Co-producing knowledge with family farming organizations: a citizen science observatory in Santarém, Brazilian Amazon

Emilie Coudel1,2,3,*, Stéphanie Nasuti3,4, Beatriz Abreu dos Santos3,4, Mariana Piva3,4, Valéria Fechine3,5 and Ricardo-Theophilo Folhes3,6

1 CIRAD, UMR SENS, F-34398 Montpellier, France
2 SENS, Univ. Montpellier, CIRAD, IRD, Université Paul Valéry Montpellier 3, Montpellier, France
3 INCT Odyssea – Observatório das Dinâmicas Socioambientais, Brasilia, Brazil
4 Centro de Desenvolvimento Sustentável, University of Brasilia, Brasilia, Brazil
5 Centro de Estudos Avançados Multidisciplinares, Universidade de Brasilia, Brasilia, Brazil
6 Núcleo de Altos Estudos Amazônicos, Universidade Federal do Pará, Belém, Brazil

Abstract – This data paper presents a survey conducted in a participatory manner in the territory of Santarém, in the Brazilian Amazon. The aim is to understand how global changes are affecting family farmers. In the study area, family farming has been confronted over the past 20 years with the rapid expansion of large-scale monocultures, especially soybean. As part of the Odyssea socio-environmental observatory, academic researchers and family farming organizations entered into a partnership to co-produce data that could be strategically useful for these organizations. A process of co-construction of the expectations allowed priorities to be established and the data collection strategy to be defined. Three levels of analysis were chosen in order to allow an integrated understanding of the dynamics of change: the Santarem Plateau territory, the rural community level (living place recognized by the farmers) and the farmers’ households. Twenty-one farmers, called community researchers, organized in 3 teams, applied a questionnaire through the KoboCollect smartphone application to 544 families in the municipalities of Santarém, Mojuí dos Campos and Belterra. Meetings were previously held in the rural communities and questionnaires were applied with representatives of 32 communities. Data was collected between April and June 2019. The community researchers and academic researchers then came together for two collective sessions of data analysis and interpretation in July and October 2019. Data was standardized and cleaned using SPSS software, between September and December 2019. The metadata and databases are available on the CIRAD dataverse.

Keywords: citizen science / community researchers / capacity building / participatory research / family farming / Amazonia / large-scale farming / soybean

Résumé – Co-construire des connaissances avec des organisations d’agriculteurs familiaux : un observatoire de science citoyenne à Santarém, Amazonie brésilienne. Cet article de données présente une enquête menée de manière participative dans la région de Santaré, en Amazonie brésilienne. L’enquête vise à comprendre comment les principaux changements globaux affectent les ménages agricoles. Dans la région d’étude, l’agriculture familiale est confrontée depuis 20 ans à l’expansion rapide de monocultures de grande échelle, notamment de soja. Dans le cadre de l’observatoire socio-environnemental Odyssea, des chercheurs académiques et des organisations d’agriculture familiale ont établi un partenariat pour co-produire les données stratégiques pour les organisations. Un processus de co-construction de la demande a permis d’établir les priorités et de définir la stratégie de collecte des données. Trois niveaux d’analyse ont été choisis, pour permettre une compréhension intégrée des dynamiques de changement : le territoire du Plateau de Santaré, les communautés rurales (lieu de vie reconnu par les agriculteurs) et les ménages agricoles. 21 agriculteurs, appelés chercheurs communautaires, organisés en 3 équipes, ont appliqué un questionnaire grâce à l’application KoboCollect pour smartphone auprès de 544 familles dans
les communautés de Santarém, Mojuí dos Campos et Belterra. Ils ont réalisé au préalable des réunions dans les communautés rurales, soumettant des questionnaires aux représentants de 32 communautés. La collecte a été réalisée entre avril et juin 2019, puis les chercheurs communautaires et chercheurs académiques ont réalisé deux sessions collectives d’analyse et d’interprétation des données en juillet et octobre 2019. Les données ont été contrôlées et nettoyées avec le logiciel SPSS, entre septembre et décembre 2019. Les métadonnées et les bases de données sont disponibles sur le datavase du Cirad.

Mots clés : science citoyenne / chercheurs communautaires / formation de compétences / recherche participative / agriculture familiale / Amazonie / grande exploitation / soja

1 Introduction

Given the rapid changes occurring in global ecosystems and the necessity to understand the complexity of their drivers, a number of socio-ecological observatories have been set up in recent decades (Bourgeron et al., 2018). The term “observatory” originally referred to systematic and long-term observations of physical phenomena to enable a better understanding and to support prediction (Piron, 1996). Today, socio-ecological observatories are framed as platforms that engage diverse researchers, policy makers and community members in science in order to coordinate diverse data streams, gather a variety of types of knowledge and allow for information planning and policy (Bourgeron et al., 2018). However, although such observatories have proved efficient in enabling a scientific understanding of the coupled human-natural systems, it is becoming evident that “it is no longer sufficient to provide lists of environmental indices or reports to inform citizens about changes in their environment” (Liu et al., 2014).

A new generation of observatories is thus now emerging, promoting a more active role of citizens in knowledge co-production processes in order to encourage community-based environmental governance (Liu et al., 2014). In particular, territorial observatories enhance the use of citizen knowledge to inform societal choices and reintroduce political debates in processes of deliberative democracy (Tonneau et al., 2017). However, as Tonneau et al. underline, to enable knowledge to truly become actionable, observatories must encourage collective learning processes, which will permit moving beyond simply making data and information available. Thus, as Liu et al. (2014) put it, raising awareness is not just about alerting the public, it involves engaging the citizens in understanding the problem so that they can make informed decisions of their own. According to Rathnayake et al. (2020), more than citizen science, which is generally focused on collecting data through citizen participation (Bonney, 2016), citizen observatories facilitate governance and decision-making by involving a broad range of stakeholders in order to ultimately bring about change in society (Rathnayake et al., 2020).

Knowledge co-production has emerged over the past decade as part of participatory and transdisciplinary research approaches, in which academic and non-academic actors are both accorded legitimacy in the knowledge production process, from their own places in society. Norström et al. (2020) define knowledge co-production as “iterative and collaborative processes involving diverse types of expertise, knowledge and actors to produce context-specific knowledge and pathways towards a sustainable future” (p. 2). Cash et al. (2003) stress that facilitating communication, translation and mediation between knowledge and action leads to production of knowledge that is more likely to be effective, since its salience (usefulness), credibility and legitimacy is enhanced. When implemented with consideration of the local context (D’Aquino, 2009), co-production processes can produce more than just knowledge; they can develop capacity, build networks and enable the implementation of actions that contribute to sustainability (Durose et al., 2012; Florin and Wandersman, 1990; Norström et al., 2020).

This renewing of the way of involving citizens in co-production strongly influenced the course of the Odyssea project that we present in this data paper. Created in 2016, the Observatory of Dynamics between Societies and their Environments (ODYSSEA, also known as ODISSEIA) brought together a network of Brazilian and European researchers who had been carrying out converging studies in Brazil on relations between societies and their environments in the face of climate change or other global changes. These researchers wished to make their research results more readily available to decision makers and social actors and proposed to set up an observatory which would monitor the main environmental changes, how social actors were adapting to them and thus inform how policies could better support adaptation. Progressively, a group of researchers (including the authors) introduced the idea of involving citizens in the co-construction of the observatory in order to raise the expectations social actors have from an observatory—so as to produce knowledge which might be useful to support their actions and agenda setting. This is how the idea of setting up a pilot observatory in a specific territory emerged.

The region of Santarém, in central Brazilian Amazon (State of Pará), was chosen for this pilot as it is host to a particularly active civil society, with NGOs, universities, unions, and associations representing very diverse populations, and since several projects of the network’s researchers had earlier been carried out there (Bommel et al., 2016; Folhes, 2018; Gardner et al., 2013). Based on the assumption that knowledge co-production requires mutual relationships of trust (Durose et al., 2012) and given the difficulty of reconciling the interests of very diverse actors, we decided to form a partnership with the Unions of Family Farmers and Rural Workers, with which we already had long-standing links through other research projects. In addition, with the aim of the inclusion of a diversity of stakeholders, we involved representatives from three municipalities in order to encourage knowledge production at a territorial level.

Over time, the researchers and the leaders of the unions of family farmers formalized a partnership to carry out a common research process that would support reflexion regarding the strategies to consolidate family farming in a context of
agribusiness expansion. One of the actions of this partnership took the form of a field survey of community leaders and households, undertaken by young community researchers using a smartphone application, KoboCollect (KoboToolbox, 2019).

In this data paper, we present the process of co-producing knowledge with the family farmer organizations at community and household levels and the two resulting databases. However, we do not consider that the “results” become available only a posteriori, i.e., when the processes of data collection, analysis, and interpretation are completed. On the contrary, our research is “process oriented” (Vandenbussche et al., 2020). The entire process was aimed at empowering the actors to define the problems they wanted to investigate, with the perspective that this knowledge would become the basis for more inclusive action strategies. Being involved at all stages of the research – in defining the focus of the research, in data collection and in information analysis – was fundamentally important to the actors; they could feel like “subjects” of the research and not its “objects”. It also enhanced learning processes and communication among citizens and between citizens and researchers, enabling the stakeholders to develop a sense of ownership of the process and to understand how the knowledge they defend was produced and validated, thus ensuring the usability of this knowledge in territorial arenas.

2 Study context: the Santarem Plateau

The study area is located in the municipalities of Santarém, Belterra and Mojuí dos Campos, at the confluence of the Tapajós and Amazonas Rivers, in the west of the State of Pará, Brazil (Fig. 1). This area, formed by an argilous plateau of high agricultural potential, is known as the Santarem Plateau.

Established by Portuguese in 1661, Santarém is one of the first towns of the Amazon region. Diverse rural communities, from many ethnic origins, made a living from fishing and slash-and-burn agriculture for centuries. In the 1960s, with the construction of highways to integrate the Amazon with the rest of Brazil, new colonizers flocked to the region, implanting pastures for cattle production (Becker et al., 1990; Côrtes and D’Antona, 2016; Hébette and Marin, 2004).

In the early 2000s, government incentives and the construction of a port by Cargill in Santarém encouraged the establishment and expansion of large-scale agriculture (Sauer, 2018). Oliveira (2001) and Nahum and Paixão (2014) report a progressive concentration of land, associated with land grabbing schemes and compulsory expulsions, which forced the disintegration of family farming and drove migrants from rural areas to urban centers.
However, the forms of resistance by and innovation of family farmers during this period are poorly documented, although a growing dynamism was evident from the expansion of fruit and vegetable production (Nahum and Paixão, 2014). This production from family farming continues to be essential for the supply of food to the region’s urban populations.

The family farming unions (called Sindicatos dos Trabalhadores Rurais, Agricultores e Agricultoras Familiares, STTRs) of the Santarém region have been taking initiatives to support family farming associations and cooperatives, undertaking campaigns to clarify the impacts that land sales can have on family farmer communities, and promoting the adoption of agroecology as an alternative to soybean monoculture and pesticide use.

Although data of agricultural production exists, estimated by the Brazilian Institute for Geographical Statistics (IBGE) through periodic rural censuses (since 1920, with the most recent census years being 1996, 2006, and 2017), there is very little data on the farmers’ practices and none regarding their well-being. Given this lacuna, the impact of recent changes (in particular of soybean expansion) on smallholders is very difficult to assess for farmer organizations and policymakers.

3 Progressive definition of the focus of the co-production with the stakeholders

We consider the co-production process in a large sense, involving the building of a strong partnership between researchers and farmer unions, co-definition of expectations, data co-production and, finally, knowledge sharing and transformation into action (Fig. 2). Although the main focus of this paper is the data produced, we find it important to contextualize the co-production process.

The partnership of researchers and social actors was built progressively and dated from well before the Odyssea project. Although in this paper, we refer simply to Odyssea, it is in fact a double project, combining the European H2020 Odyssea project and the Brazilian INCT (Instituto Nacional de Ciência e Tecnologia) Odisseia project. Both stand for: Observatory of the Dynamics of Interactions between Societies and their Environments. These projects were intended to bring together a variety of studies which analysed environmental and social change over the long term, putting them at the service of political decision making, as an observatory. This double project first started out as a proposal by the National Institute of science and Technology, called Odisseia and submitted in 2014 in response to a call from the Brazilian Ministry of Science and Technology, to continue the studies carried out as part of the Climate Network coordinated by the University of Brasilia (UnB), broadening the analysis to other dimensions of global change. To support Odisseia, another proposal, Odyssea, was submitted in 2015 by European researchers who had been collaborating for a number of years with the Brazilians in the Amazon region to a Horizon 2020 call from the European Union, to obtain funding to encourage the mobility of Europeans towards Brazil. Odyssea started in 2016, but, due to political instability in Brazil, Odisseia began only in 2018. For more information regarding these projects, consult the following websites: www.odyssea-amazonia.org and http://inct-odisseia.i3gs.org/.

Odyssea/Odisseia brought together researchers from different networks who had been undertaking projects for
many years on social and environmental change and adaptation. These researchers were from diverse backgrounds: hydrology, biology, ecology, geography, climatology, agronomy, economy, sociology, and anthropology among others. The first scientific meetings highlighted the different visions of ways to implement an observatory, ranging from an interactive repository of georeferenced data originating from research projects to a social platform to promote knowledge sharing between researchers and social actors. The Odyssea coordinators felt that these different orientations could converge but that the links between them needed to be built in a concrete case. They thus suggested experimenting with a pilot observatory at a territorial level, a level at which stakeholders would probably be interested in sharing about the changes affecting them (Tonneau et al., 2017).

Santarém was the territory most researchers from Odyssea had been involved with in previous projects, with at least five networks that could be brought together under the ambit of the observatory. Furthermore, all these projects had established partnerships with different stakeholders, in particular with the family farming unions. Researchers thus knew some of the farmer union leaders, but only on a consultative basis, to present their research results and discuss them. In Odyssea, the project coordinators wanted to initiate a co-production process from the start and identified the farmer unions as trustworthy and legitimate partner since they were well organized (with more than 500 delegates in all the rural communities) and truly represented the rural populations that were targeted by the project. One of the researchers who had lived for more than 10 years in Santarém held individual discussions which each of the farmer union presidents and explained the project and what they could gain from it. They accepted to jointly organize a workshop in September 2017 and they were then invited to a scientific seminar of the project in the city of Belém (Pará) where a formal agreement of collaboration, specifying the responsibilities of each party in the co-production process, was signed.

To initiate the co-production process, common expectations on the observatory’s focus needed to emerge. Three workshops, in 2017 and 2018, brought together community leaders, invited by the farmer unions, to discuss the main changes confronting rural communities. A series of questions, inspired by foresight methodologies (Jouvenel, 2009; Patel et al., 2007), oriented the co-definition of the stakeholders’ expectations:

1. What are the main changes that are occurring in the region?
2. What are the actions that have already been carried out to deal with these changes?
3. What futures do you expect? What future would you like to occur?
4. What are the main actions/variables that can lead to this future?
5. Which actions/variables do you have most power to influence?
6. What information do you need to carry out these actions?

Initially, the conceptual framework of the Odyssea observatory was oriented by the concepts of vulnerability and adaptation in the face of socio-environmental changes (Sherbinin et al., 2007; Wisner et al., 2014). Given that these concepts might not be meaningful for our partners, an important step was to define a common vocabulary (Beebeejaun et al., 2014; Norström et al., 2020). During a workshop in 2018, we realized that the concepts of vulnerability and adaptation were interpreted by our partners as negative— in the sense of an acceptance of fatality. The farmer unions wanted to bring visibility to family farming and its contribution to the region, and, in this way, define a positive agenda to defend. They suggested framing our common research as “Consolidating Family Farming”. This choice goes well beyond words, it involves different conceptions about the use of information and a reframing of the role of the observatory. The challenge was to produce information that, instead of supporting stakeholders in adapting to changes, would reveal the impacts of change on their livelihoods and provide them with data to advocate against these changes. We also understood that the farmer unions were eager to obtain tangible results, to transform knowledge into action, and so we accepted to take up this challenge together.

Three priority themes gradually emerged from the co-definition of expectations: land security, contamination by pesticides, and agroecology—all related to the pressure exerted by soybean expansion. The farmer union leaders underlined the lack of information regarding these themes and expressed their strong interest in collecting information on which to base their strategy, both to inform and make aware the rural population and also to be able to back up their arguments in governance arenas. In addition, the farmer union leaders expressed their mistrust of the public data from the decadal agricultural census; they felt that it under-represented the volume and diversity of family production, as well as the flows of produce directed to urban centers. They urged us to support them in collecting this data.

 Undertaking a data collection campaign was not part of the initial research objectives of the Odyssea project. From its experience from previous projects, our research team knew the huge amount of work involved in collecting and organizing data, but after reflecting about it, we saw the opportunity to conduct an experiment to produce fine data in a participatory manner, associating the data collection exercise with another demand of the farmer unions: building capacity among young community leaders. The farmer unions mobilized young people who were already involved in the unions’ activities and they became the “community researchers” we trained to collect data. These young farmers became more than just data collectors; they became proactive stakeholders of the research process (for more details, see Sect. 4.3).

Before carrying out the data collection, we defined together the objectives of the survey: to understand the current state of family farming in the territory of the Santarém Plateau and the changes that have occurred since soy was implanted, including the strategies that were developed by the farmers to resist this expansion. A common framing was defined and shared, incorporating all the issues the stakeholders considered important. In this framing, three levels interact:

- the territorial level, where the main drivers (policies, governance, soybean expansion) can be observed, but also where social actors react through their organizations and networks;
- the community level, where the pressures (mainly the proximity to soybean) and opportunities (markets, local groups) influence the context in which households evolve;
All these dynamics together determine whether rural communities are fragilized or able to resist the growing soybean pressure. This framing oriented the data collection process, as shown in Table 1.

### 4 Data co-production process

The field research began with the identification of the main local dynamics through a collective territorial zoning. This zoning enabled to orient the data collection at two different levels: the community and the household. In this part, we present the data collection strategy and describe each of the two datasets (community and household).

#### 4.1 Zoning at the territorial level to orient data collection

Caron and Cheylan (2005) defend the use of participatory methods based on cartographic representation, called “zoning according to stakeholders” (zonage à dire d’acteurs, ZADA). It enables the stakeholders to share their knowledge about the territory and identify territorial diversity and spatial dynamics. The use of this methodology has the potential to facilitate decision-making and territorial planning by allowing an integrated vision.

Based on this method, a participatory zoning of the Santarem Plateau was carried out with community leaders from each municipality. The objective of this exercise was to characterize the diversity of situations of family farming communities surrounded by agribusiness farming (Cortes et al., 2020). This participatory mapping exercise was conducted to build a general view of the territorial dynamics of the past 20 years with the soy expansion, in particular to understand where communities had disappeared and where (and whether) some of them had stood firm.

This zoning also allowed us to arrive at a sampling method for the data collection. We selected 4 zones in each municipality to represent contrasting situations (Fig. 1):

- **dominant agrobusiness (orange)**, mixed family farming and agribusiness (yellow), dominant family farming (green), and agribusiness advancing on family farming (purple). We thus composed a sub-universe of 12 zones (4 zones × 3 municipalities).

#### 4.2 Survey design

The survey was designed in a collaborative way, involving the farmers and researchers iteratively, as shown in Table 2.

#### 4.3 Data collectors and capacity building

From the beginning, we conceived the data collection campaign as a means to involve and empower young farmers linked to the farmer unions. Each of the three farmer unions was responsible for identifying 6 young leaders, male and female, as well as one field coordinator who had a good knowledge of the entire municipality.

Thus, the data collection team was composed of 18 “community researchers” (the farmer unions gave them this name), who were aged between 18 and 30, 8 of them female and 10 male. Most of these community researchers had previous links with the farmer unions, through trainings or as youth group leaders. The field coordinators were aged between 35 and 55, all male. As agreed with the farmer unions’ representatives, a monthly fellowship was offered to each of them, at the same level as undergraduate students for the community researchers, and a little over the minimum wage level for the coordinators (which is close to what family farmer union leaders earn for their work). A contract was signed between the research project and the farmer unions to cover all the monetary resources necessary to undertake the field work (wages, gasoline for the vehicles, meals served during the meetings to the participants, etc.).

In April 2019, these 18 community researchers, 3 field coordinators, several farmer union representatives, along with 7 academic researchers and 8 post-grad and post-doc fellows from the Odyssea team came together for a week of capacity building. The objective of this exercise was to introduce the future community researchers to the different aspects of conducting research: research postures, ethical considerations, presentation of the questionnaires and reformulation, how to...
apply surveys, and discovery of the KoboCollect smartphone app. A professional facilitator supported the training by devising activities that would encourage a true engagement by the community researchers, such as team motivation exercises, and theater and cultural presentations. On the fourth day of the training, we split into three groups, one per municipality, and went to a community to conduct a pilot meeting to test the focus group methodology and the community questionnaire.

Table 2. Steps in survey elaboration, application and analysis.

| Date               | Step involving farmer organizations                                                                 | Step involving researchers                                                                 |
|--------------------|------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|
| September 2018     | First planning of the data collection campaign and the main topics to be addressed                   |                                                                                          |
| December 2018      | First list of the general questions the farmers wanted addressed                                      |                                                                                          |
| January 2019       | Participatory territorial zoning to define the zones to focus on                                     | Validation in a plenary session of the main categories to be investigated                  |
| February 2019      |                                                                                                     | Elaboration of the detailed elements to be taken into account in the community and household questionnaires |
| March 2019         |                                                                                                     | Formulation of the questions and digitization in KoboToolbox                              |
| Early April 2019   | Discussion of the two surveys during the training session of the community researchers, to rephrase the questions one by one so they would correspond to the farmers’ vocabulary and representations |                                                                                          |
| Mid-April 2019     | Further adaptation of the questions after the first application of the surveys                       |                                                                                          |
| April–July 2019    | Application of the surveys by the community researchers                                             | Regular follow-up and debriefing regarding problems arising in the field                  |
| First week of July |                                                                                                     | Researchers processed all the results from the survey to calculate mean values             |
| July 2019          | First reactions regarding the field experience and the results of the survey: What was learned?     |                                                                                          |
|                    | What was surprising? What were the difficulties?                                                    |                                                                                          |
| September–October 2019 |                                                                                                     | Researchers deepen the analyses of the results, and cross variables with the support of a statistician |
| October 2019       | Debate on detailed results: what results are the most important and deserve to be presented?        |                                                                                          |
|                    | To whom should they be presented? How to present them? What narratives?                             |                                                                                          |
| December 2019      |                                                                                                     | Validation by the research team of the final version of the database and of the dictionary of variables systematized by the statistician |

Two academic fellows accompanied the field teams during the entire data collection process (and are co-authors of this paper). These fellows were from an academic background and were recruited by the academic team to be part of the INCT Odissea staff: a junior researcher who did an interdisciplinary masters about Santarem at the beginning of the Odyssea project was hired to maintain a presence in Santarem all through the project; and a post-doc from a biology background...
To offer further support all through the data collection and analysis process.

### 4.4 Sampling method and population covered by the study

The survey was applied according to a non-probabilistic sampling process by quotas. Data collection was organized according to the participatory zoning, with 4 distinct zones in each municipality (Fig. 1). Data was collected at the community and household levels.

#### 4.4.1 Community level

A first survey was conducted at the community level during meetings in each zone. The field team organized one meeting in the largest community of each zone (Fig. 1). The field coordinator had invited representatives from all the surrounding rural communities. The objective was to have the representation of as many communities as possible.

In total, 12 community meetings were organized, during which 32 community questionnaires were applied (see Sect. 5.2.1 for details on the contents of the questionnaire) and responded to by the community representatives present (sometimes 2 or 3 stakeholders answered the questionnaire together). This set represents about one-fourth of the estimated universe of rural communities of the Santarem Plateau. The sample universe is, however, not entirely clear, due to the lack of reliable official statistical information. A collective mapping undertaken by community leaders of the region in 2008 lists 436 rural communities in the three municipalities, 126 of which are in our study area (OpenStreetMap, 2019).

#### 4.4.2 Household level

The second survey was conducted at the household level, with the application of a questionnaire with heads of households (see Sect. 5.2.2 for details on the contents of the questionnaire). Among households in the rural areas, only those engaged in agricultural production were targeted.

In each zone, we aimed at interviewing households in at least 5 communities, with at least 9 interviews per community. Due to logistical difficulties (bad roads, rainy season), it would not have been possible to visit several communities on the same day and the teams thus focused on one community per day. A realistic day-objective for each pair of community researchers was to carry out 3 interviews per day. With 3 pairs per municipality, the target of 9 questionnaires per day was generally met. Households were chosen to represent a diversity of situations in each community. In total, 544 household questionnaires were applied, across 62 different communities (Tab. 3 and Fig. 1).

The research universe is not fully clear. IBGE (2018) counts 8043 family farming establishments in the whole area of the 3 municipalities, which is much larger than our area. Our best estimate is based on the family farmers of the Santarem Plateau registered with the three farmer unions, which is a total of 10 600 farmers. Often, several members of a family register at the union (husband and wife, son and father), thus the unions estimate that there are approximately 6000 farming families, which would mean that we surveyed approximately 10% of the families of the Plateau. However, our aim was not to be representative of the total population, but to compare the reality between the three municipalities and the different zones. We therefore aimed for similar sample sizes in each municipality.

### 4.5 Data collection and survey administration

For the execution of the field work, three teams were created, one per municipality (with farmers from that municipality). Each team consisted of three pairs of individuals, responsible for conducting the meetings and for filling out the questionnaires, and one field coordinator, responsible for arranging the meetings and for other logistical issues. Data collection took place consecutively in the three municipalities, with the presence and accompaniment of two academic fellows. In each municipality, the academic fellows

| Municipality          | Zone (see Map 1) | Number of households interviewed | Total |
|-----------------------|------------------|----------------------------------|-------|
| Santarém              |                  |                                  |       |
|                       | S1 Cipoal        | 48                               | 185   |
|                       | S2 Início da Curuá Una | 47                               |       |
|                       | S3 Final da Curuá Una | 45                               |       |
|                       | S4 Ituqui        | 45                               |       |
| Mojuí dos Campos      |                  |                                  |       |
|                       | M1 Pólo das hortaliças | 41                               | 197   |
|                       | M2 Pólo do Abacaxi | 45                               |       |
|                       | M4 PA Jabutí     | 56                               |       |
|                       | M5 Boa Sorte (Farinha) | 55                               |       |
| Belterra              |                  |                                  |       |
|                       | B1 Centro        | 38                               | 162   |
|                       | B2 Trevo         | 44                               |       |
|                       | B3 São Raimundo  | 35                               |       |
|                       | B4 São Jorge     | 45                               |       |
| **Total**             |                  |                                  | 544   |
carried out a detailed follow-up of the data collection during the first two weeks. They also helped facilitate all the community meetings.

The first week of follow-up in Santarem was essential for diagnosing and correcting aspects of the organization of the questionnaire and to add missing questions. Moreover, it was essential for the community researchers to become familiar in using the KoboCollect app, to learn how to present the project and the informed consent form to be signed, and to understand how to conduct the interview. After this initial follow-up period, periodic discussions were held after each zone meeting, and the academic fellows had daily contact with the teams in each municipality by phone and messaging apps.

The data collection took about one month in each municipality, starting out by Santarem in April, continuing in Mojuí dos Campos in May, and finishing in Belterra in June.

At the community level, the meetings were organized in two stages. First, focus groups were brought together around a georeferenced map of the zone, with the objective of identifying the landscape elements and land use (rivers, forests, family farming fields, soy plantations) and the main structures, such as schools, health centers, and collective production equipment (Fig. 3). The objective was to identify the main changes in the landscape, reflect upon the situation of family farming in the zone and the community’s future perspectives. Second, the community questionnaires were applied with the community representatives who were present (for details on the survey applied, see Sect. 5.2.1). After collecting the information, the groups met in plenary and a table was produced to compare the information from each community. The micro-zoning and the comparative table encouraged the community representatives to share what had been the main changes in recent years and what were the main challenges confronting family farmers.

Within each community, the households were chosen by the team coordinator, after discussion with community representatives. Each interview lasted between 60 and 90 minutes, including the presentation of the project and of the informed consent form, which was signed at the end of the interview, and the survey itself (for details on the survey applied, see Sect. 5.2.2). One member would fill out the questionnaire on the KoboCollect app on the phone, and the
other was responsible for writing down the responses on paper (Fig. 3). The pair took turns to conduct the interviews and enter responses in KoboCollect, so that everyone gained confidence in the process. At the end of each data-collection day, the pairs reviewed the notes and sent the filled-in questionnaires to be saved online.

4.6 Collective interpretation of data

After the data collection, we involved the community researchers also in the data analysis (Fig. 3). This had a two-fold objective: to refine our understanding of the data collected and to reinforce the community researchers’ sense of ownership of the data.

In July 2019, we spent a week together to discuss the first results, mainly mean values by municipality for each variable. We asked the community researchers if the results were coherent with what they had perceived, what surprised them, what might be a result of a miscomprehension of the question, etc. They also told us about what they had learnt during the data collection experience.

In October 2019, after data was cleaned and analyzed by crossing variables, we spent another week together with the community researchers, representatives of farmer unions and academic researchers to define messages and narratives that were important and to decide where (in what arenas, groups, events) they should be shared. The leaders of the farmer unions sent us material they found relevant from other projects (mainly development projects with NGOs) and we decided together to make posters that could be hung up in schools and community headquarters, where meetings took place. We focused on 4 posters to highlight the themes that had initially been identified: community life (instead of territorial insecurity), peasant production (to present numbers on what was being produced), contamination by pesticides, and finally agroecology and how it was understood in the territory. The contents of each poster were progressively fine-tuned, on the basis of collective discussions in the following months, to integrate different results from the analyses in a coherent message. We solicited the help of scientific media experts for writing simple messages and organizing them with infographics. Initially, these posters were supposed to be shared in a meeting with the different territorial stakeholders in May 2020, but due to the Covid pandemic, we had to postpone this event. The leaders of the farmer unions asked to adapt the posters to cards so that they could share them over social media, mainly WhatsApp, and we selected 6 key messages for each theme. The posters and cards can be accessed at the following link: http://inct-odisseia.i3gs.org/interface-com-a-sociedade/.

At last, in December 2020, we held an online event with social organizations, extension agents, researchers from other institutions, and decision-makers. This event represented a milestone in the process, as important organizations of the territory became aware of the potential of the results presented and as existing strategic partnership were strengthened. In August 2020, the Territorial Forum for Fighting Pesticide Effects, which brings together governmental officials and civil society representatives, invited us to present the results and define strategic orientations to take action, a process which is currently ongoing.

5 Description of data

The data consists of two complementary databases, at the community level and at the household level.

5.1 Producer

The producer of the two databases are the two linked projects:
- Odyssea, Observatory the Dynamics of Interactions between Societies and their Environments in the Amazon, which received funding from the European Union’s Horizon 2020 Research and Innovation program under the Marie Skłodowska-Curie grant agreement no. 691053: https://www.odyssea-amazonia.org;
- INCT Odisseia, Instituto Nacional de Ciência e Tecnologia/Observatório das Dinâmicas Socioambientais, no. 16-2014, with funding from CNPq/Capes/FAP-DF: http://inct-odisseia.i3gs.org/.

5.2 The databases

5.2.1 Community database

The first unit of analysis is the community, defined as a common place of living. We did not discuss the physical, social, or institutional boundaries of what “makes a community”; we instead used the local representatives’ own delimitations and denominations.

The questionnaire applied with the community representatives covered the following topics:
- historical background (creation, main dates);
- demographics (past and current population);
- collective action and local institutions;
- public infrastructure (access to water, energy, mobile phone signal, garbage disposal), access to schools, access to health services;
- local problems;
- land situation;
- agricultural activities (past and current practices);
- agroecology initiatives;
- perception of change (climatic change, social change);
- perception of agribusiness;
- difficulties and future perspectives.

In all, there are 324 variables in 12 categories.

5.2.2 Household database

The second unit of observation is the family. As for the community level, we let the respondents define the boundaries of this unit.

The questionnaire covered the following structuring themes:
– interviewees’ profile (education, migration history, professional activity);
– household composition and professional activities;
– participation in social organizations and events;
– income composition and productive activities;
– agricultural practices (including use of pesticides);
– commercialization of production;
– technical assistance and production strategies;
– land dynamics of the property;
– perception of changes and perspectives for the future.

In all, there are 786 variables in 9 categories.

5.3 Temporal coverage

Data was collected between April and July 2019. The community meetings, during which the community questionnaire was applied, took place between April and June 2019, at the rate of approximately one meeting per week. The household questionnaire was applied by the community researchers between April 3, 2019 and July 3, 2019. The data collected refers to the situation of the household and agricultural practices at the time of the interview, with the production being estimated for the past year.

5.4 Quality of data/main deviations from survey plan

As far as the research was concerned, our intent was to maintain the representativeness of the groups and subjects surveyed in the study, with the sample being based on quantification rather than on randomness. We thus opted for selection through observable characteristics. In some zones, there were fewer farmers than the target quota (45 households per zone), which limited the sample (Tab. 3). However, in most zones, the data collectors were eager to survey more households and therefore exceeded the quota.

5.5 Data processing

After applying the questionnaires, two databases were generated by KoboToolbox in Excel with the information gathered from households and community representatives. The databases were migrated to the statistical software SPSS to make it easier to construct indicators and derive analytical results for the study, and to standardize variables to allow sharing of data with external audiences.

Data consistency was maintained during three phases of standardization in order to adjust and ensure updates, following the need for concordance and identification in the database:

– deviation of absolute or relative numerical values, with adjustments of the numerical value or conversion of measurement units.

– deviation of age or date, period or historical series, such as time period since a given date in years or months, based on the period of the data survey, in pursuit of a finer level of granularity.

– deviation regarding the writing and language, corrections of grammatical errors and standardization of answers or grouping of keywords to ensure coherence between different answers of a same variable.

The conversion of the units of surface area, mass and volume required a thorough process. In Brazilian rural areas, farmers use a wide diversity of units at the same time: they vary according not only to the interviewee’s place of origin, but also to the type of area they describe or the product they refer to. During the interview, in order to facilitate the collection of information and avoid conversion errors, we let the respondents express themselves using the units of their preference. In total, 3 different units of surface, 2 units of mass, and 3 units of volume were collected, which were all converted to the metric system (hectares, kilos and liters) during database standardization.
Pricing of the produce was also challenging. It was undertaken based on complementary research and surveys of commodity prices at fairs and in food shops. Finally, we anonymized the data, deleting the name and georeferenced localization of each respondent. To compensate for the loss of the localization information, we added a calculated variable named “proximity to soy plantation”.

6 Description of the database and other files, localization of metadata and conditions of access to data

Both databases are available as dataverses (Tab. 4). Along with the databases (.csv format), we make available the questionnaires used and the dictionaries of the variables (.pdf).

7 Potential use of this co-produced data

These databases are licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License.

Use of data for publication is under embargo until July 2022 to enable prior publication by the co-producers of the data.

As we have already emphasized, we expect this data to primarily help the local actors in fostering local dynamics for the construction of strategic actions, decision-making, advocacy, or simply for updating their knowledge of local reality. In order to generate familiarity with the data, as also to impart a sense of ownership, we sought to involve the leaders of the farmer unions and the community researchers at all stages of data preparation, collection, and interpretation.

We also put emphasis on sharing this knowledge through dissemination material. We used different media and languages: thematic printed posters, to be distributed in the communities where the farmer unions operate, digital cards for dissemination campaigns via social networks, and a “results notebook” presenting the research data through graphs and maps. Various collective sessions, bringing together union leaders, community researchers and academic researchers, were held to co-design this material. It is available at: http://inct-odisseia.13gs.org/interface-com-a-sociedade/.

This data can be potentially useful to organizations and researchers wishing to understand farmer practices, impacts of climate change and agribusiness expansion on households, or general infrastructure and social dynamics in communities.

8 Limits

Two levels of limits of this data have become evident during the process: the data quality for sophisticated analyses, and the potential of the use of such data by the farmers. First, regarding data quality, we made choices to simplify the questionnaires that limit the potential of the use of the data for sophisticated statistics. We had learned from previous studies that farmers find it difficult to provide accurate quantitative information about production. So we preferred to use qualitative categories, for example, instead of asking farmers to assess family revenue, we asked them how much they estimated that they earned in a good month and in a bad month, choosing their answers from categories (less than half a minimum wage, less than a minimum wage, more than a minimum wage, more than three times the minimum wage). While this approach proved to be relevant for descriptive statistics, it limits the ability to make correlations. However, since our objective was to empower the farmers to make use of the information, and not to conduct sophisticated analyses, this has not been a significant limitation for us, but might be so for other users of the database.

Second, regarding the use that can be made of this data by the farmers, we realized that even with important translation efforts, the comprehension of the information is not straightforward.

The demand from the farmer unions to produce quantitative information determined the format of the results, which were presented as tables, graphics, and maps. We were aware that the farmers might find these formats difficult to read and so we organized several workshops to explain how to read graphs, provide an introduction to basic statistics (what is a mean value, what is a percentage, etc.). Even then, the difficulty in understanding this information limited the scope of the debates. We realized that the simpler the information (selection of one percentage value, for example), the better the debate. Statistical language, in much the same way as written text, constitutes a language specific to the academic world and is potentially excluding and elitist. And yet, paradoxically, it was precisely this effect that was sought by the leaders of the farmer unions, who see the mobilization of information formulated in a “scientific language” as an asset that can help their claims gain legitimacy in certain arenas. However, it required from us a significant reduction in the complexity of the results we are normally used to, which, we must admit, was also an interesting exercise for us as academics.

9 Conclusion: perspectives of the co-production process

The “Consolidating Family Farming” data collection campaign had two objectives. The first was to produce data, at a scale that was relevant for meeting the demands of our local partners, in particular to increase visibility of the situation of family farmers in the soybean expansion areas. The second objective was to contribute to the capacity building process of the young leaders who acted as “community researchers”.

As far as the first objective is concerned, co-producing the data proved beneficial in several ways for the quality of the data itself, imparting, as Cash et al. (2003) put it, more salience (usefulness), credibility and legitimacy. On the one hand, the social legitimacy and local knowledge of the family farming unions and the team coordinators helped overcome several difficulties in the field. The questions were translated/reformulated in a way that best allowed the family farmers to understand them, and the organization of meetings was greatly facilitated by the experience of the team coordinators. Furthermore, they helped the research to be carried out in
places not easily accessible to academic researchers, in particular where soy producers would be suspicious of academic researchers and where there exists a permanent climate of violence (a murder was perpetrated during the fieldwork in a community where the research was carried out, not with any link with our research, but it was nonetheless an unnerving incident). On the other hand, the academic legitimacy and methodological knowledge of the academic researchers helped produce rigorous and relevant information (Balazs and Morello-Frosch, 2013), which thus gained in credibility and which can thus be used by the stakeholders to reveal the situation of family farming in the territory.

As our second objective states, our results go beyond the data collected, the results of the co-production process are also important for us, as they involve a relational transformation that opens up perspectives for empowerment (Darby, 2017). Some limitations in the data find their counterpart in the capacity building of the young leaders who were involved in the research. Besides new perceptions and knowledge about their reality, the young leaders reported that “doing research” contributed to their self-knowledge, their ability to work as teams, and encouraged reflections on the importance of youth participation in organizations. They also mentioned that the experience allowed them to better understand the work of researchers and the challenges they face, thus contributing to a mutual understanding and convergence between the academic and social spheres. As academic researchers, we also learnt a great deal from this process, in particular how to adapt our methods to truly involve the social stakeholders.

Most importantly, this collaboration has helped build trust with the farmer organizations, a fundamental bedrock on which to continue to build our citizen science observatory together. We planned this co-construction as a step-by-step process, not aiming at an a priori long term observatory, which might not be relevant or meaningful for the stakeholders (Grislain et al., 2020), but rather through concrete and negotiated steps that progressively engage the stakeholders in the sensemaking (Norström et al., 2020).

Acknowledgements. We are so grateful to the farmer leaders for their trust and partnership, to the young community researchers for their enthusiasm and to all the farmer families for their warm welcoming. Many academic researchers contributed in this project, the authors are those who coordinated the process and formalized the databases. We thank all our colleagues for the rich debates with them, with special credit to Vincent Bonnal for the map elaboration. Finally, we thank the two anonymous reviewers for their valuable contributions. The project leading to this article has received funding from the European Union’s Horizon 2020 Research and Innovation programme ODYSSEA under the Marie Skłodowska-Curie grant agreement No791053 and the project INCT n°16-2014 ODISSEA, with funding from CNPq/Capes/FAPDF. The research was validated by the Research Ethics Committee (CEP) of the University of Brasilia on October 25, 2018, with the certificate n. 95385318.7.0000.5540.

References

Balazs CL, Morello-Frosch R. 2013. The three Rs: How community-based participatory research strengthens the rigor, relevance, and reach of science. Environmental Justice 6(1): 9–16. https://doi.org/10.1089/env.2012.0017.

Becker BK, Miranda M, Machado L. 1990. Fronteira amazônica: questões sobre a gestão do território. Brasília, Distrito Federal: UNB, 219p.

Beebeejauny Y, Durose C, Rees J, Richardson J, Richardson L. 2014. “Beyond text”: Exploring ethos and method in co-producing research with communities. Community Development Journal 49 (1): 37–53. https://doi.org/10.1093/cdj/bst008.

Bonnell P, Bonnet MP, Coudel E, Haentjens E, Kraus CN, Melo G, et al. 2016. Livelihoods of local communities in an Amazonian floodplain coping with global changes: from role-playing games to hybrid simulations to involve local stakeholders in participatory foresight study at territorial level. Proceedings 8th International Congress on Environmental Modelling and Software, iEMSs 2016, vol. 4.

Bonney R. 2016. Can citizen science enhance public understanding of science? Public Underst Sci 25(1): 2–16. https://doi.org/10.11177/09636625156070406.

Bourgon P, Kliiskey A, Alessa L, Loescher H, Krauze K, Virapongse A, et al. 2018. Understanding large-scale, complex, human–environmental processes: A framework for social-ecological observatories. Frontiers in Ecology and the Environment 16(S1). https://doi.org/10.1002/fee.1797.

Caron P, Cheylan JP. 2005. Donner sens à l’information géographique pour accompagner les projets de territoire : cartes et représentations spatiales comme supports d’itinéraires croisés. Géocarrefour 80 (2): 111–122. https://doi.org/10.4000/geocarrefour.1031.

Cash DW, Clark WC, Alcock F, Dickson NM, Eckley N, Guston DH, et al. 2003. Knowledge systems for sustainable development. Proceedings of the National Academy of Sciences of the United States of America 100(14): 8086–8091. https://doi.org/10.1073/pnas.1231332100.

Côrtez J, D’Antona A. 2016. Fronteira agrícola na Amazônia contemporânea: Repensando o paradigma a partir da mobilidade da população de Santarém-PA. Boletim do Museu Paraense Emílio Goeldi:Ciencias Humanas 11(2): 415–430. https://doi.org/10.1590/1981.81220160000005.

Côrtez JP, Coudel E, Piriaux M, Silva MP, Santos BA, Folhes R, et al. 2020. Quais as perspectivas da agricultura familiar em um contexto de expansão do agronegócio? Zoneamento participativo com representantes comunitários do Planalto Santarenano. Confins 45: 0–24. https://doi.org/10.4000/confins.28077.

D’Aquino P. 2009. La participation comme élément d’une stratégie globale d’intervention : l’approche « gestion autonome progressive ». Cahiers Agricultures 18(5): 433–440. https://doi.org/10.1684/agr.2009.0330.

Darby S. 2017. Making space for co-produced research “impact”: learning from a participatory action research case study. Area 49 (2): 230–237. https://doi.org/10.1111/area.12321.

Durose C, Beebeejauny Y, Rees J, Richardson J, Richardson L. 2012. Towards co-production in research with communities. Connected Communities 1–15.

Florin P, Wandersman A. 1990. An introduction to citizen participation, voluntary organizations, and community development: Insights for empowerment through research. American Journal of Community Psychology 18(1): 41–54. https://doi.org/10.1007/BF00922688.

Folhes RT. 2018. A gênese da transumança no baixo Rio Amazonas: arranjos fundiários, relações de poder e mobilidade entre ecossistemas. Boletim Goiano de Geografia 38(1): 138. https://doi.org/10.5216/bgg.v38i1.52818.

Gardner TA, Ferreira J, Barlow J, Lees AC, Parry L, Guimaraes Vieira IC, et al. 2013. A social and ecological assessment of tropical land uses at multiple scales: The Sustainable Amazon Network.
