At the time IDA discussed the project with NRL, the identification of cases was relatively accurate (i.e., use of a university-donated, genomic analyzer facilitated the efficient identification of pathogens in blood serum), as too was its temporal tagging (i.e., association of the identified case with a date of collection – though there is a difference between identifying when transmission of a pathogen took place versus when a patient makes it to a clinic or hospital). What the data lacked was an implemented means to geo-reference the reported incidence of disease within the grids established during the initial surveying of the population. Territory in Bo is not systematically organized in a manner that residents can readily provide meaningful addresses, which was the primary culprit for this initial lack of geo-referenced cases. A proposed solution at the time of the interview included having doctors present maps of the area to patients for them to use when identifying their place of residence.

HOA-Viewer

Producer: Department of State (DoS), Humanitarian Information Unit (HIU)

Type: Web-Based Tool

Purpose: Intentions for the HOA-Viewer are twofold. First, HIU wants the tool to equip users (e.g., analysts, service providers, policymakers, and so forth) with the ability to

visualize and interact with data in a manner that exploits geospatial and temporal characteristics of humanitarian crises (both the crisis events themselves as well as the circumstances preceding and following them). HIU also aspires for HOA-Viewer to be an analytic support tool by eventually infusing it with qualitative and quantitative methodological functions.

Inputs: HOA-Viewer inputs include a broad array of imagery data (theoretically, the system can capture any level of imagery data available), United Nations Humanitarian Crisis (UNHCR) Reports (unstructured text), and other geospatial data (e.g., ethnicity and population size polygons as well as event point data). Metadata each input includes geospatial and temporal components, which enable the viewer to visualize on maps various patterns of events (currently, the focus is on the representation of climate imagery data).

Information Velocity 2.0 (IV2)

Producer: Office of the Secretary of Defense, Science and Technology

Type: Web Information Harvesting Tool

Purpose: Surveying populations is an effective means for tracking attitudes and sentiment, but it is a timely process with uncertainty surrounding the conditions producing responses. Rather than survey populations directly, Web 2.0 products, such as Twitter and Facebook, provide the opportunity to track attitudes and sentiments in a populous, as expressed directly by individuals (i.e., without the response and construction biases of surveys but also without their controllability). IV2, which is currently under development as a governmental specification for currently, "commercial off the shelf" (COTS) products, plans to tap into this resource in the effort to provide AFRICOM (and by extension other global combatant commands) with the ability to track and potentially predict the occurrence of flash points associated with mass unrest throughout the African area of operations. [IV2 and similar capabilities under development, such as Mitre's Social Radar, use the examples of the London riots and the Arab Spring as cases in point for harnessing Web 2.0 technologies]. IV2 developers envision that automated reference extractions from Web 2.0 associated with Web 1.0 (e.g., newsfeeds along with company and individual profile webpages among others) will result in a broader contextual understanding, higher situational awareness, and potential ability to act than either capability alone provides.

Inputs: IV2 inputs will include Web 2.0 (e.g., Twitter, and Facebook) feeds in addition to Web 1.0 targeted page scraping, conditioned on Web 2.0 extractions. Importantly, when thinking about the application of IV2 and similar technologies, it is important to consider the informational austerity of the population in question and the targeted objective of the capability. Public opinion polling, which – when done well (e.g., according to standards followed by AfroBarometer, Gallup, and the State Department Office of Opinion Research among others) – is representative of the population in question with an identifiable degree of uncertainty (i.e., with confidence intervals on reported percentages). If the goal is to use the IV2 capability as an alternative to public opinion polling, then it will be necessary to use it on online populations that are accurate subsets of the entire population (i.e., randomly available online in a manner similar to

samples generated from randomized, stratified sampling used to construct survey populations) or to at least have determined the systematic bias distinguishing expressed online sentiments from those counterfactually gathered in person.

RiftLand

Producer: Center for Social Complexity, George Mason University

Type: Agent Based Model (ABM)

Purpose: Generally speaking, RiftLand models humanitarian crises in East Africa. Based on the description of its predecessor, RebeLand, the analytic goal of the model is to study conditions of political stability, specifically the ability of a system to withstand, various forms of stress, such as social, economic, political, or environmental. The name of the model implies that it focuses on the area in Kenya known as the Rift Valley. Following the 2007 Presidential elections, the Rift Valley was one of the areas that erupted into violence as disputed election results resonated with a long history of inter-ethnic rivalry and conflict among residents. Numerous violent events and large-scale internal displacement resulted in widespread instability throughout the Rift region. IDA infers that one goal of RiftLand is to identify regional or functional areas where government action may help to prevent future instability.

Inputs: RiftLand, as a "real world" version of RebeLand, is an attempt at modeling an entire polity. According to documentation for RebeLand, some of the basic inputs required for doing this include a range of geospatial information (e.g., provincial boundaries, topography and land cover, location and size of cities, location as well as type and amount of natural resources), location along with type and composition of military (state and non-state) forces, climate data (e.g., rainfall/drought, wind, and temperature), hydrology, and so forth.[39] Corresponding data requirements for RiftLand, beyond the basic descriptive characterizations of the local population, are not yet documented for public consumption.

Modeling societal effects of naturally-occurring or manmade phenomena require values for those actions as well as data on the relational mapping between changes in these values and outcomes of interest. Other implicit inputs to RebeLand include how changes in community context and individual wellbeing affect recruitment of rebel and other anti-state groups. Authors emphasize the characterization of community issues relative to government activity. Abstractly, it is possible to work through the analysis of this problem without "real world" data, but linking the two (i.e., determining what the definition of an issue and its relevant dimensions are for coding in a dataset for ingestion to the model) is necessary for accurate modeling. Documentation for RebeLand does not explicitly identify contextual and relational data as necessary inputs, but it is clear that the utility of RiftLand depends upon capturing this information along with the descriptive data already identified as inputs.

---

[39] Claudio Cioffi-Revilla and Mark Rouleau, "MASON RebeLand: An Agent-Based Model of Politics, Environment, and Insurgency" *International Studies Review* 12, 31-52, 2010.

Unnamed

Producer: Naval Postgraduate School (NPS), Operations Research Department

Type: Web-Based Data Visualization Tool (with future possibilities for analysis development)

Purpose: This tool under development at NPS intends to make survey data more accessible to end-users who are not well versed in the handling and exploitation of survey data. Currently, exploitation of raw, survey data requires some facility with software tools, such as those in the Microsoft Office suite (Excel and Access) or more traditional statistical analysis platforms (e.g., R, Stata, SPSS, and Gauss to name a few). Even those capable of using such programs find it difficult to visualize and understand calculated results geographically, because doing so necessitates the additional skills required to work either mapping functionalities within the aforementioned platforms (mainly the alternate packages available in R) or to import and manipulate them within a geospatial analysis platform, such as Esri's ArcGIS suite. The product under development at NPS seeks to overcome both hurdles for end-users who do not have time to develop the necessary skill sets but nonetheless need the data and the insights it brings.

Inputs: The tool ingests survey data, which makes the quality of its outputs entirely dependent upon that of its inputs. This means it is sensitive to common survey data issues, such as sample construction, question validity, timeliness, along with a host of others. Efforts made to resolve these problems will translate directly into the quality of insights the NPS visualization tool provides.

Appendix B: Spreadsheet of Data Sources

This spreadsheet contains a list of the data sources that the IDA team encountered over the course of this study. These are sources that were a) cited by developers as inputs they used for their MS&T; b) referenced by interviewees as known data sources; or c) discovered by IDA in its research and attendance at various Africa-related events. While this project focuses on qualitative data, there were numerous quantitative data sources that were identified during the course of the study. They are also included in this list in an effort to capture as many data sources as possible.

The descriptive data in this spreadsheet has been organized to facilitate the entry of each data source into the Datacards portal. The columns, for example, correspond with the fields required by Datacards. The sources in this list have been compared with Datacards holding, and, where they are already included, they are indicated as such to prevent duplication. They are nonetheless included here in case this list offers any additional descriptive information that may be added to the Datacards entry. Entries are ordered first by inclusion in DataCards (those which are *not* included are listed first), then by the scope of their geographic coverage (those which contain data for Africa only are listed first, followed by entries that contain worldwide data).

The list presented below (in hard copy) shows the most important descriptive fields where the most information is known. The electronic file includes the remaining fields that exist in Datacards. Due to time constraints, IDA did not populate most of these fields, though where the information was easily available, it is included.

Title	Recognized authority	Data Category	Scope of geographic coverage	Data source format(s)	Description of data	Data storage location	Limitations on Access to the data source	In Data-cards?
92 surveys	DataFirst Resource Unit, University of Cape Town	Public Perception/ Atmospherics	Africa		Contains digital archive of 92 survey datasets.	www.datafirst.uct.ac.za/home/	Varies	No
Language Resources	Open Language Archives Community	Religion/Culture/ Language/Ethnicity	Africa		The Open Language Archives Community is an international partnership of institutions and individuals who are creating a worldwide virtual library of language resources by: (i) developing consensus on best current practice for the digital archiving of language resources, and (ii) developing a network of interoperating repositories and services for housing and accessing such resources.	www.language-archives.org/area/africa	Public	No
Interactive Infrastructure Atlases	African Development Bank Group	Infrastructure/ Construction/ Technology Terrain	Africa	Geo-PDFs, ARCGIS .shp	Geo-PDFs containing a large number of infrastructure layers that can be switched on and off in any desired combination to create a map customized to the user's interest. For example, the user could create a map showing ICT and power transmission infrastructure only, or a map showing the location of road and rail networks. In addition to the maps, ArcGIS files have been placed on the website.	http://www.infrastructureafrica.org/tools/data	Public	No
Global Administrative Areas	Robert Hijmans	Terrain	Africa	Shapefile, ESRI, GoogleEarth (kmz), R (SpatialPolygon sDataFrame)	GADM is a spatial database of the location of the world's administrative areas (or administrative boundaries) for use in GIS and similar software. Administrative areas in this database are countries and lower level subdivisions such as provinces, departments, bibhag, bundeslander, daerah istimewa, fivondronana, krong, landsvæðun, opština, sous-préfectures, counties, and thana. GADM describes where these administrative areas are (the "spatial features"), and for each area it provides some attributes, such as the name and variant names.	www.gadm.org/home	Public	No

B-2

Title	Recognized authority	Data Category	Scope of geographic coverage	Data source format(s)	Description of data	Data storage location	Limitations on Access to the data source	In Data-cards?
The Africa Soil Information Service (AfSIS)	Center for International Earth Science Information Network (CIESIN), Earth Institute, Columbia University	Terrain Weather/Climate/Envir onment/Wildlife	Africa	GeoTIFF (.tar.gz)	The Africa Soil Information Service (AfSIS) Web Map Service supports the Open Geospatial Consortium (OGC) OpenGIS Web Map Service (WMS) Implementation Specifications and dynamically produces maps of georeferenced data. Support of this international standard opens the AfSIS map collection to users who can access its contents via machine-to-machine interaction.	www.ciesin.columbia.edu/afsi s/bafsis_fullmap.htm#	Public	No
Mapping Malaria risk in Africa (MARA)	Malaria Research Programme of the South African Medical Research Council	Health/Nutrition/Diseas e	Africa	.xls, PDF	Geo-referenced data and maps of endemic malaria models for Africa as well as Population, administrative & topo maps of Southern Africa. Data on malaria prevalence data, malaria distribution maps and estimated populations at risk.	http://www.mara.org.za/	Public	No
African Geodetic Reference Frame (AFREF)	Regional Centre for Mapping of Resources for Development (RCMRD)	Terrain Imagery	Africa	Selected maps and imagery available	The African Geodetic Reference Frame (AFREF) was conceived as a unified geodetic reference frame for Africa to be the fundamental basis for the national and regional three-dimensional reference networks fully consistent and homogeneous with the International Terrestrial Reference Frame (ITRF).	http://www.rcmrd.org/index.ph p?option=com_content&view =section&layout=blog&id=6&lt emid=54	Public	No
Multiple	Center for Strategic Counterterrori sm Communicatio ns (CSCC)	Terrorism	Africa and Middle East		Through media analysis (twitter, etc.) the CSCC produces analysis of violent extremist media strategies, including the appeal of Al Shabaab, Al Shabaab's preferred media outlets, Recruitment themes, Best methods to counter AS's message (validated through MOEs), etc.		Upon request	No
Database on dams	Aquastat (Food and Agricultural Organization, UN)	Agriculture/Food Security Natural Resources/Energy/Sani tation/Water	Africa and Middle East	.xls	Contains geo-referenced information on around 1300 dams in Africa and 1100 dams in the Middle East	http://www.fao.org/nr/water/aq uastat/dams/index.stm	Public	No

Title	Recognized authority	Data Category	Scope of geographic coverage	Data source format(s)	Description of data	Data storage location	Limitations on Access to the data source	In Data-cards?
International Political Interactions (IPI) Project		Violence/Security Military Operations/Activities/C apabilities	Africa: Nigeria, Democratic Republic of Congo, Zimbabwe Asia: Indonesia, Pakistan, South Korea Latin America: Argentina, Brazil, Chile, Colombia, Mexico, Venezuela Europe: Belgium, Hungary	.csv, .xls	The IPI project was designed to measure political conflict and cooperation within societies through the coding of political event reports from international, regional, and local sources. These events were coded on two ten point scales which reflect the severity of various cooperative and conflicting statements and actions. This scaled events data can be used to calculate the volume and intensity of political conflict and cooperation within the domestic polity. In addition to facilitating the calculation of general levels of political conflict, the IPI coding scheme allows the examination of the dynamics of interaction among specific groups within the society. IPI gives scholars the ability to track interactions among social groups and between the state and social groups.	http://mailer.fsu.edu/~whmoor e/garnet-whmoore/ipi/cases.html	Public	No
State Department Surveys	Office of Opinion Research (OOR), U.S. Department of State	Public Perception / Atmospherics	As requested	PDF	OOR collects opinion data worldwide and has three analysts dedicated to sub-Saharan Africa. Quantitative and qualitative data are collected and OOR has recently begun some editorial opinion analysis of local news sources. Time range is snapshots. However, there are certain questions which are included in most surveys, which address: perceptions of the U.S., the most urgent issue in a given country (answer is the economy), ethnic tolerance, among others. These are typically asked on an annual basis, or however often OOR conducts research in that country. As such, they are able to infer some general trends in opinion, but their methodology for collecting these data are not nearly as rigorous as Gallup.	OOR's government website.	"Opinion Analysis" available to USG on request. Raw data not available due to licensing restrictio ns but excerpts and analysis available on request.	No

Title	Recognized authority	Data Category	Scope of geographic coverage	Data source format(s)	Description of data	Data storage location	Limitations on Access to the data source	In Data-cards?
Civil Affairs Operating System (CAOS)	95th Civil Affairs Brigade, US Army Special Operations Command	Other	Chad Mali Mauritania Senegal Morocco Niger Nigeria HOA	Website	Central repository for all Civil Affairs (CA) data collected in the field by CA forces. Mainly infrastructure and key leader data.	https://uroc.usace.army.mil/caos	CAC required	No
Violent Intranational Conflict Data Project (VICDP)	Florida State University	Violence/Security Military Operations/Activities/Capabilities	Colombia Nigeria Peru Sri Lanka Zimbabwe Some data were collected for both Lebanon and the Philippines	Lotus 123	The Violent Intranational Conflict Data Project (VICDP) was created by Will H. Moore in 1992 to produce events data for the study of violent intranational conflict.	http://dvn.iq.harvard.edu/dvn/dv/wmoore/faces/study/StudyPage.xhtml;jsessionid=8fa9172f3035e7300f46cfbcc98c?globalId=hdl:1902.1/10687&studyListingIndex=0_8fa9172f3035e7300f46cfbcc98c	Public	No
Kenya OpenData	Government of Kenya	Public Perception / Atmospherics	Kenya	.csv, .json, .pdf, .rdf, .rss, .xls, xml	The 2009 census, national and regional expenditure, and information on key public services are some of the first datasets to be released. The website is a user-friendly platform that allows for visualizations and downloads of the data and easy access for software developers.	https://www.opendata.go.ke/browse	Public	No
Foiben-Taosaintanin'i Madagasikara (FTM)	Institut Géographique et Hydrographique National, Madagascar government	Terrain Imagery	Madagascar		FTM provides geographic information on Madagascar; it offers traditional products for the general public (road maps, city maps) as well as products for professionals from the environment, tourism, rural development, urban planning sectors.	http://www.ftm.mg/home.htm	Upon request	No
National Mapping and Remote Sensing Center	Government of Mozambique	Terrain Imagery	Mozambique		land cover maps from Spot and Landsat imagery	http://www.cnt.nat.tn/en/index.php	Public	No
Council for Geosciences	Government of South Africa	Terrain Imagery Infrastructure/Construction/Technology	Primarily South Africa	PDF	Provides analysis and access to geographic and environmental data including multiple databases, maps (geological, mineral, poverty, infrastructure), and research reports.	http://www.geoscience.org.za/index.php?option=com_content&view=article&id=269&Itemid=337	Public	No

Title	Recognized authority	Data Category	Scope of geographic coverage	Data source format(s)	Description of data	Data storage location	Limita-tions on Access to the data source	In Data-cards?
Human Sciences Research Council	Government of South Africa	Governance/State-Society Relations Development Programs/ Reconstruction Health Nutrition/Disease Labor/Employment Migration/Immigration/ Demographics/ Population	Primarily South Africa		HSRC conducts large-scale, policy-relevant, social-scientific projects for public-sector users, non-governmental organisations and international development agencies. Raw data available (upon request in most cases) on: Democracy, Governance and Service Delivery, Economic Performance and Development, Education and Skills Development, HIV/AIDS, STIs and TB , Human and Social Development, Population Health, Health Systems and Innovation.	http://www.hsrc.ac.za	Upon request	No
Regional Remote Sensing Unit (RRSU)	SADC	Terrain Imagery	SADC	Interactive maps, GIS datasets, Satellite Imagery	This key sub-regional organisation provides data for mainly food security analysis. It also hosts and distributes LandSat imagery for free	www.sadc.int/geonetwork	Upon request	No
Southern Africa Regional Poverty Network (SARPN)	Southern Africa Regional Poverty Network (SARPN)	Government Financial Statistics Migration/Immigration/ Demographics/Populati on Labor/Employment	SADC	PDF	Data on poverty, papers on poverty on the different SADC countries, statics and reports	http://www.sarpn.org/index.php	Public	No
Master Narratives	Open Source Center (USG) and Monitor 360	Religion/Culture/Langu age/Ethnicity Other	Selected case studies in Africa include: Somalia, South Sudan, Algeria, Egypt, and Al Qaeda.	Text	Master narratives are the historically grounded stories that reflect a community's identity and experiences, or explain its hopes, aspirations, and concerns. These narratives help groups understand who they are and where they come from, and how to make sense of unfolding developments around them.	www.opensource.gov/portal/server.pl/community/master_narratives/1203	USG and contractors	No
Spatial and Land Information Management, Department of Water Affairs	Government of South Africa	Terrain Imagery Natural Resources/Energy/Sani tation/Water Weather/Climate/Envir onment/Wildlife	South Africa	.shp	SLIM provides spatial and land information services, which include: Aerial mapping, hydrographic surveying, GIS, precise engineering, control surveys, GPS surveys and a cadastral and topographic information services.	http://www.dwaf.gov.za/BI/Systems	Upon request	No

Title	Recognized authority	Data Category	Scope of geographic coverage	Data source format(s)	Description of data	Data storage location	Limitations on Access to the data source	In Data-cards?
South African boundaries	Municipal Demarcation Board, Government of South African	Other (administrative boundaries)	South Africa	Google earth files, PDF	Ward, district, municipal, and provincial boundaries.	http://www.demarcation.org.za	Public	No
Tanzania'a Socio Economic Database	National Bureau of Statistics, Government of Tanzania	Migration/Immigration/ Demographics/Population Labor/Employment	Tanzania		TSED is a comprehensive socio-economic database system where the data is organized by sectors and themes and provides updated time-series as well as multiple estimates from various sources, disaggregated data down to district, Ward, Village level, by sex and urban/rural strata wherever these are available. TSED also allows for user-friendly analysis, through table, graph and map options and provides an opportunity to calculate composite indices.	http://www.tsed.org/home.aspx	Public	No
The Penn State Event Data Project (Formerly Kansas Event Data System (KEDS))	Penn State	Violence/Security Military Operations/ Activities/ Capabilities	Worldwide		Nominal or ordinal codes recording the interactions between international actors), project papers.	http://eventdata.psu.edu/data.html	Public	No
World Poll	Gallup	Public Perception / Atmospherics	Worldwide		Collects basic data on law and order, food and shelter, institutions and infrastructure, jobs, well-being and brain gain. Additional Africa-oriented survey.		Dataset file available at cost	No
Global Land Grabs Dataset	GRAIN	Land Ownership/Real Estate	Worldwide	.xls	This dataset documents 416 recent, large-scale land grabs by foreign investors for the production of food crops. The cases cover nearly 35 million hectares of land in 66 countries.	http://www.grain.org/article/entries/4479-grain-releases-data-set-with-over-400-global-land-grabs	Public	No
Findings on the Worst Forms of Child Labor	U.S. Department of State	Black Markets/Trafficking (Drug/Arms/Human)/ Prostitution	Worldwide	PDF	The report covers 144 countries and territories and provides information on worst forms of child labor in goods and services, as well as information on country efforts, including laws, enforcement, policies and programs. In 2009, the report began to offer findings and recommended actions to prevent or reduce the worst forms of child labor.	http://www.dol.gov/ilab/programs/ocft/tda.htm	Public	No

B-7

Title	Recognized authority	Data Category	Scope of geographic coverage	Data source format(s)	Description of data	Data storage location	Limitations on Access to the data source	In Data-cards?
Gateway to Land and Water Information	Asia-Pacific Network on Integrated Plant Nutrient Management (APIPNM)	Terrain Natural Resources/Energy/ Sanitation/Water	Worldwide		The purpose of Gateway is to provide information on the state, condition and trends of rural land and water resources at national, regional and global scales. This network essentially consists of land and water "reports". Reports are compiled at different levels (global, regional, national, and sub-national) and are linked together as part of a global system of information on the Internet. Gateway acts as clearinghouse and facilitator, providing a useful entry point for land and water information, including FAO's databases on Soil, Agro-ecological Zone, Fertilizer and Water/Irrigation (AQUASTAT).	http://www.apipnm.org/swlwp nr/reports/rc_codes.htm#sf	Public	No
Internet Data Sources for Social Scientists	Data Warehouse / Repository	Portal for social science data	Worldwide		The CISER Data Archive maintains a collection of social and economic datasets, about 27,000 online files and thousands of studies on CD-ROMs and DVDs. It's a centralized source for numeric data files: their acquisition, storage, maintenance, and use. We support the research activities of social science faculty, students, and staff at Cornell University. The collection includes federal or state censuses, files based on administrative records, public opinion surveys, economic and social data from national and international organizations, and studies compiled by individual researchers.	www.ciser.cornell.edu/ASPs/d atasource.asp	Public	No
Crowdvoice	Crowdsourced	Civil Society/Non-Governmental Organizations Civil Strife/Protests Migration/Immigration/ Demographics/Populati on Children/Youth Other	Worldwide	Website	CrowdVoice is a user-powered service that tracks voices of protest from around the world by crowdsourcing information.	www.crowdvoice.org/	Public	No

B-8

Title	Recognized authority	Data Category	Scope of geographic coverage	Data source format(s)	Description of data	Data storage location	Limita-tions on Access to the data source	In Data-cards?
Database of World Rivers and their Sediment Yields	Aquastat (Food and Agricultural Organization, UN)	Agriculture/Food Security Natural Resources/Energy/Sani tation/Water	Worldwide	.xls	This database contains data on annual sediment yields in worldwide rivers and reservoirs, searchable by river, country and continent. The database was compiled from different sources by HR Wallingford, UK, on behalf of the FAO Land and Water Development Division.	http://www.fao.org/nr/water/aq uastat/sediment/index.asp	Public	No
Earth Trends	World Resources Institute	Agriculture/Food Security Natural Resources/Energy/Sani tation/Water	Worldwide		EarthTrends is a comprehensive online database, maintained by the World Resources Institute, that focuses on the environmental, social, and economic trends that shape our world	www.wri.org/publications/data -sets	Public	No
Marine Corps Country Handbooks	Marine Corps Intelligence Activity (MCIA)	Governance/State-Society Relations Religion/Culture/ Language/Ethnicity Military Operations/Activities/ Capabilities Economics/Banking/ Finance Migration/Immigration/ Demographics/ Population	Worldwide	Website	Handbooks provide basic reference information on counties, including geography, history, government, military forces, and communications and transportation networks. This information is intended to familiarize military personnel with local customs and area knowledge to assist them during their assignments overseas.	https://www.intelink.gov/mcia/ handbook.htm	USG and contrac-tors	No
Quality of Elections Dataset (QED)	Duke University	Politics/Voting/ Elections	Worldwide	SAS, SPSS, Stata, ASCII.	This data set codes the quality of national level legislative and presidential elections in 172 countries from 1978 to 2004. The overall election quality variables as well as the seven categories of violence and irregularities in the pre-election period and on Election Day are the same categories as were also coded in the DIEM data set.	http://sites.duke.edu/kelley/da ta/	Public	No
The Post-Internal War Accommoda-tion and Repression (PIWAR) Data Project Page	College of Wooster	Violence/Security Military Operations/Activities/ Capabilities	Worldwide	.xls	Provides data on accommodation, repression, opposition activity, ethnic fractionalization and other conflict data.	http://www3.wooster.edu/polis ci/mkrain/PIWAR.html	Public	No

Title	Recognized authority	Data Category	Scope of geographic coverage	Data source format(s)	Description of data	Data storage location	Limitations on Access to the data source	In Data-cards?
Electoral Commission of Zambia	Government of Zambia	Politics/Voting/ Elections	Zambia	PDF	Zambian election results including: • Public Notice - Presidential Election Result • Public Notice - National Assembly Elections Results • Summary - Allocation of National Assembly Seats • List of Elected MPs • 2011 LGE Results • 2011 Summary of Allocation of Local Government Seats • 2011 Certified Register of Voters	http://www.elections.org.zm/index.php	Public	No
National Water Supply and Sanitation Council	Government of Zambia	Natural Resources/Energy/Sanitation/Water	Zambia	PDF	Research-based reports on the Zambian water sector	http://www.nwasco.org.zm/index.php	Public	No
Zimbabwe Administrative Boundaries	Government of Zimbabwe	Other (administrative boundaries)	Zimbabwe	ArcInfo	Zimbabwe Administrative Boundaries	http://www.grida.no/prog/global/cgiar/dtlf/GNVd1129.htm	Upon request	No
Natural (Agro-Ecological) regions of Zimbabwe	Government of Zimbabwe	Natural Resources/Energy/Sanitation/Water Weather/Climate/Environment/Wildlife	Zimbabwe	ArcInfo	Natural (Agro-Ecological) regions of Zimbabwe	http://www.grida.no/prog/global/cgiar/dtlf/GNVd1128.htm	Upon request	No
Armed Conflict Location and Event Dataset (ACLED)	Trinity College	Violence/Security	50 developing countries	.xls	These data contain information on: • Dates and locations of conflict events, • Specific types of events including battles, civilian killings, riots, protests and recruitment activities • Events by a range of actors, including rebels, governments, militias, armed groups, protesters and civilians; • Changes in territorial control • Fatalities	www.acleddata.com/trends	Public	Yes

Title	Recognized authority	Data Category	Scope of geographic coverage	Data source format(s)	Description of data	Data storage location	Limitations on Access to the data source	In Data-cards?
Afrobarometer	Afrobarometer	Public Perception / Atmospherics	Africa	Spreadsheet Web-based Statistical software	The Afrobarometer is an independent, nonpartisan research project that measures the social, political, and economic atmosphere in Africa. Afrobarometer surveys are conducted in more than a dozen African countries and are repeated on a regular cycle. Because the instrument asks a standard set of questions, countries can be systematically compared. Trends in public attitudes are tracked over time.	http://www.afrobarometer.org	Public	Yes
Climate Change and African Political Stability (CCAPS) Mapping Tool	University of Texas	Weather/Climate/ Environmental/Wildlife Other	Africa		The CCAPS mapping tool provides an online data portal to enable researchers and policymakers to visualize data on climate change vulnerability, conflict, and aid, and to analyze how these issues intersect in Africa. The mapping tool, which uses Esri technology, allows users to select and layer any combination of CCAPS data onto one map to assess how myriad climate change impacts and responses intersect. By integrating the various lines of CCAPS research, as well as other existing datasets, the CCAPS mapping tool aims to provide the most comprehensive view yet of climate change and security in Africa. CCAPS and its partner AidData will release thematic mapping tools throughout the spring and summer of 2012. Climate Change Data, Disaster Response Data. All data collected by CAPPS will be made available in the public domain, including qualitative case studies that collect governance data on government structures, political processes, budget data, and disaster infrastructure to assess the capacity and resilience of African states. SCAD will be added to this mapping tool in spring 2012	http://ccaps.aiddata.org/	Public	Yes

B-11

Title	Recognized authority	Data Category	Scope of geographic coverage	Data source format(s)	Description of data	Data storage location	Limitations on Access to the data source	In Data-cards?
EIS-AFRICA	The Environmental Information Systems (EIS) Program	Terrain	Africa	.PDF, .xls	EIS-AFRICA is a pan-African organization developing African capacity to generate, manage, disseminate and use geo-spatial and environmental information to enrich policy debate and support decision-making.	http://www.eis-africa.org	Public	Yes
Map Library	Map Maker Trust	Terrain Imagery	Africa	Map Maker DRA, ESRI Shape file, MapInfo MIF, or GoogleEarth KMZ. Lat/Longs in TIFF.	The Map Library is a source of public domain basic map data concerning administrative boundaries in developing countries.	www.maplibrary.org/stacks/Africa/index.php	Public	Yes
Famine Early Warning Systems Network (FEWSNET)	U.S. Agency for International Development United States Geologic Service (USGS)	Humanitarian Assistance Natural Resources/Energy/Sanitation/Water Natural Disasters Weather/Climate/Environment/Wildlife Imagery	Africa	.shp	The U.S. Agency for International Development (USAID) Famine Early Warning Systems Network (FEWS NET) is an information system designed to identify problems in the food supply system that potentially lead to famine or other food-insecure conditions in sub-Saharan Africa, Afghanistan, Central America, and Haiti. The USGS FEWS NET Data Portal provides access to geo-spatial data, satellite image products, and derived data products in support of FEWS NET monitoring needs throughout the world. This portal is provided by the USGS FEWS NET Project, part of the Early Warning and Environmental Monitoring Program at the USGS Earth Resources Observation and Science (EROS) Center.	www.fews.net/Pages/default.aspx and http://earlywarning.usgs.gov/fews/africa/index.php	Public	Yes

Title	Recognized authority	Data Category	Scope of geographic coverage	Data source format(s)	Description of data	Data storage location	Limitations on Access to the data source	In Datacards?
African Centre of Meteorological Application for Development (ACMAD)	African Centre of Meteorological Application for Development (ACMAD)	Weather/Climate/Envir onment/Wildlife Imagery	Africa	Satellite imagery, Weather maps	ACMAD's mission is the provision of weather and climate information and for the promotion of sustainable development of Africa (notably within the context of national strategies for poverty eradication), in the fields of agriculture, water resources, health, public safety and renewable energy. ACMAD carries out its mission through; capacity-building for the 53 National Meteorological Services (NMSs) of its Member States, in weather prediction, climate monitoring (extreme events…), transfer of technology (telecommunications, computing and rural communication) and in research.	www.acmad.net/index_en.ph p	Public	Yes
Social Conflict in Africa Database (SCAD)	University of North Texas and the College of William and Mary	Violence/Security	Africa	.xls	SCAD includes protests, riots, strikes, inter-communal conflict, government violence against civilians, and other forms of social conflict not systematically tracked in other conflict datasets. SCAD currently includes information on over 7,900 social conflict events from 1990 to 2011.	http://strausscenter.org/scad. html	Registrat ion Required	Yes
Development Workshop Publications	Development Workshop Angola	Violence/Security Governance/State-Society Relations Development Programs/ Reconstruction Education/Training Politics/Voting/ Elections Land Ownership /Real Estate Civil Society/ Non-Governmental Organizations	Angola	PDF	Development Workshop has been engaged in ongoing research in a number of areas. These include such key issues as community survival and coping mechanisms, land tenure (helped to found the Rede de Terra - Land Network), gender and livelihoods in the informal sector, community services access and peace and conflict risks. DW has acquired a strong capacity in rapid community appraisal techniques and is one of the first institutions in Angola to build a geographic information system (GIS) into their monitoring strategies	http://dw.angonet.org	Public	Yes

Title	Recognized authority	Data Category	Scope of geographic coverage	Data source format(s)	Description of data	Data storage location	Limita-tions on Access to the data source	In Data-cards?
Botswana National Atlas	Department of Surveys and Mapping, Government of Botswana	Criminal Justice Development Programs/ Reconstruction Human Rights/ Gender Equality Economics/ Banking/Finance Terrain	Botswana	PDF	A geographical description of Botswana through sets of thematic maps complemented by explanatory text, statistical information in graphical form and pictures and photographs	http://www.atlas.gov.bw/	Public	Yes
CARPE Data Explorer	Central Africa Regional Program for the Environment (CARPE) (USAID program)	Agriculture/Food Security Natural Resources/Energy/ Sanitation/Water Terrain Imagery Data Warehouse/Repository	Burundi Cameroon Central African Republic Democratic Republic of the Congo Gabon Guinea Republic of the Congo Rwanda Sao Tome and Principe	Interactive maps, GIS datasets, Satellite Imagery	CARPE and its collaborators are hoping to increase the availability of digital maps and data for the Central Africa region, bring spatial information into policy planning and decision-making, provide a tool for training and education, and raise the quality of presentation and communication of environmental information. The data should be used in conjunction with higher resolution data where available, especially for analyses at the local level.	http://congo.iluci.org:8080/ge onetwork/srv/en/main.home	Public	Yes
Multipurpose Africover Databases on Environmental resources (MADE)	Food and Agricultural Organization, UN	Infrastructure/Construct ion/Technology Weather/Climate/Envir onment/Wildlife Terrain	Burundi DR Congo Egypt Eritrea Kenya Rwanda Somalia Sudan Tanzania Uganda	.shp	MADE is the baseline information for GIS users looking for environmental data. MADE is a a set of detailed/homogeneous land cover and environmental information that can be used by a large community of specific end-users. This database is functional to the new database management trends, reduces costs and improves efficiency at local, national and regional levels.	http://www.africover.org	Public	Yes
HIU Maps and analytical products	Humanitarian Information Unit (HIU), Department of State	Terrain Imagery Religion/Culture/Langu age/Ethnicity	Focus on Sudan and DRC. Other relevant countries	PDF, .csv, .xls, .shp.	HIU uses secondary data from various sources (UNOCHA, UNHCR, UNPKO, UN Habitat, FAO, Fewsnet, etc) to produce maps and written analytical products. HIU is an inter-agency office providing U.S. Government officials and other aid organizations with geographic data and analysis to prepare for and respond to complex humanitarian emergencies worldwide.	https://hiu.state.gov/Pages/Ho me.aspx	Public	Yes

Title	Recognized authority	Data Category	Scope of geographic coverage	Data source format(s)	Description of data	Data storage location	Limitations on Access to the data source	In Data-cards?
Malawi Socio Economic Database	National Statistics Office, Government of Malawi	Migration/Immigration/ Demographics/Populati on Labor/Employment	Malawi		The Malawi Socio-Economic Database (MASEDA) is the first comprehensive and up-to-date socio-economic database on the situation of human development in Malawi for use by government institutions, the donor community and civil society counterparts. MASEDA was created by UNICEF and the National Statistics Office (NSO) in collaboration with Malawi's development partners. It was developed to enable the civil society organisations, international organisation, government departments, academic institutions as well as policy makers to have access to the information for their own purposes pertaining to socio-economic development in Malawi.	http://www.maseda.mw/	Public	Yes
Air Pollution Information Network Africa (APINA)	Institute of Environmental Studies (IES) at the University of Zimbabwe	Weather/Climate/ Environment/Wildlife	Primarily Southern Africa; some APINA activities span Africa		The main role of APINA is to form a strong link between the air pollution scientific community and policy makers at national and regional levels. It acts as a conduit of knowledge and data derived in the scientific programmes and existing research to influence policy and decision-makers in matters related to air pollution. APINA acts as a link between different networks and programmes on air pollution in Africa.	http://www.sei-international.org/rapidc/apina.htm	Upon request	Yes

Title	Recognized authority	Data Category	Scope of geographic coverage	Data source format(s)	Description of data	Data storage location	Limita-tions on Access to the data source	In Data-cards?
Agricultural Geo-Referenced Information System (AGIS)	South African government	Terrain Imagery Weather/ Climate/Environment/ Wildlife	South Africa		The Agricultural Geo-referenced Information System (AGIS) strives to offer a one-stop information service for the agricultural sector in South Africa. Using interactive WEB-based applications, AGIS provides access to spatial information (maps), industry specific information and decision support tools. AGIS currently provides access to the following sets of maps: • Orientation information including the 1:250 000 and 1:50 000 topo-cadastral maps, location of towns, roads, rivers, and administrative areas, as well as farm boundaries. • The natural resources atlas includes soils, natural vegetation and climate information, as well as land capability on a national scale. • Demographic information includes the national censuses as well as lifestyle segmentation data in a spatial context. • Foot and mouth outbreak areas indicate the areas affected by recent foot and mouth outbreaks.	http://www.agis.agric.za/agis web/agis.html	Public	Yes
Cultural Mapping for Military Training, Planning and Operations (CMAP)	Army Geo-spatial Center, US Army Corps of Engineers	Terrain Imagery Religion/Culture/ Language/Ethnicity	Worldwide	geoPDF	Data compiled from various sources. Baseline layers of data in geoPDFs include: Ethnicity/Tribe/Clan distribution, languages, belief systems, formal boundaries, infrastructure, natural environment, land usage, resource development activities, imagery, demographics. Analysis support layers include: conflict areas, refugee/IDP camp locations, humanitarian activities, water rights, religious structures/sacred sites, traditional trade routes, caves, neighboring country activities. Products also include: Engineering Route Studies, Urban Tactical Planner, Water Resources.	https://cac.agc.army.mil/produ cts/cultural_mapping/	USG and contracto rs	Yes

B-16

Title	Recognized authority	Data Category	Scope of geographic coverage	Data source format(s)	Description of data	Data storage location	Limitations on Access to the data source	In Data-cards?
AidData	AidData (Brigham Young University, the College of William and Mary, and Development Gateway) World Bank	Development Programs/Reconstruction	Worldwide		Geo-coded foreign aid funding flows from donor nations to recipient countries. Includes the year, donor, recipient, title of project, financial commitment, purpose, and source of aiddata.	http://www.aiddata.org/content/index/data-search	Public	Yes
Human Rights Reports	U.S. Department of State	Human Rights/Gender Equality	Worldwide	PDF	Cover internationally recognized individual, civil, political, and worker rights as set forth in the Universal Declaration of Human Rights.	http://www.state.gov/j/drl/rls/hrrpt/	Public	Yes
Trafficking in Persons Reports	U.S. Department of State	Black Markets/Trafficking (Drug/Arms/Human)/Prostitution	Worldwide	PDF	Assesses governments on their efforts to combat trafficking in persons as defined in the UN Trafficking Protocol, including conduct involved in forced labor and trafficking of adults and children for commercial sexual exploitation.	http://www.state.gov/j/tip/rls/tiprpt/	Public	Yes
International Religious Freedom	U.S. Department of State	Religion/Culture/Language/Ethnicity	Worldwide	PDF	Describe the status of religious freedom, government policies violating religious belief and practices of groups, religious denominations and individuals, and U.S. policies promoting religious freedom.	http://www.state.gov/j/drl/rls/irf/	Public	Yes
Language Resources	Ethnologue (a property of SIL International, originally the Summer Institute of Linguistics, Inc.)	Religion/Culture/Language/Ethnicity	Worldwide	Charts/figures	An encyclopedic reference work cataloging all of the world's 6,909 known living languages. Contains statistical summaries, country indices, bibliographies, and information on endangered languages.	www.ethnologue.com/home.asp	Public	Yes
Mapping for Results	World Bank	Terrain Development Programs/Reconstruction	Worldwide	.csv, .shp	Data on sector distribution, list of financed activities, and number/volume of active financed activities of Bank financed activities under all 12 "lending" product lines.	www.maps.worldbank.org	Public	Yes
MapStory	MapStory	Terrain Other	Worldwide		Collection of animated maps depicting selected time series datasets. MapStory is a social media platform that enables a global community of experts to crowd source socio-cultural data within a geospatial and temporal framework.	www.mapstory.org	Registration required	Yes

B-17

Title	Recognized authority	Data Category	Scope of geographic coverage	Data source format(s)	Description of data	Data storage location	Limita-tions on Access to the data source	In Data-cards?
GeoNetwork	Food and Agriculture Organization (FAO), United Nations	Terrain Development Programs/ Reconstruction Data Warehouse/ Repository	Worldwide	.shp, .lyr, .avl, xls, .csv, .gmz, Rdata, ESRI personal geodatabase, ESRI file geodatabase	The FAO GeoNetwork provides Internet access to interactive maps, satellite imagery and related spatial databases maintained by FAO and its partners.	http://www.fao.org/geonetwork/srv/en/main.home	Public	Yes
ReliefWeb	Office for the Coordination of Humanitarian Affairs (OCHA) United Nations	Development Programs/ Reconstruction Humanitarian Assistance	Worldwide	Text, maps	Latest situation updates, data on complex emergencies and natural disasters by country, maps, financial tracking database for complex emergencies and natural disasters by year.	www.reliefweb.int/countries	Public	Yes
Environmental Data Explorer	United Nations Environment Program (UNEP)	Agriculture/ Food Security Natural Resources/Energy/ Sanitation/Water Natural Disasters Weather/Climate/ Environment/Wildlife Terrain Imagery	Worldwide	.xls	The Environmental Data Explorer is the authoritative source for data sets used by UNEP and its partners in the Global Environment Outlook (GEO) report and other integrated environment assessments. Its online database holds more than 500 different variables, as national, subregional, regional and global statistics or as geospatial data sets (maps), covering themes like Freshwater, Population, Forests, Emissions, Climate, Disasters, Health and GDP. Display them on-the-fly as maps, graphs, data tables or download the data in different formats.	www.geodata.grid.unep.ch/#	Public	Yes
Measure DHS	U.S. Agency for International Development	Health/Nutrition/ Disease Migration/Immigration/ Demographics/ Population	Worldwide	STATA System file, Flat data, SAS System file, SPSS System file.	The Demographic and Health Surveys (DHS) program has collected, analyzed, and disseminated accurate and representative data on population, health service provision, malaria, HIV, and nutrition through more than 260 surveys in over 90 countries.	www.measuredhs.com/	Registration required	Yes

B-18

Title	Recognized authority	Data Category	Scope of geographic coverage	Data source format(s)	Description of data	Data storage location	Limitations on Access to the data source	In Data-cards?
SocioEconomic Data and Applications Center (SEDAC)	SocioEco-nomic Data and Applications Center (SEDAC), Center for International Earth Science Information Network (CIESIN), Earth Institute, Columbia University	Economics/Banking/ Finance Natural Resources/Energy/ Sanitation/Water Civil Society/ Non-Governmental Organizations Migration/Immigration/ Demographics/ Population Terrain	Worldwide		Data on Agriculture, climate, conservation, governance, hazards, health, infrastructure, land use, marine and coastal, population, poverty, remote sensing, sustainability, urban, and water	sedac.ciesin.columbia.edu/da ta/sets/browse	Public	Yes
Action Aid International	Action Aid International	Governance/State-Society Relations Humanitarian Assistance Education/Training Health/Nutrition/Diseas e Human Rights/Gender Equality	Worldwide	PDF	Provides written products on food, HIV/AIDS, emergencies, conflicts, governance, women's rights, Climate changes, capacity building for farmers and communities, harvests.	http://www.actionaid.org/publi cations	Public	Yes
AQUASTAT main country database	Aquastat (Food and Agricultural Organization, UN)	Agriculture/Food Security Natural Resources/Energy/Sani tation/Water	Worldwide	.csv	Contains data on over 70 variables, searchable by country or by region per 5-year period.	http://www.fao.org/nr/water/aq uastat/data/query/index.html?l ang=en	Public	Yes
Center for International Forestry Research (CIFOR)	Center for International Forestry Research (CIFOR)	Agriculture/Food Security Natural Resources/Energy/Sani tation/Water	Worldwide	PDF	CIFOR's strategic research is focused on policy issues that will enable more informed, productive, sustainable and equitable decisions about the management and use of forests.	http://www.cifor.cgiar.org	Public	Yes

B-19

Title	Recognized authority	Data Category	Scope of geographic coverage	Data source format(s)	Description of data	Data storage location	Limitations on Access to the data source	In Data-cards?
Political Instability Task Force (PITF)	Center for Global Policy, George Mason University	Violence/Security Governance/State-Society Relations Civil Strife/Protests	Worldwide	.xls	PITF has compiled annual information on each of four types of political instability events for all countries with a total population of 500,000 or greater, covering the period 1955 to the most current year; these events include ethnic wars, revolutionary wars, genocides and politicides, and adverse regime changes (transitions to more open and democratic forms of governance are not considered instability events). Case lists of each of these four event types are recorded in separate Excel spreadsheets. Each annual record for each event includes three measures of magnitude and a composite magnitude score (cases of genocide and politicide have only a single measure of magnitude for each annual record). For purposes of general instability analyses these four event types are consolidated into periods of (complex) instability when distinct instability events overlap in time or occur in sequence. Sequential events are treated as a single episode if less than five years elapsed between the end of one event and the start of the next event.	http://globalpolicy.gmu.edu/political-instability-task-force-home/pitf-problem-set-annual-data/	Public	Yes
Correlates of War (COW)	University of Michigan	Violence/Security Military Operations/Activities/Capabilities Economics/Banking/Finance	Worldwide	.csv	Cross-national and cross-time data sets, codebook and listings of civil wars , Interstate and extrastate wars and links to other quantitative international relations data.	http://www.correlatesofwar.org/datasets.htm	Public	Yes
Minorities at Risk	University of Maryland	Violence/Security, Human Rights/ Gender Equality Religion/Culture/ Language/Ethnicity Civil Society/ Non-Governmental Organizations Civil Strife/Protests	Worldwide	.sav, .xls	The MAR project currently maintains data on 283 politically active ethnic groups. The centerpiece of the project is a dataset that tracks groups on political, economic, and cultural dimensions. The project also maintains analytic summaries of group histories, risk assessments, and group chronologies for each group in the dataset.	http://www.cidcm.umd.edu/mar/data.asp	Public	Yes

B-20

Title	Recognized authority	Data Category	Scope of geographic coverage	Data source format(s)	Description of data	Data storage location	Limitations on Access to the data source	In Data-cards?
Refworld	United Nations High Commissioner for Refugees (UNHCR)	Migration/Immigration/ Demographics/ Population Humanitarian Assistance	Worldwide	Text	Refworld contains a vast collection of reports relating to situations in countries of origin, policy documents and positions, and documents relating to international and national legal frameworks.	www.refworld.org	Public	Yes
Institute for Study of Violent Groups (ISVG) Relational Database	Institute for the Study of Violent Groups (ISVG)	Violence/Security	Worldwide Given constraints imposed by limited resources, daily "skims" of high- and mid-profile stories are included. Deep-dive collection performed on Mexico and Thailand (not Africa).	.csv, .xls	This relational database captures incidents reported in the open source (mostly through media reports) pertaining to terrorist, extremist, and criminal activities. It captures violent (bombings, armed assaults, arson, WMD attacks, high-jacking, and kidnapping/hostage taking) and non-violent events as well as details relating to the responsible individuals, organizations, and security operations. The ISVG violent extremism and terrorism database currently contains information on over 223, 000 incidents, whereby 43,000 unique individuals have links to over 3,000 distinct groups. The database is geospatially and temporally tagged supporting the production of link association, temporal, statistical and geospatial visualizations. Event data are transformed into the most appropriate format (typically .csv) for integration into the following tools: Palantir Government, i2 Intelligence-led Operations Suite, FMS Sentinel, nFIDeL, Tableau, Omniskope, uReveal.	http://www.isvg.org	Upon request	Yes

REPORT DOCUMENTATION PAGE

1. REPORT DATE *(DD-MM-YYYY)* 28-02-2013	2. REPORT TYPE IAD Final	3. DATES COVERED *(From - To)* 01-10-2011 – 31-01-2013

4. TITLE AND SUBTITLE A Qualitative Data Collection Strategy for Africa	5a. CONTRACT NUMBER DASW01-04-C-0003
	5b. GRANT NUMBER — — —
	5c. PROGRAM ELEMENT NUMBER — — —

6. AUTHOR(S) Ashley N. Bybee, Dominick E. Wright	5d. PROJECT NUMBER — — —
	5e. TASK NUMBER AI-55-3061
	5f. WORK UNIT NUMBER — — —

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Institute for Defense Analyses 4850 Mark Center Drive Alexandria, Virginia 22311-1882	8. PERFORMING ORGANIZATION REPORT NUMBER P-4942 H 2012-001540

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Glenn A. Fogg Director, Rapid Reaction Technology Office (703) 746-1343 2231 Crystal Drive, Suite 900 Arlington, VA 22202	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) — — —

12. DISTRIBUTION / AVAILABILITY STATEMENT

   Approved for public release; distribution is unlimited (18 March 2013).

13. SUPPLEMENTARY NOTES
— — —

14. ABSTRACT

This document presents a qualitative data collection strategy (QDCS) to address gaps in qualitative data that exist in models, simulations, and relevant tools currently being used to analyze the African continent. It is based on the findings of two previous phases of research that identify the most pressing qualitative gaps and the collection of insights from native African researchers to identify the types of data that have the greatest explanatory power in the African context.

15. SUBJECT TERMS
Africa, qualitative, data, social science, models, simulation, data collection

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Ashley Bybee
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified	Unlimited	68	19b. TELEPHONE NUMBER *(include area code)* 703-845-2288