Data Collection for Seed System Network Analysis

Christopher E. Buddenhagen (cbuddenhagen@ufl.edu), James C. Fulton, Kelsey F. Andersen, Karen A. Garrett (karengarrett@ufl.edu)

Plant Pathology Department, University of Florida, Gainesville, FL 32611-0680, United States

Institute for Sustainable Food Systems, University of Florida, Gainesville, FL 32611-0680, United States

Emerging Pathogens Institute, University of Florida, Gainesville, FL 32611-0680, United States

Abstract

We present survey questions useful for describing agricultural seed systems. The questions are designed so that they can be used for standardized comparisons among seed systems, addressing both networks for seed movement and networks for the communication of information related to variety selection and integrated pest management. This approach provides information that can be used in multilayer network analyses of how information influences seed system success. Also provided are example data sheets with field descriptors that should provide for straightforward statistical analysis after data collection.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2806v2 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

Introduction

The structure of seed system networks can have important effects on the success of the systems for germplasm distribution, conservation, and protection against the spread of seedborne pathogens (Pautasso et al. 2013). Seed systems are often described in terms of some aspects of their structure, without full consideration of network structure. A fuller description of the network is also useful for comparing across seed systems, as part of analyses to understand why systems may be more or less successful. A standardized set of data is needed to do this. Standards for common terminology for describing seed systems have been developed (Almekinders et al. 1994, Sperling et al. 2013), and here we describe common *data* to be collected to describe networks. To some extent, our discussion emphasizes vegetatively-propagated crops such as banana, plantain, cassava, potato, sweetpotato, and yam. However, most of the methods are applicable to any seed system. The basic format for generating the network is to record the movement of seed (vines, tubers, sucker, rhizomes) or products (potatoes, yams, bananas, etc.) to and from the different stakeholders. At the study outset, the researcher should decide if food production needs to be included in the network. For example, in potato seed systems there are networks of tubers produced for seed and those produced for consumption (food products). Harvest sold or used for food and consumption may provide important indicators about livelihoods of farmers.

Impact Network Analysis (INA) is a multilayer network analysis that looks at the effects of a technology – such as a new variety, a management strategy, or other types of information – as it influences the outcomes of linked socioeconomic networks and biophysical networks (Garrett 2012, 2018). The better the understanding of network structure, in terms of the connections (links) between the stakeholders, the more effective INA will be. Stakeholder identity is a useful type of data, and can make estimation of larger network structures more straightforward. Collecting other descriptive variables supports answers to questions related to volume, cost, and distance. The addition of data about farmer decision making and the quality of their information sources is also useful, e.g., loan sources, integrated pest management, and variety selection.

Sampling efforts can vary in their level of complexity, based on access to the stakeholders, the degree of realism desired in the resulting networks, and the extent to which explanatory variables are important, e.g., gender, age or economic status (Lumley 2010). Surveys could be carried out with pen and paper, or with the aid of computers or phone applications (such as open data kit https://opendatakit.org/). The method here has been constructed to be flexible based on resource constraints.

Because the subjects of these studies are humans, special permissions are generally needed to ensure privacy and fair treatment. For example, many institutions have an Institutional Review Board to evaluate proposed surveys, and permission is needed prior to carrying out the study and prior to publishing.

Sampling Strategy

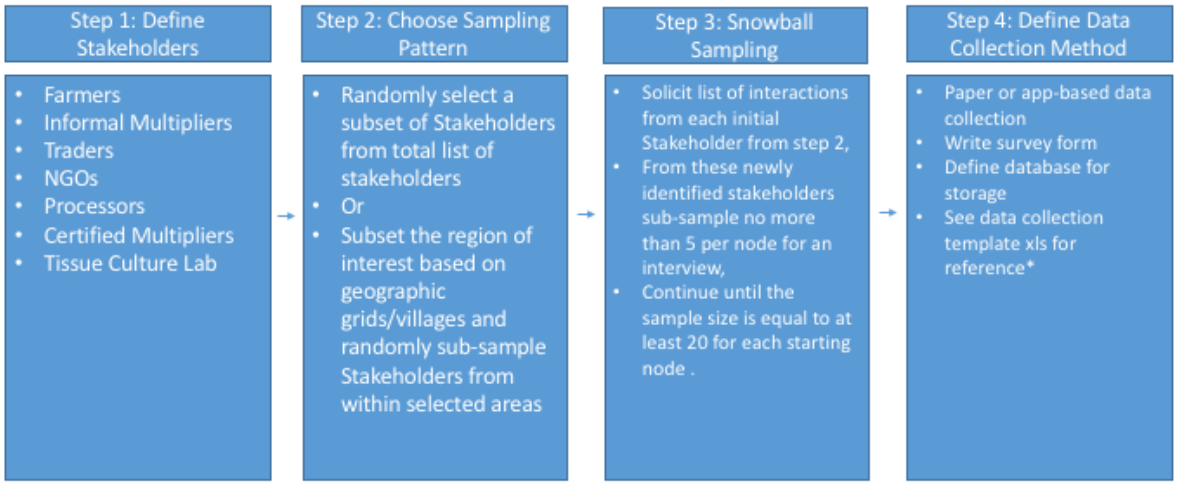
To characterize a seed system in the region of interest, we suggest a stratified sampling method that will provide information about the main stakeholder types that are present. This would

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2806v2 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

include: farmers, multipliers, certified seed providers, NGOs, traders or sellers, and food processors (large buyers of food product). The specificity of relationships to be documented can also be varied. If specific stakeholders are a focus, identification of all the specific person to person transactions is useful. If broad patterns for transactions between types of stakeholders are of interest, questions might focus broadly on the categories of seed sources. Surveys could include transactions related to both seed and food product sales, as the two are interrelated in many systems and both important for livelihood.

Sample strata or regions could be selected based on an objective criterion (e.g., random, stratified, a complete census, snowball sampling, or other) but should be aimed at answering an important question. For example, what is the difference between seed networks in sparsely populated rural areas versus densely populated regions around important population centers? Sampling effort could be scaled to available time and financial resources. Sampling could be carried out by traveling to individual stakeholders to interview them, by using a questionnaire (paper, computer, or application), or by inviting them to a single location for a workshop. The time window of interest should be determined by the researcher (for example, all transactions in the previous 12 months).

The following figure roughly describes the process, for a case where snowball sampling is applied:



A key consideration is to address gender. It may be important that the data reflect actual rates of participation by women and men in a project, for example. Then objective sampling methods should be employed to ensure the probability of sampling women and men will provide information about their relative occurrence. Alternatively, if the research question is specifically gender focused and a gender is under-represented, explicit effort to sample subjects in the rare gender may be necessary. Gender differences in seed use and production of ware potato were detected in an Ecuadorian cooperative (Buddenhagen et al. 2017) but not in a system in Ethiopia, where seed access was most influenced by farmer wealth (Tadesse et al. 2017). There may be interactions among social descriptor variables that need to be considered in the context of gender-focused studies.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2806v2 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

Provided below is an example questionnaire, with some emphasis on snowball sampling strategies. A spreadsheet designed for collecting this data is available upon request from the authors.

Important introductory concepts

In this system for data collection, every unique seed or product transaction or exchange of information will have its own unique entry or row on the spreadsheet, including cases where the farmer uses/consumes his/her own seed or product. If a farmer receives seed from multiple sources, or shares seed/product with multiple stakeholders, then each of these cases gets a separate entry. As much as possible the stakeholders identified should be sufficiently well described that, if two data providers are talking about the same person or entity, this will be clear. This may be difficult to achieve, for example, when seed is obtained at a market from an unfamiliar person. A good strategy in this case is to get other information about the seller's relevant characteristics, or to lump all anonymous transactions from that market. For example, in this case the seed source could be "Gulu Market", or "Anonymous Trader Huancavelica". The more specific (or disaggregated) your data are, the more options there will be for analysis.

Valuable references for further understanding of sampling for network characterization include Borgatti et al. (2013), Frank (2011), and Marsden (2011). Al Hasan et al. (2017) categorized network sampling methods in a very approachable and useful way. When implementing seed network studies, a similar framework can be used to help identify the best sampling method. Their first consideration was the *a priori* information available. When sampling to characterize seed networks, there are likely to be one of two scenarios. In the first, we have a complete sampling frame wherein the individuals of interest in our network are known and their exchanges measured. Or, we may encounter a scenario where the network residents and their exchanges are unknown, but are accessible. In the latter case, we may begin with a single or several known network residents who will serve as initial starting points. Beginning with these individuals, the network can be enlarged by learning of network partners. The second important sampling consideration is whether the objective is to characterize the attributes of individual nodes in the original graph or to sample a representative subgraph of the larger network. The approach taken requires trade-offs between ensuring against certain biases, maintaining consistency of node and network characteristics between the sub-graph with the larger graph, and resource availability.

Acknowledgements

This work was undertaken as part of, and funded by, the CGIAR Research Program on Roots, Tubers and Bananas (RTB) and supported by CGIAR Fund Donors http://www.cgiar.org/about-us/governing-2010-june-2016/cgiar-fund/fund-donors-2/

References

Al Hasan, M., N. Admed, and J. Neville. 2017. Network sampling: methods and applications. https://www.cs.purdue.edu/homes/neville/courses/kdd13-tutorial.html.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2806v2 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

Almekinders, C. J. M., N. P. Louwaars, and G. H. Debruijn. 1994. Local seed systems and their importance for an improved seed supply in developing countries. Euphytica **78**:207-216.

Borgatti, S. P., M. G. Everett, and J. C. Johnson. 2013. Analyzing Social Networks. SAGE Publications.

Buddenhagen, C. E., J. F. Hernandez Nopsa, K. F. Andersen, J. Andrade-Piedra, G. A. Forbes, P. Kromann, S. Thomas-Sharma, P. Useche, and K. A. Garrett. 2017. Epidemic network analysis for mitigation of invasive pathogens in seed systems: Potato in Ecuador. Phytopathology **107**:1209-1218.

Frank, O. 2011. Survey sampling in networks.*in* J. Scott and P. J. Carrington, editors. The SAGE Handbook of Social Network Anaysis. SAGE Publications, London.

Garrett, K. A. 2012. Information networks for plant disease: Commonalities in human management networks and within-plant signaling networks. European Journal of Plant Pathology **133**:75-88.

Garrett, K. A. 2018. Impact Network Analysis: a framework for evaluating the effects of information and other technologies through linked socioeconomic and biophysical networks. bioRxiv.

Lumley, T. 2010. Complex Surveys: A Guide to Analysis Using R. John Wiley & Sons, Hoboken, New Jersey.

Marsden, P. V. 2011. Survey methods for network data.*in* J. Scott and P. J. Harrington, editors. The SAGE Handbook of Social Network Analysis. SAGE Publications, London.

Pautasso, M., G. Aistara, A. Barnaud, S. Caillon, P. Clouvel, O. T. Coomes, M. Delêtre, E. Demeulenaere, P. De Santis, T. Döring, L. Eloy, L. Emperaire, E. Garine, I. Goldringer, D. Jarvis, H. I. Joly, C. Leclerc, S. Louafi, P. Martin, F. Massol, S. McGuire, D. McKey, C. Padoch, C. Soler, M. Thomas, and S. Tramontini. 2013. Seed exchange networks for agrobiodiversity conservation. A review. Agronomy for Sustainable Development **33**:151-175.

Sperling, L., O. Ortiz, and G. Thiele. 2013. Seed Systems. Conceptual frameworks for guiding practical interventions. CGIAR.

Tadesse, Y., C. J. Almekinders, R. P. Schulte, and P. C. Struik. 2017. Tracing the seed: Seed diffusion of improved potato varieties through farmers' networks in Chencha, Ethiopia. Experimental Agriculture **53**:481-496.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2806v2 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

Survey Questionnaire Example

Include consent and participation phrasing here. This should inform the subject that they are free to withdraw from the study at any time. The purpose of the survey should be communicated clearly. Subjects should be reassured that any reports or publications will use anonymized information about the participants, and any of the information recorded will be kept private and in a safe place. Include signing of agreement to participate. This requires Institutional Review Board approval for human subjects.

Date of interview:

Demographic data for the interviewee

Name/Organization:
Gender:
Age/Age Category:
Location:
Stakeholder type: Farmer, NGO, Trader, Multiplier, Store, Market, etc
Identifier Code:

Farmer context

How many fields do you own? How many fields do you manage? What is the total land area in cultivation?

What is the amount of seed you used/planted/sold last season?

Note: These questions set up the interviewer to examine that the farmer information provided is accurate. For example, if a farmer reports two hectares of land but describes seed amounts appropriate for one hectare, the interviewer could ask them to re-estimate.

SEED SURVEY

Define season/time of interest and/or dates for each transaction. It may be reasonable to ask for information for the three most recent seasons, although the quality of information may decline with time.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2806v2 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

For this season/time, from whom did you obtain seed?

	Name/Organization	Type	Gender	Crop	Variety	Volume Units:	Price	Place of Transaction	Transaction Date	Planting Date	Repeat trans.	Quality
1												
2												
3												
4												
5												

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2806v2 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

How much seed from your own farm did you reuse?

	Crop	Variety	Volume Units:	Date	Quality
1					
2					
3					
4					
5					

For this season/time, who did you give or sell seed to?

	Name/Organization	Gender	Type	Crop	Variety	Volume Units:	Price	Place of Transaction	Transaction Date	Repeat trans.	Quality
1											
2											
3											
4											
5											

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2806v2 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

PRODUCT

(For example, ware potato as opposed to seed potato)

For this season/time, from whom did you obtain product?

	Name/Organization	Gender	Type	Crop	Variety	Volume Units:	Price	Place of Transaction	Transaction Date	Repeat trans.	Quality
1											
2											
3											
4											
5											

How much product did you consume?

	Crop	Variety	Volume Units:	Date
1				
2				
3				
4				
5				

For this season/time, who did you give or sell product to?

INFORMATION

	Name/Organization	Gender	Type	Crop	Variety	Volume Units:	Price	Place of Transaction	Transaction Date	Repeat trans.	Quality
1											

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2806v2 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

2											
3											
4											
5											

Information sources related to variety adoption

These questions address information sources that influence variety adoption. Define the time period of interest, where the last three seasons may be a reasonable time interval.

For this season/time, from what sources did you obtain information about variety traits?

	Name/Organization/Source	Type	Gender	Crop	Transaction Date	Info Types *	Trust (0-5, where 5 is highest)
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							

Info Type examples: a) yield, b) nutritional value, c) disease resistance,

	Name/Organization/Source	Type	Gender	Crop	Transaction Date	Info Types *	Trust (0-5, where 5 is highest)

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2806v2 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

1							
2							
3							
4							
5							
6							
7							
8							
9							
10							

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2806v2 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

Information sources for integrated pest management

For this season/time, from what sources did you get information about IPM?

	Name/Organization/Source	Type	Gender	Crop	Transaction Date	Info Types *	Trust (0-5, where 5 is highest)
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							

*Types include: 1 pest and disease recognition, 2 chemical control, 3 crop rotation, 4 roguing, 5 positive selection, 6 planting/harvest time

What diseases and pests do you consider most when making variety selection or IPM decisions (order by importance)?

	Name of disease/pest
1	
2	
3	
4	
5	
6	
7	
8	

In the last season, which diseases and pests most impacted yield/quality of your crop (order by importance)?

	Name of disease/pest
1	
2	
3	
4	
5	
6	
7	
8	

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2806v2 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

9			**9**		
10			**10**		

Yield

What date was your last harvest?

In your opinion, was your yield in the last harvest:

1) higher than expected,
2) as expected
3) lower than expected?

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2806v2 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

Spreadsheet Field Explanations

The following describe the fields from the spreadsheet provided as supplemental material. Tab headers are indicated.

TAB: StakeholderSeedProduct

This tab includes data about the stakeholders in the system, where stakeholders include all system actors. It lists both the interviewee, and the stakeholders that they identify as being sources or destinations for seed and products for the general market (e.g., ware potato).

Stakeholder

The name of the stakeholder. This could be either the person being interviewed, or a different stakeholder identified by the subject as being the source of the product or seed. Its name should allow data from different stakeholder interviews to be matched when they identify the same stakeholder by name.

StakeholderCode

This provides a code for actors in the system that provides an anonymized alternative name for the stakeholders, generally a requirement for publication and data sharing. In some cases it may be acceptable to use the non-anonymized names of institutions.

InterviewedOrNamed

This field identifies whether the stakeholder was the interviewee/survey subject, or was named by the interviewee during the survey. If a stakeholder is both interviewed and named, this is recorded. The question is getting at the amount of detail that is available about that stakeholder. A food processor might be interviewed and supply a list of farmers who supplied a product, but probably only a fraction of the named farmers would end up being interviewed.

Gender

Code for people is M or F. Agencies or other entities could be designated N. If the person at the agency is identified, gender could be indicated.

Age category

Age category of interviewee.

StakeholderType

This could involve a classification system that matches the system being evaluated. This could include farmers, NGOs, multipliers, certified seed providers, research agencies, farmer cooperatives, traders, markets, and companies (Sperling et al. 2013).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2806v2 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

Location

This could involve more than one field, but could be based on any relevant naming system or hierarchy of geographic names, e.g., village, town, districts, counties, etc.

Latitude and Longitude

The latitude and longitude of stakeholder named.

TAB: SeedProductTransactions

This describes the links or ties between the stakeholders in terms of seed or product transactions and collects some relevant descriptive information about those transactions.

DateofInterview

Date that survey was completed.

Interviewee

This is the name of the stakeholder who is the *subject of the interview*. This should be a unique identifier, often the full name should be adequate (although more descriptive information may be necessary). Each Stakeholder should be assigned a unique ID (e.g., if there is Mary Jane from Village A and Mary Jane from village B, each needs a unique ID). *Please note:* the name of the interviewee (person being questioned) will correspond to one of the stakeholder fields, either StakeHolderFrom or StakeHolderTo. There will need to be both From and To information from each interviewee. See explanation for those fields and the example questionnaire.

The goal is that if any two stakeholders receive or send seed from/to the same stakeholder, then the data will be complete enough to identify that they share the connection.

Crop

The crop you are gathering information about.

Variety

This will be the variety name (common and commercial if available), if there are synonyms between names these should be identified.

EndUse

This mostly focuses on whether the product being transacted is intended for use as *seed*, or as a food *product*. There may be more types than seed or product, for example, seed, food, industrial.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2806v2 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

DateofTransaction

Date of the transaction (if known, or rough estimate of time window)

StakeHolderFrom

The name of the stakeholder. This could be either the person being interviewed, or a different stakeholder identified by the subject as being the source of the product or seed. Its name should allow data from different stakeholder interviews to be matched when they identify the same stakeholder by name. For example, if a farmer being interviewed will use their own seed, then StakeHolderFrom and StakeHolderTo will be identical, but if the farmer sources seed from a government agency, then the government agency name would be recorded here.

StakeholderTo

The name of the stakeholder. This could be either the person being interviewed, or a different stakeholder identified by the subject as being the source of the product or seed. Its name should allow data from different stakeholder interviews to be matched when they identify the same stakeholder by name. For example, if you are interviewing a farmer they will and they use their own seed then StakeHolderFrom and StakeHolderTo will be identical. For example, if a farmer being interviewed will use their own seed, then StakeHolderFrom and StakeHolderTo will be identical, but if the farmer sources seed from a government agency, then the farmer name would be recorded here.

Quality

This would be some measure of quality for the seed or product in each transaction. This could be omitted if not available. It could be a simple rating such as high, medium or low, or it could involve common ratings used in the seed system or markets. For seed it might be related to specific certified seed classifications.

Volume

The volume of product (seed or food sold). Units for each will differ. For example, if sweetpotato vines are sold for seed, the number or bushels sold may be important, but if the product is sweetpotatoes sold for food, the units might be kilos. Thus, units will differ for each transaction type.

VolumeUnits

Kilos, vines, suckers, banana stalks, etc. Units will differ within a crop depending on the transaction characteristics, for example sweetpotato vines are sold for seed, but the product is sweetpotatoes sold at the named market units will differ for each transaction type.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2806v2 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

PricePerUnit

This is the amount spent/received for the product expressed in terms of the per unit cost. If the units quintals, for example, this would express the price for one quintal. Currency type needs to be recorded. A zero would be recorded if the product is given away for free or if the seed is reused from a previous season. Data entry should distinguish between free seed and lack of information about price per unit.

Currency

Qualifier of "price"

TransactionDate

This records the date for the transaction in as specific a way as possible, depending on the system. It could be Year, Month, Day, or a combination of Season (e.g., summer) and Year.

Planting/HarvestDate

This could be actual or *anticipated* dates. Both planting and harvest dates could be recorded in each case. This indicates the date for the transaction in as specific a way as possible. It may be useful to record, as a minimum, the planting date for seed and the harvest date for product. It could be Year, Month, Day, or a combination of Season (e.g., summer) and Year.

TransactionLocation

This identifies one of the stakeholders as the location if the transaction occurred there, for example at the farm of one of the stakeholders. This means the location information on the StakeholderSeedProduct tab can be used. Otherwise this is a third location different from the location of the stakeholders. For example, this would record where a trader or farmer travelled to buy or sell a product. So, if Trader 1 sells to Farmer 1 at the Trader's stall in a market then Trader 1 is written here. If Trader 1 travels to a market in another town to sell then the location of the market in the other town would be recorded.

TransactionLat/TransactionLon

This is the location where the seed or crop product is being sold or received, to be specified if it is *not the specific location of either of the main stakeholders*. The stakeholder locations should already be specified on the StakeholderSeedProduct tab. It may be that some of this data is generated after the interview/questionnaire process is completed. For example, if a specific well-known market is frequently named, the location can be generated by the researcher. If it is an obscure location like a stall on the side of a road, it would be worth the effort to get the information during the interview process.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2806v2 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

TAB: InformationSourcesStakeholders

This records data about farmer information sources that may influence farmers' use of the seed system, or the risk of disease spread and ultimately yield. Information sources include people, institutions, publications, or the internet. In the example questionnaire, we focus on variety adoption and information about the management of pests and diseases (IPM).

SourceOrStakeholder

The name of the stakeholder. This could be either the person being interviewed, or a different stakeholder identified by the subject as being the source of the information. The name should allow data from different stakeholder interviews to be matched when they identify the same stakeholder by name. In this case, information sources are treated as nodes in the network, and are also included even if they are not people or institutions involved directly in seed trade. For example, the internet could be an information source for some farmers and would be a node and recorded here.

InterviewedOrNamed

This field identifies whether the stakeholder was the interviewee/survey subject, or was named by the interviewee during the survey. If a stakeholder is both interviewed and named, this is recorded. The question is getting at the amount of detail that is available about that stakeholder. A food processor might be interviewed and supply a list of farmers who supplied a product, but probably only a fraction of the named farmers would end up being interviewed.

Gender

Code for people is M or F. Agencies or other entities could be designated N. If the person at the agency is identified, gender could be indicated.

StakeholderSourceType

This would be used to classify stakeholders (e.g. farmer, friend, family) and information sources (e.g. publication, training, internet) into groups useful for comparison.

Location/Latitude/Longitude

If the location fields can be applied for the stakeholders or information sources then the spatial data could be relevant for information sources too.

TAB: InformationTransactions

This records data about farmer *links* to information sources that may influence farmers' use of the seed system, or the risk of disease spread and ultimately yield. Information sources include people, institutions, publications, or the internet. In the example questionnaire, we focus on variety adoption and information about the management of pests and diseases (IPM).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2806v2 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

DateOfInterview

When the subject was asked questions about their sources of information.

Interviewee

This is the name of the stakeholder who is the *subject of the interview*, or completing the questionnaire. This should be a unique identifier, often the full name should be adequate (although more descriptive information may be necessary). Each Stakeholder should be assigned a unique ID (e.g., if there is Mary Jane from Village A and Mary Jane from village B, each needs a unique ID). *Please note:* the name of the interviewee (person being questioned) will correspond to one of the stakeholder fields, either StakeHolderFrom or StakeHolderTo. Both From and To information are needed from each interviewee. See explanation for those fields and the example questionnaire.

The goal is that if any two stakeholders receive or send seed from/to the same stakeholder, then the data will be complete enough to identify that they share the connection.

InformationSourceFrom

This records where the farmer is getting information from about variety use and pest and disease management.

InformationSourceTo

This would usually be the same as the Interviewee field, but describes who receives the information.

InformationTransactionDate

This field records when the information was gained or accessed. Dates may be less important for the informational aspect of the study.

Crop

The crop that is the focus of the study.

IPMInfoType

Subjects should be prompted to think about the information sources they had for the following: pest and disease recognition, chemical control, crop rotation, roguing, positive selection, planting/harvest time. (This field would receive an “NA” when the information sources relate solely to VarietyInfo.)

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2806v2 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

VarietyInfo

This records variety names for which farmers have received information about variety traits. This could also include varieties they have heard of, but have little information about. (This would be filled with an “NA” if the information sources relate solely to IPM.)

YieldInfo

This relates to each of the varieties named and is here imagined as a yes or no field. A “yes” would mean that the farmer received information about the yield performance of the variety from the source indicated.

ResistanceInfo

This relates to each of the varieties named and is here imagined as a yes or no field. A “yes” would mean that the farmer received information about the resistance of the variety to pests and diseases from the source indicated. It may be of interest to code for different levels of information.

QualityInfo

This relates to each of the varieties named and is here imagined as a yes or no field. A “yes” would mean that the farmer received information about the marketable qualities of the variety from the source indicated, e.g., color, texture, taste, starch levels, nutritional aspects. It may be of interest to code for different levels of information or types of information.

Trustworthiness

This is the subject’s assessment of the information source’s trustworthiness, or reliability on scale of 0-5 (0 is none, 5 is very trustworthy).

Tab: PestDiseaseImportance

This section documents the most important pests and diseases for farmers, either in terms of impact on yield in the last season or in terms of features that influence their pest and disease management decisions. The two lists may or may not be similar. For example they may select varieties resistant to a particular disease, but have another pest or disease reducing yield in the last season.

Interviewee

The farmer recording the information.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2806v2 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

Date

Date of the interview.

PestOrDiseaseYield/PestOrDiseaseIPM

Lists the most important diseases for the farmer in terms of last season's yield and variety selection, respectively.

RankMostImpactOnYield

This ranks the pests and diseases in terms of their importance for the last season's yield. The pathogen would be listed in the PestOrDisease and ranked.

RankIPMDecisions

This ranks the pests and diseases in terms of their importance for IPM decisions. For example, if farmers selected a variety because of its resistance to a pathogen then the pathogen would be listed in the PestOrDiseaseIPM and ranked.

Tab: LastYieldAssessment

This is a just a simple assessment of the last season's (or the last few season's) yield in terms of farmer's expectations.

Interviewee

Name of interviewee.

DateInverviewed

Date of interview.

DateOfHarvest

This records the date of the harvest being assessed, and should at least include the last season's harvest, and potentially the date of other harvests.

YieldAssessment

This records whether the farmer's last yield met expectations, or was higher or lower than expected, e.g. higher, lower, met.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2806v2 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018