To add/combine with the summary that Erica prepared

ESCOP Social Sciences Subcommittee

A Vision for Social Sciences Research Utilizing "Big Data"

At the ESCOP meeting in 2017 the Ag Economics and Rural Sociology Group brainstormed guided by the following questions

1. What does Big Data mean to us?

The questions and discussion raised about with the first question focused on the one had on what are the characteristics of data in the social sciences, and on the other what big data means in terms of being able to access and translate data into meaningful and actionable knowledge; and type of infrastructure is required to process big data, what would be the nature of collaborations among producers of the data and researchers in order to access the data for research, and issues of ethics, equity in access, and who is left out of big data, as well as how data may drive the agenda rather than the problems driving the data that is needed.

The characteristics of data needed are a function of the questions and approaches used in the social sciences, and these are many. It is more often the case in the social sciences that instruments/questionnaires are developed informed by theories to be able to have the variables needed; or generated through experiments. The researchers develop instruments and conduct the surveys, or work with large data sets either accessed through public organizations (ACS, Census Data, NASS, are examples) or purchased from companies that gather the data.

Conducting surveys (Human Subjects research) require IRB training and approvals, and seek to gather information that larger datasets do not gather. What type of infrastructure and skills is required to work with Big Data and how would, universities be able to access it. Concerns were brought up about equity as some universities have more resources than others. Is Big Data public? What should be the "rules" of access so it is equitable. Would NIFA or other agencies support initiatives to ensure that this is the case?

What is big data for the social sciences? Is it the data that we need? If Big Data is the Census for example, what are the mechanisms in place to access the data, and how costly is it? Would research funding using Big Data compete for resources with the research funded by NIFA for example?

<u>Some of the statements on issues raised follow:</u>

There is value in developing partnerships because otherwise some may be left out of the game.

What types of research agendas does pushing forward Big Data encourage?

How relevant is big data to what we do in the social sciences?

There are data infrastructure needs definitely, but why Big data?

We need to be able to sustainably collect data and use it for research.

We also need the geo coding more accessible so that it is connectable.

ACS and the Census of Ag and other data bases already collected by other public agencies. How are these accessed, are there issues, lessons learned?

The nature of social sciences means specific characteristics with ethical dimensions, therefore IRB, identifiers?

Is it about big farmers (who benefit from these type of data) rather than responding to the needs of human beings?

Small institutions matter and should not be left behind.

Equity seems to be a big issue: Is it large for operations? Is sufficient data being generated? Who is left out?

How can we use big data to benefit the economic viability of rural communities? What is the data that can actually accomplish this?

Need to include the community piece – rural communities

IF we can use the data to improve the economic conditions then it would be a plus

Asymmetrical knowledge: private vs public, and small versus large institutions

How do we get information collected to the public, the people that need to know and use it?

How do we get the real data to the people?

2. What "I" like?

Through the discussion of Question 1 we realized that we need small data. We also need complex data that links issues, at multiple scales, and units of measure that could be used to address complexity.

The groups also liked the fact that there is potential to collaborate in the social sciences to define ways of making big data relevant to inclusion.

Some of the question identified that we need to answer are:

- What is the research needed to make this happen?
- What are the research questions that we need to ask?
- What if we can take this big data to demonstrate the gaps and therefore what needs to be done to increase sustainability?

3. What "I" wish?

That the big data that we want is capable of being aggregated, spatially defined, and more accessible so we can use it for analysis in complex systems.
And that the data be in the public domain rather than private, because it is expensive.

4. What if we could answer the following:

- What if we had a central federal data that compiles all the data so it would be accessible?
- Could or can this data address informing policy at the local level?
- How can big data inform sustainability in rural communities?
- How can big data contribute to health, sustainability, food security in rural and poor communities?
- **New layer of goals** that involve people: (See figure 1)
    - Improving Human Beings
    - Strengthening Families
    - Enhancing communities
    - Improving nutrition and health for youth

The group identified additional questions we need to answer:

- How do we prevent from repacking the federal data through the private sector, which ends up being very expensive?
- What type of Data infrastructure is required?
- Assess the differences between conducted surveys, in terms of response rates and completed questions.
- The issue of access and asymmetric knowledge?
- We have to adhere to IRB?
- The usability of big data for smaller institutions (universities): It is a challenge for some of the institutions in terms of access, operability, inconvenience, and you don't know how useable.
- Capacities of small institutions may be a barrier to use

Figure 1. A Layer that must be present in consider Big Data