Variables as Currency: Linking Meta-analysis Research and Data Paths in Sciences

[1]**Lynne Davis** lynne@ucar.edu

[1]**Hua Qin** huaqin@ucar.edu

[2]**John D'Ignazio** jadignaz@syr.edu

[1]**Patricia Romero Lankao** prlankao@ucar.edu

[1]**Matthew Mayernik** mayernik@ucar.edu

[3]**Peter Alston** alstonp@edgehill.ac.uk

[1]University Corporation for Atmospheric Research and National Center for Atmospheric Research
[2]Syracuse University, [3]Edge Hill University

ABSTRACT

Meta-analyses are studies that bring together data or results from multiple independent studies to produce new and over-arching findings. Current data curation systems only partially support meta-analytic research. Some important meta-analytic tasks, such as data selection and integration, are not well supported in current data curation systems. To design tools and services that more fully support meta-analyses, we need a better understanding of meta-analytic research. This includes an understanding of both the practices of researchers who perform the analyses and the characteristics of the individual studies that are brought together. In this study, we make an initial contribution to filling this gap by developing a conceptual framework linking meta-analysis with data paths represented in published articles selected for the meta-analysis. The framework focuses on key variables that represent primary/secondary datasets or derived socio-ecological data, contexts of use, and the data transformations that are applied. We introduce the notion of using variables and their metadata as a type of currency that leverage larger bodies of relevant source data. Handling variables in this manner provides an equalizing factor between data from otherwise disparate data-producing communities such as bridging between different kinds of research (e.g., researchers who depend on qualitative versus quantitative methods). We conclude with implications for exploring data integration and synthesis issues and system development.

Keywords

Meta-analysis, Data, Interdisciplinary Science, Concepts, Variables, Urban Vulnerability, Data Curation, Data Integration

INTRODUCTION

A key motivator for open data systems is the possibility that data from disparate studies might be re-purposed, combined, and synthesized to produce new research results that would not be possible by looking at any single study (Interagency Working Group on Digital Data, 2009; Riding the Wave, 2010; Michener, et al., 2011). One important research methodology for bringing together data or results from multiple sources to produce new and over-arching findings is meta-analysis. Meta-analyses gather multiple studies that have a common theme or topic in order to examine patterns across those studies and develop an integrated set of findings (Glass, 1976; Cooper & Hedges, 2009; Rudel, 2008). Integrating and synthesizing data from multiple independent projects into a meta-analysis, however, is a time and energy intensive process. Meta-analyses might take on many different forms, depending on the goals, methods, and kinds of data available, but meta-analyses do show common features. Researchers performing such a study are faced with significant comparability work (Sandelowski, Voils, & Barroso, 2007). They must compare the research topics and methods of the individual studies to be combined, decide which studies should and should not be included, and develop data practices and data structures that enable disparate data or results to be brought together.

Current data curation systems only partially support meta-analytic research. Key tenets of open data systems, such as openly accessible collections, standardized metadata descriptions, and sophisticated search algorithms enable data to be discovered, accessed, and compared (to the extent that metadata descriptions allow). Other meta-analytic tasks, such as data integration and decision making are less well supported in current data curation systems. In part, this is because such tasks rely on human expertise and judgment; different analysts might make different decisions about what to include or exclude from a meta-analysis. At the same time, our understanding of the practices that researchers use to perform meta-analyses is not highly developed. For example, while we know some detail about the elements of scholarly literature used by scientists to synthesize evidence or look for larger effects in a set of results (Blake & Pratt, 2006), the role of data in the researcher's work to establish important decision points needs more attention. To design data curation tools and services that more fully support meta-analyses, we need a better understanding of meta-analytic research. This includes an understanding of both the practices of researchers who perform the analyses and the characteristics of the individual studies brought together.

In this study, we make an initial contribution to filling this gap by developing a conceptual framework linking meta-analytic concepts with the data from individual published articles that are used to derive those concepts. We analyze meta-analyses using the notion of a "data path": the set of stages and linkages where data plays a role, e.g., from its source, through transformations, to presentation in a study.

Our framework emphasizes how key variables represent primary/secondary datasets or derived socio-ecological data, contexts of use, and the transformations that are applied to data. In published articles, the main sources of data for meta-analyses, key variables are typically represented as tables, figures, graphs, etc. Mons, et al., (2011) note that tables and figures are prime components of published articles to be curated as "nanopublications". In addition, Atwood, et al., (2010) and Briscoe, et al., (2011) note how new pdf markup schemes can enable tables and figures to serve as sources of additional information for users and for document retrieval systems respectively.

Managing and publishing the data paths of entities like spreadsheets and graphs have been studied and attempted by information and computer professionals for some time, typically in the context of "workflow" systems (Georgakopoulos, Hornick, & Sheth, 1995; Ludascher, et al., 2006; De Roure, Goble, & Stevens, 2007). A workflow is a precise step-by-step description of a scientific procedure that acts as a script for the coordination of research tasks. Gil (2009) describes how workflows can serve as a currency that enables scientific exchanges. As McPhillips, et al., (2009) note, however, workflows need to be as accessible and usable by researchers who are domain experts as they are for technical experts. Studies of scientific data practices have shown that individuals within most science domains do not use workflow tools and have little experience in sharing their data or processes in structured forms (Borgman, et al., 2007; Cragin, et al., 2010). Our study examines data paths in a meta-analytic research setting where formalized workflow tools are not used, and might be hard to implement because of the variability in research methods and data sources.

We develop the notion of using variables and their metadata as a type of currency to leverage larger bodies of relevant source data. Handling variables in this manner provides an equalizing factor between data from otherwise disparate data-producing communities, such as bridging between researchers who depend on qualitative or quantitative research methods. We conclude with implications for exploring data integration and synthesis issues and system development.

BACKGROUND

Data Conservancy at National Center for Atmospheric Research (NCAR)

Our work was funded as part of the Data Conservancy project. Data Conservancy, funded by the National Science Foundation's DataNet initiative (Sandusky, et al., 2009), is building data curation software services that respond to the needs of particular research communities (http://dataconservancy.org/). Designing data curation services for particular research communities, however, requires developing an understanding of the data practices and needs within those communities. Our work is part of the needs assessment thrust within Data Conservancy. Our aim is to characterize the role of data within a particular research community by simultaneously collaborating with and studying the data practices of a group of interdisciplinary social scientists who study urban vulnerability and resilience to climate change. From this characterization, we can contribute to the development of data curation and preservation infrastructure that enables the conduct of new, interdisciplinary science.

This paper derives from our work with this urban vulnerability (UV) research group at the National Center for Atmospheric Research (NCAR) (Davis, Alston, & D'Ignazio, 2011). For the analysis described in this paper, we are drawing from observations and interviews of NCAR social scientists, from direct personal experience in performing meta-analytic urban vulnerability research, and from examinations of published papers in the literature that relate to urban vulnerability. The model we introduce in the next section illustrates our working understanding of how data and variables are used in this setting.

Urban Vulnerability at NCAR

To understand and contribute to aiding the data practices of any group of scientists, it is necessary to understand their research emphasis and vocabulary (White, 2010). Urban areas are centers of social, economic and technological changes, providing key sources of human adaptation and innovation. Urban areas are also hotspots of environmental pollution and particularly at risk from natural hazards that climate change is expected to exacerbate. These opposing roles highlight the importance of cities in the arena of global climate and environmental change and have spawned an increasing number of studies on urban vulnerability and resilience in recent decades.

Existing UV research is based on conflicting theories and paradigms (Romero Lankao & Qin, 2011) that result in disparate methods and ways of collecting, organizing, and presenting data. The situation poses two sets of interrelated challenges: the first refers to a tenet put forth by some scholars that urban vulnerability is context specific, that is, whether it is possible to identify repeated processes and patterns of vulnerability and resilience across urban areas as diverse as New Orleans in the United States and Lusaka in Africa. The second challenge relates to the fact that different lineages of urban vulnerability research exist and define incompatible methods, concepts and data. The challenge relevant to data curation research and development involves developing a series of tools, techniques, heuristics, and workflows. Such mechanisms would enable database integration across the disciplinary domains involved in the analysis of urban vulnerability and resilience (e.g., atmospheric science, climate modeling, health, sociology, and urban planning.)

The vulnerability of urban areas and populations to hazards of climate change is an example of a complex system (Miller & Page, 2007) with interdependent, interacting variables acting in a changing or adapting environment at a time and place. A variable in this field of study (e.g., average air temperature, age group, and proportion of population living below poverty level) measures a concept (e.g., temperature, age, and poverty) related to a specific vulnerability dimension such as hazard, sensitivity, and adaptive capacity.

A short description of the urban vulnerability model developed by the NCAR research group for their meta-analysis provides the framework for the data-related aspects of urban vulnerability in terms of "variables-as-currency." As illustrated in Figure 1, urban vulnerability to environmental change describes a complex and dynamic reality comprised of several dimensions. Romero Lankao and Qin (2011) developed a single, unified model of these dimensions from their synthesis of the diverse lineages of urban vulnerability research. Urban vulnerability, or the potential for people in urban areas to be negatively impacted by climate change, is a function of: a) hazards, (i.e., probable or looming perturbations and stresses to a system); b) exposure, i.e., the extent to which urban populations are in contact with, or subject to hazards; c) sensitivity, i.e., the degree to which subsets of urban populations are susceptible to hazards with patterns of susceptibility often based on demographic characteristics or medical conditions; and d) adaptive capacity, or the ability to respond to and recover from the negative consequences of climate change based on access to resources, assets and options people draw on to moderate potential damages, to cope with the consequences, or to introduce policy changes to expand the range of variability with which it can cope, see Figure 1.

Each of the dimensions of urban vulnerability has different variables and concepts. For instance, hazards are defined by variables having magnitude and frequency, while sensitivity and adaptive capacity are defined by such concepts as age, preexisting disease conditions, income, dwelling type and quality and access to social networks and health services. These three terms, variable, concept, and dimension, are central to our framework introduced in the next section, and have the following definitions:

- Variable - a measurable phenomenon with a particular unit of measurement (e.g., hazard magnitude)

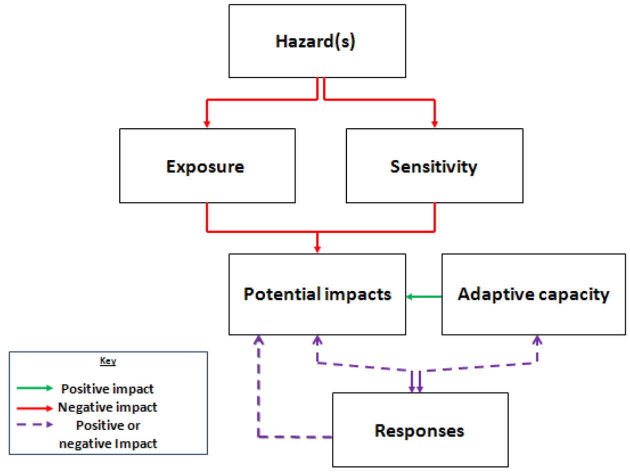

Figure 1: Model of urban vulnerability to global climate change (Romero Lankao and Qin, 2011)

- Concept - a feature of analytic interest that derives from variables (e.g., income)
- Dimension - an analytical construct used to understand and integrate concepts. Our dimensions are: hazards, exposure, sensitivity, adaptive capacity and adaptation

At first glance, age or income would seem to be examples of concepts that are easily quantified as variables, but upon closer inspection, their use within a published research article is qualified in ways that must be considered. To generalize what determines vulnerability within urban areas requires a deeper inspection of how conceptual dimensions or concepts of urban vulnerability are operationalized within the different lineages of this research community. Their meanings need to be comparable and sufficient quantities of data need to be compared so evidentiary patterns can emerge.

The DC effort at NCAR to scrutinize how the UV group examined scholarly articles on urban vulnerability to climate change provided rich examples of how heterogeneous datasets associated with publications might be characterized for recombination/reuse. Such examination can aid information scientists, data librarians and data system designers to know how to capture, curate, and find data which is often heterogeneous, and how to preserve their value for reuse and repurposing. The articles provide summaries and representations of relevant data, and were used as source materials for meta-analysis research. The studies represented by each paper explored relationships between climatic events and their impact on urban populations often in statistical terms. The scientists used variables according to methods that connected data and urban vulnerability dimensions and concepts.

Data related to urban vulnerability to global climate and environmental change originate from both environmental and social science domains, including such information about urban areas as temperature, carbon emissions, mortality rate, population size and density, age structure, gender composition, education attainment, income level, and gross domestic product. Existing sociodemographic, economic, and environmental data on urban areas involve different spatial and temporal scales as well as different characteristics depending on how concepts are defined and respective variables measured. In this interdisciplinary field, primary or secondary datasets related to individual research projects are usually represented in tables and figures within scholarly publications, while the datasets themselves are not included in the publication process. As such, the potential value of these highly processed and embedded datasets for future analysis, synthesis, and preservation in research, cannot easily be actualized, as observed in our study of the UV group's meta-analysis project.

CONCEPTUAL FRAMEWORK

We develop a conceptual framework of data use in urban vulnerability research. The goal is to improve our understanding of the needs of this interdisciplinary research field for the development of data curation and preservation infrastructure that enables the conduct of new science. Figure 2 illustrates the two conceptual linkages hypothesized as important in understanding the relationship between meta-analysis and data paths in scientific practices. One starts with concepts and uses variables as links to the source data, the other starts with the data and uses variables to link to concepts. Variables with contextual information in published articles serve as a bridge between urban vulnerability concepts and their underlying data, and expose primary or secondary datasets and critical data relationships to preserve for reuse. Later sections explain the details in the figure.

Getting from Concepts to Data

To fully understand and be able to compare and contrast the highly concept-based science behind urban vulnerability requires that the concepts themselves be examined in terms

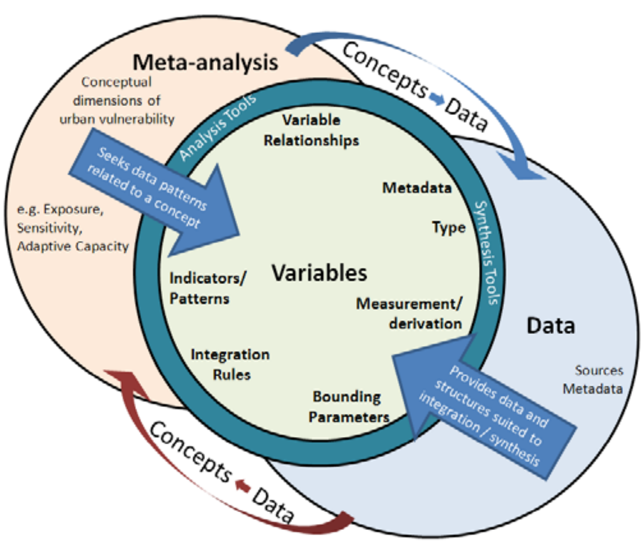

Figure 2: A conceptual framework of data use in urban vulnerability research.

of the data and variables that support them. Concepts of value to the conduct of a meta-analysis in urban vulnerability stem from verifiable variable relationships.

Meta-analyses in social sciences usually start from a concept-based analytic framework which guides the readings of academic publications to focus on data and research findings. Collections of academic articles are the subject of meta-analysis. They are treasure troves of related information qualifying and quantifying data and offer rich summaries of primary/secondary datasets and variable relationships as derived data. As researchers scan articles, they are looking for references to concepts that align with their framework, and could therefore be extracted. Such references could be found anywhere in an article, but arguably, the findings and methods sections may be the most promising. We will focus on those sections because they represent the important juncture useful for both the researcher seeking conceptual references, and for tracing the use of data in support of those references.

Findings of studies give results of a process that realize some relationship between data variables. For example, the result may indicate a positive or negative relationship. Tables, graphs, figures, etc. illustrate relationships between variables; captions and narrative describe the findings. If one were to question the value, validity or definition of a concept reported as a result, one could trace the finding's path "back" to understand the input variables involved, assess their source, and the processes applied to them.

Figure 3 shows data represented in published articles and their importance for meta-analyses. Each of these sections offers something for the subsequent meta-analytic researcher that helps him/her decide about the applicability and value of this work and the potential for reusing its data. The "Data" column shows the types of information in a publication that are of value to meta-analyses. All these points revolve around variables included in a study, and are also useful information about the source primary or secondary data. The next column "Value to Meta-Analysis Researcher" gives clues as to how the extracted "data" can be used in a meta-analysis.

Getting from Data to Concepts

While researchers conducting a meta-analysis may identify more naturally with the concepts-to-data perspective, data managers and data curation professionals may more readily relate to ideas in this section in which we speak to data processing and a mapping between data and concepts. Within urban vulnerability research publications are numerous examples of variables connecting data with urban vulnerability concepts and dimensions. Figure 4 shows a few of them to represent aspects of this data-to-concepts process.

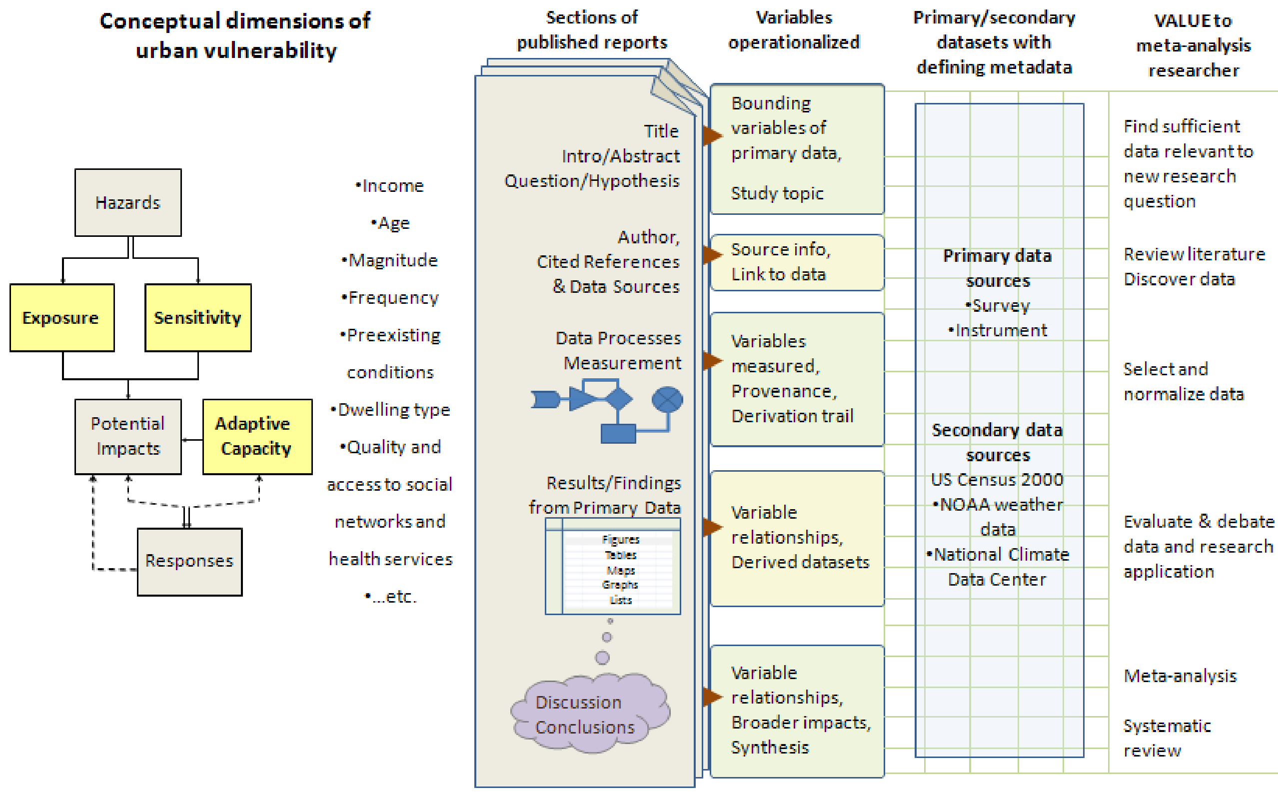

Figure 3: Data represented in articles and their importance for meta-analyses.

There may be a way to improve access to relevant data for meta-analysis by exposing primary/secondary datasets and variable relationships within data curation systems and specifics about them and their formulations. Further, researchers want to know something about how a result was produced before building on those results, so information about its derivation would be valuable. If data repositories included such information about primary/secondary datasets and derived data, search criteria could request articles having variables of specific interest and metadata could include links back to data sources. This would give researchers access to a larger number of relevant articles from which to choose and potentially decrease time spent in the search and selection process.

Figure 4 shows a general data mapping of published articles. Secondary datasets such as NOAA weather data or US Census data often reflect spatial and temporal bounds and identify the unit of analysis. One of the first processes seen in the data path is one that selects those secondary data and/or collects primary data according to parameters that suit the new research. Processes vary widely and can be iterative. They perform calculations that

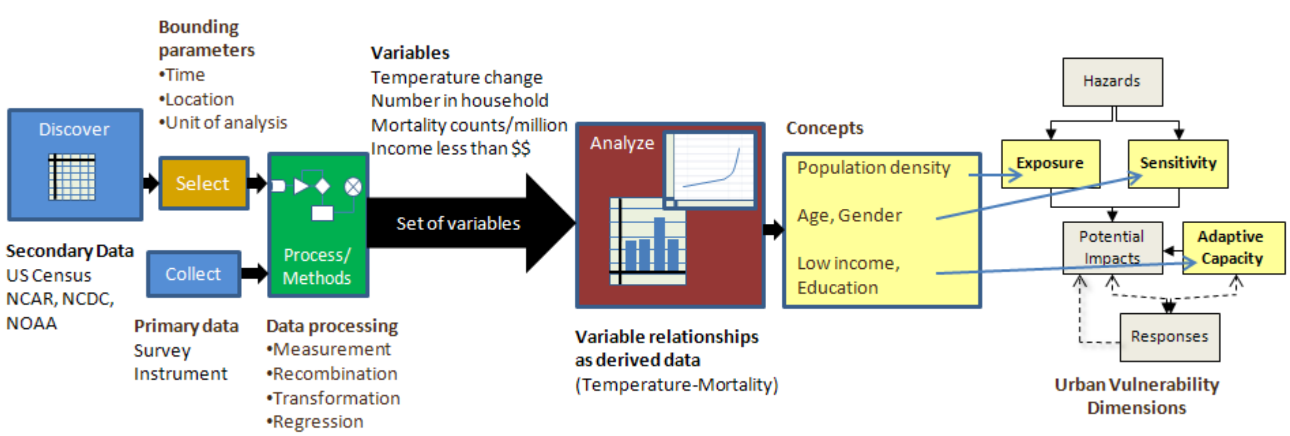

Figure 4: Getting from Data to Concepts: Data and workflows support derived relationships.

prepare data for integration. Having been prepared, it is at this point where variables take on a context of use and their names begin to appear as labels in tables and graphs. Integration and analysis can now be performed on them. Through analysis, variables are interrelated, continue to carry real data substance, and allow scientists to apply them to conceptual models downstream (e.g., Urban Vulnerability Dimensions) in meaningful ways. In this manner, a data path can be discerned and informational aspects of study paths can be captured and used for metadata assignment as part of data curation.

Variables as Currency

In our conceptual model, a variable with associated metadata is a compound component that represents data within a context of use. Data described with units (temperature in degrees Centigrade) but without further context are largely unusable to study more complex issues, like urban vulnerability. At a minimum, researchers need to know the bounding parameters of datasets. Bounding parameters of importance to meta-analyses are unit of observation, spatial scale, time period studied, plus bounds of particular interest to urban vulnerability, such as geographic area involved, hazard type, and impacted urban sector. Further, the way time and place are represented is important when considering their integration. In a published paper, interactions of variables are qualified by the bounding parameters established for the study and are reflected in the derived data presented. Having access to what those data bounds are aids in deciding whether a study can be included in the meta-analysis or how much human effort may be in store to make differing parameters compatible. Figure 4 illustrates a variable within a broad associative structure. Variables (with associated metadata) serve different purposes: some represent what is being measured such as average daily temperature or percentage of population over age 64 or percentage of population with income below poverty level, and some represent a synthesis of variables typically referred to as indices, that result from, for example, a statistical analysis, that shows a composite measure of human thermal comfort. One other group is relational, which serves as a common key between otherwise disparate datasets allowing them to be associated, aggregated, or integrated. However integration happens, it requires that the data to be combined be made homogeneous or relational.

From a data curation perspective, and from a reuse perspective, there are a number of issues related to each of these purposes that when resolved, will help to actualize the potential to leverage previous work in meta-analyses. Some of these are data normalization (e.g., day to week, city to metro area), heterogeneity (inconsistency in name, method, unit, type, source), measurement characteristics (degree of hotness or coldness) and issues surrounding variable relationships (expressing variable correlation, variable associations.) There are also issues of scientific community-related data standards and practices.

Picking out key variables within a published paper can often be simply a matter of looking at the labels, axes, or headings of figures, charts, graphs, tables, etc. Figures, charts, etc., when available within a research article, are frequently where researchers look to clarify and/or evaluate the role of and relationships between variables used in the study. Methods and discussion sections often name data sources and explain the workflow or rules for how those data were acquired, transformed or combined for use in research projects. These derived data thus could serve as a data proxy to the direct use of secondary source data, producing reliable and comparable results with a huge savings in time and effort. Derived data, as reported in these data tables and graphs, is then in a form that can be referenced by researchers to evaluate a study and to inform their meta-analysis. Typically, though, these are not, themselves, accessible in databases, or available in a readily reusable form.

Referring again to Figure 2, the center circle represents variables in their structured sense. It represents data in tables, graphs, figures, etc. using variable names (axes) and relationships, e.g. statistical significance or notable patterns where they may be related to determinants of urban vulnerability. Our model depicts the place of variables acting as a medium of exchange between the value of data and where it is applied conceptually in a meta-analysis. Or, in the reverse direction, selected variables of interest could be used to find and leverage larger bodies of relevant source data, improving representation and significance.

To further illustrate, variables, which by most definitions are something that varies, when combined with contextual metadata become a bounded entity with value for a purpose as illustrated in Figure 5. As such, variables might be thought of as value-laden tokens exchangeable between science domains… i.e., scientific currency. Scientific communities can form their own collections (banks) of them with attributable community-derived rules and standards so they may be transparently exchanged and reinvested as deemed appropriate, conceivably across disparate domains of scientific research. Variables might be

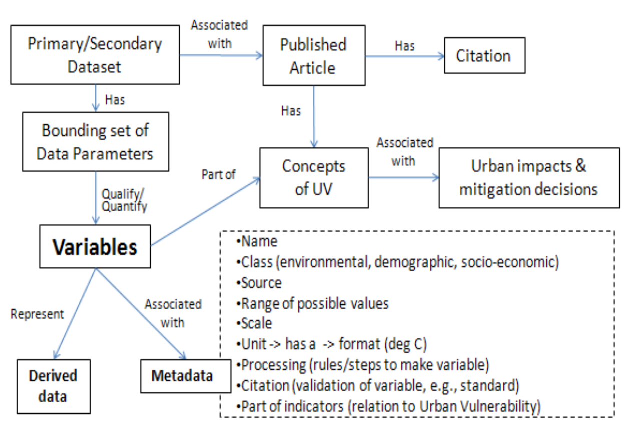

Figure 5: Variables in a structure

associated with source data and published accounts of specific transformations (workflows) to provide a record of provenance. Certain variable associations (derived data) might also become quantifiable metrics for assessing value for a specific purpose.

Another interesting aspect of the concept of using variables as currency in data use resides in the transformation of data variables. If sufficient characteristics and calculation steps of variables can be identified and described in published articles, it becomes feasible to think about automating some of those processes. Results that are exchangeable, meaningful, and trustworthy, are something of great interest to broad research communities. Transformation rules become interesting for curation in association with variables when we think of them as an exchange method for our variable "currency" that might be generic across science domains. Conversely, variables could reflect calculations or treatments applied by certain practices when domain-specific rules are applied. The idea of sharing derived data in its published form as tables, graphs, etc. and having search access to supporting primary/secondary datasets should be very interesting to researchers, provided adequate metadata accompanied them. In such a form, it is conceivable that data from disparate research projects may be made accessible, comparable, and integrable.

As a next step, we want to evaluate this conceptual framework in actual practice to facilitate the conduct of interdisciplinary meta-analyses, particularly those researching the determinants of urban vulnerability to climate change. This includes an in-depth analysis of data paths within a sample of published studies selected for meta-analysis. Doing this for a collection of articles will provide a foundation for better understanding the interrelationships between concepts, variables and data. We aim to test the viability of using contextualized variables to help apply available data to inform new research in an unintended domain.

We are also applying the "variables-as-currency" idea in designing a new tool for the combination of quantitative primary/secondary datasets and derived data that are related to concepts in urban vulnerability research (see Birkland (2011) for details). The primary function of this tool is to synthesize heterogeneous quantitative datasets from individual researchers to produce new datasets based upon specific selection and transformation criteria. Table 1 provides an example of the kinds of inputs and output parameters that may be pertinent to this tool. Data will fall into one of the five dimensions within the conceptual model of urban vulnerability (e.g., hazard, exposure, sensitivity).

Each concept (e.g., temperature, age, income) has its own set of relevant measures (variables). Most of the input data to the quantitative synthesis tool will be tabular in nature, in the excel spreadsheet format. The tool will categorize this data into appropriate concepts and domains (e.g., social, demographic, economic, environmental). With this in place, the tool will allow selection, transformation, and integration (fusion) of the data according to a set of user-supplied parameters (e.g. unit of observation, geographic region). The output of this process will be a novel dataset in the form of a file containing tabular data, which will be suitable for visualization or statistical analysis within a variety of outside tools (e.g., SPSS).

CONCLUSIONS

Meta-analyses require that data and results from multiple individual studies be brought together in combination. Our discussion illustrates how researchers who perform meta-analyses extract variables from individual studies as a type of currency that enable data, results, and concepts to be exchanged and combined. Data curation systems might better facilitate data integration and combination by enabling variable-based ways of comparing and combining individual data sets.

Input				**Output**
Variables	**Concepts**	**Dimension**		**Parameters**
Maximum/minimum/average daytime temperature	Temperature	Hazard		Unit of observation
Population per square mile/kilometer	Population density	Exposure		Geographic region
Years of life	Age	Sensitivity	**Synthesis**	Time period
Annual household income; per capita annual income	Income	Adaptive Capacity		Domains
Daily counts of deaths; number of deaths per 1,000 population per year	Mortality	Impacts		…

Table 1: Examples of input types and output parameters of the proposed quantitative synthesis tool.

Digital data are widely recognized to have great potential for reuse. Cataloging primary or secondary research datasets are already central data curation system functionalities for a wide range of research communities. In emphasizing the centrality of variables in the process of data selection and integration, we suggest that not only the variables, but also the relationship between them, i.e., derived data, become an important candidate for curation as well.

With the new data management planning mandate for NSF grants (NSF, 2011), the need to understand how data and data processes fit within scholarly practices is now being institutionalized at a policy level. As one component of scholarly data practices, what is currently published in the form of tables, figures, charts, maps and the associated data compilations are valuable sources of data for new research, particularly meta-analytic research. We argue that meta-research would benefit more from such data if it were curated as inter-related sets of variables and associated metadata, or curated separately as variable collections that are linked to their author and published work. As a collection, primary/secondary datasets and derived data (tables, figures, etc.) could be annotated as data resources. They also provide metadata through which searches could be performed to uncover variable use in unanticipated domains and publications, increasing the potential for interdisciplinary collaborations and citation.

ACKNOWLEDGEMENTS

This work is supported by the National Science foundation (NSF) HPCC, 9139, 7785, 7726. Any opinions, findings and conclusions, recommendations or omissions expressed are those of the authors and do not necessarily reflect the views of NSF. We gratefully acknowledge the Research Applications Laboratory at NCAR; Mary Marlino, Director, NCAR Library and Integrated Information Systems, UCAR; Sayeed Choudhury, Data Conservancy PI, Johns Hopkins University; and Aaron Birkland, Cornell University, for their contributions to this work.

REFERENCES

Attwood, T. K., Kell, D. B., McDermott, P., Marsh, J., Pettifer, S. R., & Thorne, D. (2010). Utopia documents: linking scholarly literature with research data. *Bioinformatics*, 26(18): i568 –i574. http://dx.doi.org/10.1093/bioinformatics/btq383

Birkland, A. (2011). RS-Cities Proof of Concept #1 Plan and Justification: Initial IRD perspective. http://dlsciences.org/research/DataConservancy/RS-cities_POC_01_plan--ird.pdf

Blake, C., & Pratt, W. (2006). Collaborative information synthesis I: A model of information behaviors of scientists in medicine and public health. *Journal of the American Society for Information Science and Technology*, 57(13): 1740-1749.

Borgman, C.L., Wallis, J.C., Mayernik, M.S., & Pepe, A. (2007). Drowning in data: digital library architecture to support scientists' use of embedded sensor networks. In JCDL '07: *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries.* ACM.

Briscoe, T., et al. (2011). Intelligent Information Access from Scientific Papers. In W. Bruce Croft (Ed.) *Current Challenges in Patent Information Retrieval* (pp. 329-342), Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-642-19231-9_16

Cooper, H, & Hedges, L.V. (2009). Research synthesis as a scientific process. In H.M Cooper, L.V. Hedges, & J.C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 3-16). New York: Russell Sage Foundation.

Cragin, M.H., Palmer, C.L., Carlson, J.R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* 368(1926): 4023-4038. doi:10.1098/rsta.2010.0165

Davis, L.K., Alston, P., & D'Ignazio, J. (2011). Repurposing data across disciplines: a study of data reuse issues between climate science and social science. In *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries (JCDL '11)*. New York, NY: ACM, 433-434. http://doi.acm.org/10.1145/1998076.1998171

De Roure, D., Goble, C., & Stevens, R. (2007). Designing the myExperiment Virtual Research Environment for the Social Sharing of Workflows. In *International Conference on e-Science and Grid Computing* (pp. 603-610). Third IEEE International Conference on e-Science and Grid Computing (e-Science 2007).

Gil, Y. (2009). From Data to Knowledge to Discoveries: Scientific Workflows and Artificial Intelligence. *Scientific Programming*, 17(3). http://dx.doi.org/10.3233/SPR-2009-0261

Glass, G.V. (1976). Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, 5(10): pp. 3-8.

Georgakopoulos, D., Hornick, M., & Sheth, A. (1995). An overview of workflow management: From process modeling to workflow automation infrastructure. *Distributed and Parallel Databases*, 3(2): 119-153. doi:10.1007/BF01277643

Interagency Working Group on Digital Data to the National Science and Technology Council. (2009). *Harnessing the Power of Digital Data for Science and Society*. Networking and Information Technology Research and Development (NITRD) Program. http://www.nitrd.gov/About/Harnessing_Power_Web.pdf.

Ludaescher, B., Lin, K., Bowers, S., Jaeger-Frank, E., Brodaric, B., & Baru, C. (2006). Managing scientific data: From data integration to scientific workflows. In A.K. Sinha (Ed.), *Geoinformatics: data to knowledge* (pp. 109-130). Boulder, CO: Geological Society of America.

McPhillips, T., Bower, S., Zinn, D., & Ludäscher, B. (2009). Scientific workflow design for mere mortals. Future Generation Computer Systems, 25(5): 541–551. http://dx.doi.org/10.1016/j.future.2008.06.013

Michener, W. K., Porter, J., Servilla, M., & Vanderbilt, K. (2011). Long term ecological research and information management. *Ecological Informatics*, 6(1): 13-24. doi:10.1016/j.ecoinf.2010.11.005

Miller, J.H., & Page, S.E. (2007). *Complex adaptive systems : an introduction to computational models of social life*. Princeton, N.J.: Princeton University Press.

Mons, B., van Haagen, H., Chichester, C., den Dunnen, J. T., van Ommen, G., van Mulligen, E., Singh, B., et al. (2011). The value of data. Nature genetics, 43(4), 281–283. http://dx.doi.org/10.1038/ng0411-281

National Science Foundation (NSF). (2011). *Dissemination and sharing of research results*. http://www.nsf.gov/bfa/dias/policy/dmp.jsp

Riding the Wave: How Europe Can Gain From the Rising Tide of Scientific Data. (2010). Final report of the High Level Expert Group on Scientific Data. A submission to the European Commission. European Union. http://ec.europa.eu/information_society/newsroom/cf/document.cfm?action=display&doc_id=707.

Romero Lankao, P., & Qin, H. (2011). Conceptualizing urban vulnerability to global climate and environmental change. *Current Opinion in Environmental Sustainability*, 3(3): 142-149. doi: 10.1016/j.cosust.2010.12.016

Rudel, T. K. (2008). Meta-analyses of case studies: a method for studying regional and global environmental change. *Global Environmental Change*, 18(1): 18–25.

Sandelowski, M., Voils, C. I., & Barroso, J. (2007). Comparability work and the management of difference in research synthesis studies. *Social Science & Medicine*, 64(1): 236-247. doi:10.1016/j.socscimed.2006.08.041

Sandusky, R. J., Palmer, C. L., Allard, S., Cragin, M. H., Cruse, P., Renear, A., & Tenopir, C. (2009). The DataNet partners: Sharing science, linking domains, curating data. In *Proceedings of the American Society for Information Science and Technology*, Vol. 46. doi:10.1002/meet.2009.1450460141

White, H. C. (2010) Considering Personal Organization: Metadata Practices of Scientists. *Journal of Library Metadata*, 10(2/3): 156-172. doi:10.1080/19386389.2010.5063961