Philos Theor Pract Biol (2022) 14:16

RESEARCH ARTICLE

Process-Sensitive Naming: Trait Descriptors and the Shifting Semantics of Plant (Data) Science

Sabina Leonelli*

This paper examines classification practices in the domain of plant data semantics, and particularly methods used to label plant traits to foster the collection, management, linkage and analysis of data about crops across locations. Such methods crucially inform research and interventions on plants and agriculture. The efforts required to share data place in sharp relief the forms of diversity characterizing the systems used to capture the biological and environmental characteristics of plant variants: particularly the biological, cultural, scientific and semantic diversity affecting the identification and description of plant traits, the methods used to generate and process data, and the goals and skills of those with relevant expertise—including farmers and breeders. Through a study of the Crop Ontology (which explicitly recognizes and negotiates diversity) and its application to cassava breeding, I argue for a *process-sensitive* approach to the naming of plant traits that focuses on documenting environmental processes instead of biological products. I claim that this approach can foster reliable linkage and robust re-use of plant data, while at the same time facilitating dialogue between data scientists, plant researchers, breeders, and other relevant experts in ways that crucially inform agricultural interventions. I conclude that the study of data semantics and related descriptors constitutes a productive and underexplored way to think about the epistemic import of naming traits within plant science. The effort to articulate semantic differences among plant varieties and methods of data processing can generate newly inclusive ways to develop and communicate biological knowledge. In turn, such practices have the potential to defy existing understandings of systematisation and hierarchies of expertise in biology, thus bolstering the extent to which plant science can support biodiversity and sustainable agriculture.

Keywords

data science • phenomics • diversity • computational ontologies • databases • plant traits • classification • taxonomy • precision agriculture

---

*Exeter Centre for the Study of the Life Sciences (Egenis), University of Exeter, EX4 4PJ, Exeter UK, S.Leonelli@exeter.ac.uk https://orcid.org/0000-0002-7815-6609

Received 25 January 2022; Revised 6 August 2022; Accepted 8 August 2022
doi:10.3998/ptpbio.16039257.000000

1 Introduction: Naming Traits to Circulate Data

The current combination of climate change, armed conflicts, economic crisis and biodiversity depletion constitutes a severe test for agricultural systems around the world, with global crop production projected to drop dramatically unless humans find ways to re-imagine food chains and how natural resources are managed. The collection and interpretation of data about plants plays an important role in the development of scientific and technological insights that may inform cultivation strategies; yet how plants are identified, studied and portrayed within large data systems continues to be a matter of controversy, with traditional taxonomy proving insufficient as a scaffold for gathering and organizing key information about how plants relate to the environment. This paper argues for a process-sensitive approach to the naming of plant traits, one which aims to embrace and document the diversity of existing systems for the classification, management, and description of plant data—thereby improving existing understandings of plant-environment interactions as well as fostering dialogue between data scientists, plant researchers, breeders, farmers, and other relevant experts.

The starting point for this argument is consideration of the classification challenges that emerge in relation to contemporary efforts to produce, disseminate and re-use data about plant traits ranging from genetic to morphological, environmental and cultural (i.e., relating to the use made of plants within human societies). This domain is often referred by data curators and participants in plant research as *plant data semantics*. It is a key contention of this paper that focusing on this domain constitutes a productive and underexplored way to explore the scientific and philosophical significance of practices used to name, describe and classify the properties of organisms within biology and its applications, and particularly the manners in which multiple knowledge systems are incorporated into (or excluded from) crop science. The effort to share data about crops across locations places in sharp relief the diversity of the biological and environmental characteristics being studied, as well as the methods used to generate and process data, and the background, goals and skills of those with relevant expertise—including breeders, farmers and indigenous communities. Such diversity makes attempts to assemble and analyze data into a crucible for confrontation among semantic systems, thus exemplifying not just the technical challenge of integrating data sources, but the conceptual and socio-political challenge of bringing widely different forms of knowledge about plants into productive dialogue with each other.[1] A broader question underpinning such concerns is the extent to which naming practices used in plant research affect not only how humans identify diverse species or varieties—which has been a long-standing issue in conservation biology and agroecology (e.g., Nazarea 1998)—but also how humans interact with the plant world. In what follows, I show how efforts to articulate and manage differences among plant data practices may generate newly inclusive ways to develop and communicate biological knowledge, thereby taking better account of existing forms of diversity in plant cultivation and research than traditional taxonomy. In turn, such data practices—if adequately managed—have the potential to defy existing understandings of systematisation and hierarchies of expertise in biology, thereby fostering novel approaches to the use of plant science towards supporting biodiversity and sustainable agriculture.

My discussion centers on the epistemic practices used to label crop data collected through field trials around the world in ways that foster data linkage and re-use across locations and for a variety of purposes. I examine the production, circulation and re-use of phenomic data collected on cassava (*Manihot esculenta*), a tuber whose roots provide essential sustenance to millions of people across central Africa, South America and South-East Asia. Scientific research on cassava

---

[1] A parallel and more visible debate on knowledge integration is happening within ecology at large (Albuquerque et al. 2021) and has been promoted by ethnobotanists for decades (Ludwig 2016).

has been long overlooked in favour of high-yield species used as food in the Global North, such as maize, wheat and rice. Efforts to understand cassava biology and biodiversity have intensified over the last decade, thereby providing an excellent example of crop research supported by the accumulation and management of large volumes of data from highly heterogeneous sources. I focus on the data practices promoted by researchers at the International Institute for Tropical Agriculture (IITA) in Ibadan, Nigeria, in collaboration with local breeders and international experts in data curation.[2] I highlight their contributions to developing CassavaBase (NextGen Cassava 2022), one of the key databases worldwide for access to data on cassava, and the Crop Ontology (Crop Ontology Community of Practice 2021), a semantic system devised to capture and share highly diverse data around cassava and other crops. My analysis of this case is informed by: archival research on the history of plant trait descriptors (particularly those used by the Consultative Group for International Agricultural Research, or CGIAR); published sources on the Crop Ontology and CassavaBase; and semi-structured interviews and informal discussions with researchers involved in developing these resources. These interactions took place in person as well as over email and digital platforms between 2016 and 2019, including a visit to IITA in July 2017 during which I interviewed both the developers of CassavaBase and some of the researchers and breeders involved in providing feedback to those efforts.[3]

The paper is structured as follows. In the second section, I briefly discuss recent developments in plant science and the ways in which such developments connect with contemporary attempts to collect data generated through field trials and link them with other types of plant data. I introduce the notion of data linkage, now recognized as a main strategy for such initiatives, and point to the importance of semantic systems such as plant trait descriptors in enabling data analysis across locations and infrastructures. The key challenge for trait descriptors, I argue, is tackling the interrelated forms of diversity that underpin plant data collection, linkage and re-use: namely *biological* diversity among plant variants and their environments, *cultural* diversity in the communities—and respective expertise—involved in the study of plants, *scientific* diversity in methods and approaches used to study plants, and the resulting *semantic* diversity in the naming systems used to describe traits. The third section zooms into one specific attempt to capture phenomic data on cassava, the IPGRI system of trait descriptors, and discusses the key challenges emerging from the use of such descriptors in the field, each of them stemming from an inability to manage biological, cultural, scientific and semantic diversity. The fourth section examines how Crop Ontology and CassavaBase curators confront these challenges, with particular attention to the ways in which they structured their semantic system to encompass and document the various forms of diversity underpinning the use of data on cassava traits. In the fifth section I argue that these curators have shifted their focus from capturing *biological products*, which involves documenting plant traits with marginal attention to the specific circumstances of plant development and data collection, to capturing *environmental processes*, i.e., the interactions

---

[2] I chose this field site due to its innovative work towards developing data infrastructures to collect and disseminate phenotypic data on cassava, its efforts to make it comparable and interoperable with relevant data from other sites, its collaborative links to international data semantics initiatives such as the Crop Ontology, and its engagement with local breeders' communities as significant constituents and sources of expertise for such research. See Leonelli (2022) and Curry and Leonelli (under review) for historical background on such efforts.

[3] Some of the transcripts from these interviews (those which interviewees agreed to release as Open Data) are available on Zenodo (https://zenodo.org/communities/datastudies). All interviews and fieldwork were carried out with ethical clearance from the Ethics Committee of the College for Social Science and International Studies at the University of Exeter. Interview transcripts are labelled with a sequence of letters and numbers indicating whether the interviewee was a principal investigator or a researcher, the number associated to the interviewee and the order in which the interview was taken (e.g., P_12_B indicates the second interview with principal investigator number 12 in my sample).

between plants, humans and wider ecosystems that have given rise to the objects being evaluated. The resulting semantic system underscores and documents the importance of know-how and sensory perception for the practices of data collection used to describe and measure phenotypic traits. To this aim, it uses a mix of qualitative and quantitative information (including imaging) to capture tactile traits such as the consistency of the cassava pulp and information about local uses of cassava, such as the taste of foods produced through boiling or mashing the roots. This naming system thereby acknowledges and underscores the limitations of using particular tools, language or numbers to capture plant traits; and the significance of local conditions of crop use (including consumption) towards identifying relevant traits and pertinent data in the field. This emphasis in turn fosters consideration of traits and naming criteria pertaining to forms of expertise not typically integrated within plant taxonomy, such as traits favoured by breeders due to their relevance in local markets or traits singled out by farmers as relevant to cooking cassava. In the final section, I articulate the broader biological and philosophical significance of this approach, which I call *process-sensitive naming*, including its implications for the semantics of plant science and biology more generally, the function and goals of such semantics in relation to agricultural development, and its practical implementation in data systems.

2 Semantic Diversity and the Prospects of Plant Data Linkage

Over the last two decades, plant science has sought to apply fundamental insights, models and techniques developed through laboratory studies of model organisms to research on crops (Henkhaus et al. 2020; Ankeny and Leonelli 2020), and particularly to research aimed to breed high-yield varieties that can thrive in the fast-changing environmental conditions created by climate change (Atlin et al. 2017; Taylor 2018; Harfouche et al. 2019). This move has been accompanied by a growth in efforts to: (1) move research outside of the standard laboratory environment and into hybrid spaces such as field stations, farm platforms and smart glasshouses, which are construed to better capture features of the natural environment and research the effect of environmental stressors on plant growth; (2) integrate agronomic research with 'basic' plant science, so as to harness cutting-edge insights into molecular mechanisms and related technologies to increase food security; (3) study plant species of economic and cultural interest to parts of the world other than Europe and the United States, such as cassava, yam and bambara groundnut, with the hope to bring new and more resilient source of food to the global market; and (4) increase knowledge about gene-environment interactions, using morphological traits as conduits to understand the impact of genetic modifications and/or environmental changes on plant structures and behaviours.

These trends greatly expand the forms of expertise of potential relevance to the scientific study of plants, in particular by highlighting the scientific value of knowledge arising from agronomic practices such as farming and breeding—and related data. Relatedly, these trends rely heavily on the availability of infrastructures where plant data collected in different parts of the world can be shared among researchers, which in turn require semantic systems through which such data can be organised and retrieved. The nature and content of such systems has been a matter of debate at least since the 1950s, when the potential of using digital infrastructures to collect and disseminate data first started to be recognised by international organisations such as the Food and Agriculture Organisation (FAO) and the CGIAR (Curry and Leonelli under review). A key question in this domain: Which naming practices[4] can best underpin current

---

[4] In what follows, naming practices are taken to encompass all decisions made around how a particular biological entity or process (ranging from a group of organisms to a specific trait) is identified, described and conceptualised. Naming practices thus encompass assumptions around the relations between the entity in question and its envir-

efforts towards plant data systematization and circulation—and related standards and infrastructures?

The most prominent effort towards a standard naming system is of course the Linnaean nomenclature currently formalised by the Schenzen Code (the International Code of Nomenclature for Algae, Fungi, and Plants). The Code aims to develop and certify plant names as common, unique and distinctive reference for a given taxon, which ensures that no confusion can arise over which species a given name refers to and underpins any further investigation of that species, including the tagging and organisation of data around it. The identification and description of traits is here a by-product of the identification of species, since differences among traits are a crucial means of distinguishing one taxon from another. Linnaean taxonomy does not in fact aim to name traits, but rather to store and organise traits found in the broader anatomy and physiology literature in order to use them to differentiate among species. This system is by no means infallible or universally recognised; there continues to be a lively debate over the validity of this system as the main reference point for naming plants (de Queiroz 1988; Franz and Thau 2010; Conix 2019) and the possible alternatives to what some commentators called 'taxonomic anarchy' (Garnett and Christidis 2017). In this paper I shall not consider the details of these discussions, focusing instead on the fact that they mostly happen in relation to the naming of *species* as a key taxonomic unit (Hauber 2019), which in turn involves deciding how to group organisms (and thereby what constitute salient differences and communalities between them, often with a strong focus on genetic characteristics) as well as selecting a type specimen to represent each grouping (Witteveen 2015, 2019). Much less attention has been paid to the biological adequacy and significance of the ways in which plant *traits* are identified, described and named within and across species, especially in relation to their local environmental and cultural context (Minelli 2019). Within plant and agricultural science, this is a salient gap given the enormous variation typically found across specimens belonging to the same species (as evident in the plethora of names used to refer to sub-species clusters, which include ecotypes, variants, landraces, cultivars, accessions, strains, and forms—among others), the diverse origins of such variation, and the relevance of such variation to how plants are used by humans.

Gray nomenclature fills this gap by highlighting the relevance of folk taxonomies used to name plants traits around the world, including within indigenous and traditional communities as well as farmers, breeders, and local markets.[5] Organizations such as the Taxonomic Databases Working Group, or Biodiversity Information Standards, do sterling work in evaluating possible challenges and alternatives to the existing Linnaean nomenclature and adapting the system to new observations emerging from heterogeneous sources. Their goal is however, by and large, to disambiguate reference to species, thereby still seeking to produce a widely accepted, common semantic system under which species names can be agreed and understood by all relevant expert communities and fruitfully applied to different types of crops.[6] In a similar spirit, efforts geared towards taxonomies of plants grown by humans (such as the International Code of Nomenclature for Cultivated Plants, or Cultivated Plant Code) also include a firm commitment to using Linnaean nomenclature as a reference point (Brickell et al. 2016).

This commitment is not necessarily compatible with the aspiration to acknowledge the various forms of diversity characterizing the description of plant traits, rather than species, and ensure that such diversity is reflected in the naming systems used to inform plant data infrastructures. A key concern is the *biological diversity* characterizing the organisms in question,

---

onment, often including its causal and functional role within a broader biological system.

[5] For a detailed discussion of the significance of including folk taxonomies into gray nomenclature, see Kendig 2020.

[6] For a critical discussion of taxonomies geared towards consensus, see Sterner, Witteveen and Franz (2020).

which is manifested in the extensive variability of the traits to be described, and the myriad ways in which such traits may respond and adapt to different and ever-changing environments (Farnham 2007). Understanding plant traits in the light of species-level classifications does not capture the rich differences among sub-species groupings of crops, whose morphological characteristics can vary so widely as to prompt different human uses of those variants (e.g., as cooking ingredients, forage for animals, building materials, starting points for pharmaceutical research) and different methods to research them (including different ways of selecting and storing germplasm samples). As emphasized by a recent review, "the world over, farmer and cultivator groups have selected and bred crop varieties, especially in tropical areas, which have not entered gene banks and are much more dynamic and constantly evolving. Documenting and databasing such information has challenges in effective integrating wild relatives, landraces, cultivars, vernacular names, cultivation practices and crop traits into a consistent taxonomic backbone" (Rajagopal et al. 2017). Indeed, the diversity of traits associated to any one species is of capital interest to breeders, whose key task is to identify and promote traits of specific interest to the environments, cultures and markets within which they operate. Beyond the challenge posed by biodiversity, there is *cultural diversity* in the motives, incentives and approaches to plants across the globe and across history, as mediated through the variety of expertise brought to the study of plants and related attitudes to plants as resources for human survival (such as food cultures or medicinal uses), which of course shaped the very evolution of plant life to fit human consumption; and, relatedly, *scientific diversity* in the data practices and research perspectives on biological and human diversity—and their interrelations—that underpin the development, management and use of plant data infrastructures.

These forms of diversity have long marred attempts to identify 'standard' or 'universal' names for plant traits (Nazarea 1988; Brush 2004; Franz and Thau 2010). Data of relevance to plant research are produced by groups with diverging goals and different cultures of data generation and exchange, which may include not only biologists but also nutritionists, agrotech and pharmaceutical businesses, farming communities, consumer groups and those engaged in conservation efforts. While some of these stakeholders are content with Linnaean taxonomy as an adequate system for naming plant traits, many others employ different and not obviously overlapping semantic systems inspired by traditional knowledge, existing practices of seed and germplasm banking, crop trade and the transnational regulations overseeing the circulation of plant genetic materials. In other words, the biological, cultural and scientific diversity characterizing this domain give rise to *semantic diversity* in the ways in which plant traits are named. While species-oriented classifications such as Linnaean nomenclature continue to play a seminal function in attempts to order and retrieve data on the best standardized varieties of major crops, they are less helpful in capturing data relating to fine-grained, locally contextualized forms of diversity, which are however fundamental to future uses of plants within human societies.

Hence semantic diversity in plant trait naming systems is not only an epistemic problem caused by the *quantity* of data and knowledge available thanks to digital technologies. It is also an epistemic issue with the *quality* of the data and knowledge incorporated into data infrastructures, and more specifically with the ways in which such quality is evaluated.

Ethnobotanical research is fostering understanding of semantic systems used within indigenous knowledge, thus contributing "systematic documentation—or memory banking—of indigenous practices of local farmers associated with traditional varieties of staple and supplementary crops" (Nazarea 1998, 5). Such documentation is of enormous value to scientific efforts of understanding biodiversity and its potential for supporting food security (Murphy 2007); and yet, the disparity and lack of connections across naming systems makes it hard to consider such data on a par with data coming from other sources, including scientific labs and field experiments

(Frison 2018). This situation reflects concerns around what forms of biological, cultural and scientific diversity are being captured and privileged which are common to many areas of biology, but are particularly pronounced in plant science given the entrenched focus on high-yield breeds and monocultures promoted by the so-called Green Revolution (Scott 1999; Kloppenburg 2004; Bonneuil 2016), and the related legacy of colonial models of agricultural development that give the superiority of Western, scientific approaches for granted (Benjamin 2009; Ludwig 2016). In particular the central role played by genetics in the determination of what constitutes a *valuable* trait, and for which purposes, cannot be underestimated.[7]

Accordingly, much of the discourse around plant research and the governance of plant genetic resources is grounded on the assumption that the Global North is gene-poor but technology-rich, while the Global South is gene-rich but technology-poor; and yet, "it is only by conceiving technology through a narrow lens as a system of doing that is generated, validated, and legitimised by Western science that the above equation can be sustained" (Nazarea 1988, 117). Developments in the technologies and governance of data exchange around the world, not least due to the increasingly digitalization of plant data, may offer a way to overcome the existing monism around 'best naming practices' by making semantic diversity more visible and reducing the problems it creates towards circulating and mining data. A starting point is the opportunity offered by such technologies to place different naming systems in relation with each other without necessarily needing to incorporate them all into a common standard, thus preserving the specific features of each approach (including its history and fit to particular biological entities and scientific goals). Today data management experts place emphasis on understanding data integration as an exercise in linkage rather than unification of data sources: in other words, integration is not about bringing all data together as a single body of evidence, but rather it is about *making it possible to link and compare datasets reliably and efficiently* (Coppens et al. 2017; Williamson et al. 2021). Data linkage understood in this way is crucial to any attempt to visualise and access not only data, but also the information required to mine, model and interpret data reliably (so-called metadata). The emphasis on data linkage aims to take account of the diverse perspectives underpinning data generation and processing, making discussions around how to link data semantics into useful platforms for different forms of plant expertise to confront each other and discuss ways to ensure enough compatibility among the goals and assumptions of respective data sources and formats as to enable meaningful comparison. Algorithms, infrastructures and classification systems geared towards data linkage aim to facilitate inter-dataset searches and thereby the overarching interoperability of data resources (Williamson and Leonelli 2022). This in turn can open up ways to side-step existing knowledge hierarchies such as those structuring the field of taxonomy.[8]

In what follows, I explore such potential through detailed discussion of a concrete case, which is the naming practices associated to the collection of data on cassava. I argue that the shift to digital systems of data management focused on contextualising information about crops, combined with a strong emphasis on engagement across several stakeholders, has produced a semantic system with the potential to link different perspectives on the crop—and related forms of diversity—in ways that are productive for all involved. As I shall point out, this has involved

---

[7] While I am unable to examine the prominence of genetic approaches to the evaluation of crop value within the scope of this paper, historians of biology and agriculture have provided substantive documentation of this phenomenon and of its significance for contemporary plant and agricultural research (see Scott 1999; Harwood 2012; Saraiva 2013; Fullilove 2017; and Curry 2019, among many others).

[8] In this respect, this paper builds on the understanding of taxonomy as an information science pioneered by Kendig and Witteveen (2020).

a shift of emphasis from attempting to name plant traits to capturing the relation between traits and their (biological, cultural and scientific) contexts.

3 Challenges from the Field: Using IPGRI Descriptors for Data Collection

The puzzles created by the plurality of semantic systems associated to plant traits, rather than species, is particularly evident within contemporary phenomics, a field dedicated to the analysis of trait data with the aim to study development and gene-environment interactions. Phenomics is typically characterised as the measurement and comparison of organism morphology under varying genetic and/or environmental conditions. Within plant science, it is heir to the study of plant morphology long associated with botany and phenotypic taxonomy, which it complements with insights from contemporary work on plant structures and functions (including molecular and cellular levels of organisation). By bringing together results from laboratory and field studies, phenomics aims to broaden its focus "from the initial characterization of single-plant traits in controlled conditions towards 'real-life' applications of robust field techniques in plant plots and canopies" (Walter et al. 2015). A key aim for phenomics research is to increase the yield and efficiency of plants growing in changing environments (e.g., in drought conditions), thereby helping to tailor plant traits to specific environments of growth and contributing to the development of precision agriculture (Coppens et al. 2017).[9]

To this aim, phenomics builds on data arising from field trials, glasshouses, and laboratory work on living organisms, which document not only their immediately visible features, but also traits only accessible through intervention and specialised imaging techniques (such as tissues, proteins, metabolic pathways) as well as environmental factors of relevance to plant development (such as data on the soil, climate, other organisms and microbiomes with which plants interact). Such complex data are also triangulated with data extracted from specimens kept by seed and germline banks, such as digital sequence information. The study of plant phenomics thus involves enormously complex efforts of data collection and analysis, where plant traits need to be broadly and multiply defined to match the diversity of materials, goals and methods encompassed by the data. There are countless parameters of potential relevance ranging from the information about the soil, relevant microbiomes, plants at different stages of development, changing climatic conditions and so forth; and no universal approaches to identifying and labelling relevant traits.

One of the most influential efforts to develop a consistent naming system for plant traits to date has been the descriptor system developed by the CGIAR (and particularly the International Plant Genetic Resources Institute, now known as Bioversity) in collaboration with FAO (Bioversity International 2007).[10] This system, widely known as IPGRI descriptors, was initiated in the 1980s as a separate system to traditional Linnaean taxonomy, whose aim was to help govern the conservation and transnational movement of plant genetic materials for specific human uses (as exemplified by seed banks and trade around plant varieties, respectively) through the standardization of the ways researchers would refer to the morphology of plants (Gotor et al. 2008; Curry 2018; Curry and Leonelli forthcoming).[11] Its most straightforward

---

[9] For a pointed critique of this approach, see Miles (2019).

[10] This system is flanked by other, complementary systems for crop description focusing on specific types of traits, including for instance descriptors for genetic resources held in gene banks, for crop wild relatives *in situ* and for farmers' knowledge of plants (Biodiversity International 2007, 2021).

[11] Indeed, FAO provides a general definition of descriptors that matches the emphasis on plants—and particularly crops—as genetic resources: "providing an international format and a universally understood language for plant genetic resources data […] targeted at farmers, curators, breeders, scientists and users and facilitate the ex-

initial application was in the collection of phenomic data from field trials conducted by CGIAR institutes around the world. This was gradually expanded to trials conducted by other research institutes, within both public and private institutions; and starting from 2001; the descriptors were integrated into the Passport system regulating the transnational exchange of plant genetic materials, thus establishing itself as an important tool for trade and for the exchange of data of commercial significance (Gotor et al. 2008).

As a specific example of how IPGRI descriptors work in practice, consider the descriptors developed to annotate data on cassava, a portion of which is reported in figure 1. They consist of a list of characters (traits) with associated attributes (e.g., color). Each attribute is assigned a number, which makes it possible to quantify traits by convention. When considering a trait such as "color of the stem cortex", for instance, researchers have a choice between the following attributes, each of which reflects a different stage in the development of cassava: 1–light yellow; 2–light green; 3–green; and 4–dark green. By choosing and noting "3", researchers assign a numerical value to what would otherwise be a quality of the crop, by resorting to a visual assessment (which in turn relies on their familiarity with the range of colors that a cassava stem typically assumes during its life cycle) and a standardized association between that assessment and a numerical symbol. These qualitative-turned-quantitative assessments can thus complement—and be analyzed together with—more straightforwardly quantitative measurements such as the length of the stem and the width and number of the roots. The numbers produced using trait descriptors can also facilitate the analysis of imaging data, such as photographs of crops at various stages of development, thereby supporting the broader aims of phenomics.

Given their apparent simplicity and applicability to the collection of data from field trials, it may seem that using IPGRI descriptors presents no significant epistemic challenges. After all, there is no mistaking what a stem or a leaf is, and it seems clear from the descriptor table what each number/trait combination is supposed to represent. However, as I discuss in the remainder of this section, those engaged in this practice in the field have encountered various challenges requiring highly localized decision-making processes, which are tightly connected to difficulties in handling the forms of biodiversity I discussed in the previous section.

The first challenge stems from biological diversity, and more specifically from the *observed variability across crop specimens*. Researchers have difficulties in determining: what counts as an individual or a relevant group (sometimes called 'variety', 'accession', 'ecotype', or 'strain'), especially in cases of clonal reproduction; what traits should be regarded as representative of such individual/group, particularly given the enormous variation in phenotypes depending on soil, climate and nurture; and which individual plants could be considered to be representative of a plot—a decision typically taken by each data collector in the field on the basis of their own experience and judgement.[12] Consider the problems posed by the extraction of data from a field as in the case of the IITA cassava trial pictured in figure 2. This photograph shows the stage of the trial in which researchers have excavated most roots from the ground and are preparing for data collection. Given the number of plants involved, it is not possible to precisely excavate all the cassava roots grown within the trial; and not all excavated roots can be subjected to close evaluation and measurement. Researchers and their assistants thus constantly make decisions around which roots to use as representative samples and how to set the boundaries around any one plot of land, to ensure that the plants grown for one trial are not confused with the plants grown for another. Because the descriptor system does not document such decisions and the extent to which they may diverge, measurements taken by different people at different locations

---

change and use of resources" (FAO website, accessed August 2018).

[12] For a relevant discussion of the complexity of describing individuals in the realm of ecology, see Trappes (forthcoming).

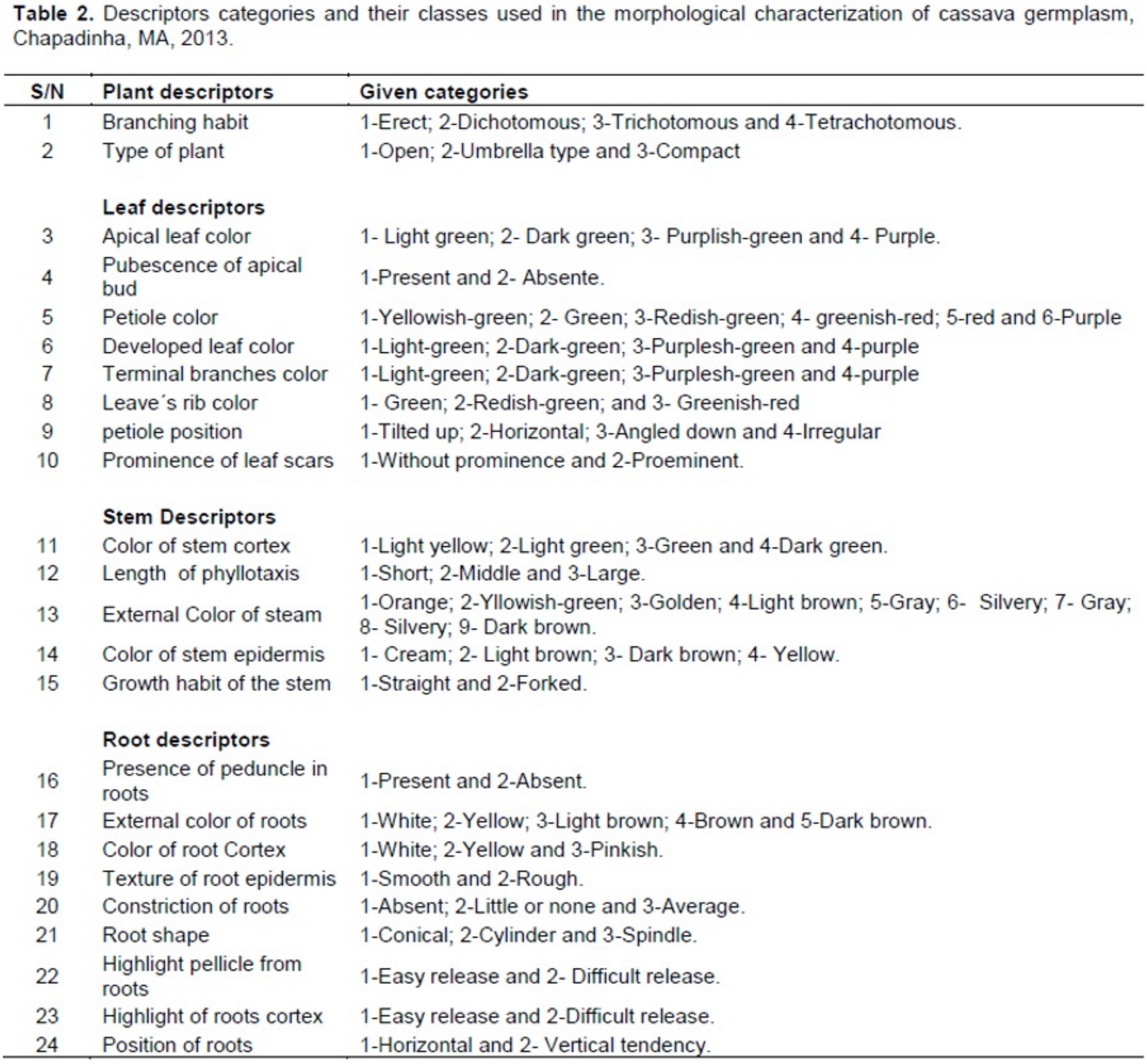


Table 2. Descriptors categories and their classes used in the morphological characterization of cassava germplasm, Chapadinha, MA, 2013.

S/N	Plant descriptors	Given categories
1	Branching habit	1-Erect; 2-Dichotomous; 3-Trichotomous and 4-Tetrachotomous.
2	Type of plant	1-Open; 2-Umbrella type and 3-Compact
	Leaf descriptors	
3	Apical leaf color	1- Light green; 2- Dark green; 3- Purplish-green and 4- Purple.
4	Pubescence of apical bud	1-Present and 2- Absente.
5	Petiole color	1-Yellowish-green; 2- Green; 3-Redish-green; 4- greenish-red; 5-red and 6-Purple
6	Developed leaf color	1-Light-green; 2-Dark-green; 3-Purplesh-green and 4-purple
7	Terminal branches color	1-Light-green; 2-Dark-green; 3-Purplesh-green and 4-purple
8	Leave´s rib color	1- Green; 2-Redish-green; and 3- Greenish-red
9	petiole position	1-Tilted up; 2-Horizontal; 3-Angled down and 4-Irregular
10	Prominence of leaf scars	1-Without prominence and 2-Proeminent.
	Stem Descriptors	
11	Color of stem cortex	1-Light yellow; 2-Light green; 3-Green and 4-Dark green.
12	Length of phyllotaxis	1-Short; 2-Middle and 3-Large.
13	External Color of steam	1-Orange; 2-Yllowish-green; 3-Golden; 4-Light brown; 5-Gray; 6- Silvery; 7- Gray; 8- Silvery; 9- Dark brown.
14	Color of stem epidermis	1- Cream; 2- Light brown; 3- Dark brown; 4- Yellow.
15	Growth habit of the stem	1-Straight and 2-Forked.
	Root descriptors	
16	Presence of peduncle in roots	1-Present and 2-Absent.
17	External color of roots	1-White; 2-Yellow; 3-Light brown; 4-Brown and 5-Dark brown.
18	Color of root Cortex	1-White; 2-Yellow and 3-Pinkish.
19	Texture of root epidermis	1-Smooth and 2-Rough.
20	Constriction of roots	1-Absent; 2-Little or none and 3-Average.
21	Root shape	1-Conical; 2-Cylinder and 3-Spindle.
22	Highlight pellicle from roots	1-Easy release and 2- Difficult release.
23	Highlight of roots cortex	1-Easy release and 2-Difficult release.
24	Position of roots	1-Horizontal and 2- Vertical tendency.

Figure 1: Extract from IPGRI trait descriptors for cassava.

can be uneven, leading to unreliable comparisons across the resulting datasets.

This brings me to a second challenge associated to the use of trait descriptors, which relates to what I called scientific diversity: more specifically, the *diversity of skills and measuring methods* used for traits data collection across locations. When confronted with circumstances of data collection such as pictured in figure 2, researchers need to make practical decisions around how to count the leaves and flowers on the available plants. This typically involves making an estimate rather than counting every single item, given their abundance and the need to collect data relatively quickly to keep up with plant growth and the pace of research (figure 3). Evaluation of qualities such as the color, firmness and consistency of the cassava roots are made through recourse to the senses, particularly vision and touch, in ways that can vary widely depending on who is carrying out the assessment (figure 4). And any attempt to quantify the length of roots and stems depends on how plants are harvested, with technicians adopting different approaches. For instance, there can be different ways to cut the stem away from the rest of the plant, resulting in a stem being measured as longer and the root system being shorter and lighter due to the weight and length assigned to the stem, or vice versa (shorter stem and longer/heavier roots). When participating in debates among technicians, researchers and breeders at IITA and else-

Figure 2: Cassava field trial at IITA, Ibadan, July 2017. Copyright of the author.

where, I witnessed several disagreements over what counts as a 'standard cut'. This was being explicitly debated due to the difficulties that may arise when comparing stem length across field trials that take place in different locations, where habit and tacit knowledge around harvest may differ considerably.

Another challenge lies in the *temporality of data collection* and demonstrates the tight interrelation between biological and scientific diversity, with scientific methods construed to match as closely as possible the variability of the biological entities being studied. The time of flowering and harvest can vary considerably within and across cassava trials, making it impossible to standardize data collection and requiring researchers to continuously monitor the fields. Again, researchers need to make situated decisions about when to measure plants, which depend on their familiarity with the growth patterns of the varieties at hand as well as of the environment and soil in which plants are developing—and the circumstances under which such decisions are made are not documented by IPGRI descriptors. This matters because depending on when data collection is carried out, the quality of the relevant traits may change considerably: a stem cortex may shift from light green to green, for instance—and thus be measured as 3 instead of 2.

Last but not least, there is the issue of *which traits to evaluate and measure* in the first place, which raises the question of which traits matter to whom, and for which purpose—thus bringing in concerns around cultural diversity. For example, the consistency and color of cassava pulp (figure 4) are valuable traits for local markets and consumption, since these are the qualities that prospective consumers look for. These same traits are however not considered relevant for taxonomic classification, and are therefore left aside in scientific evaluations that privilege genomic sequencing as the gold standard for establishing significant differences among variants. It is

Figure 3: Try counting leaves and flowers here, for instance. Copyright of the author.

not uncommon for two types of cassava roots, well-recognized within local markets as having different taste and therefore as suited to forms of cooking, to be declared one and the same because they have similar genetic markers—a taxonomic decision that does not fit the experience and observations of breeders, farmers and consumers of cassava, and can be problematic when used to inform future field trials. To avoid such situations, researchers working on the ground in Nigeria place great emphasis on the need to consult farmer and breeder communities to discuss which cassava traits—whether or not they are widely recognized as 'scientifically relevant'—should be privileged when investigating whether and how a given variant tolerates drought or specific pathogens. This is especially important since a cassava field trial can last up to seven years, and therefore represents a significant investment not just for plant researchers but also for communities using these crops as their food staples.

To sum up, the value of IPGRI descriptors lies in their being fixed reference points, hierarchically organized through a stable structure, and narrowly focused on measuring plant traits as context-independent entities. This structural stability and narrow focus is what makes them effective standards and benchmarks for researchers of different backgrounds looking to identify a given variety and validate its taxonomy before entering it into in situ or in vitro collections. At the same time, this very stability and focus prevent these descriptors from being able to capture: the biological diversity exhibited in the countless, variously adapted and constantly evolving forms of plant life, the scientific diversity in the methods and skills used by data collectors responsible for measuring and implementing descriptors in the field, and the cultural diversity manifested in existing ideas around what constitutes a valuable trait. IPGRI descriptors are therefore of limited use to researchers studying plant environmental responses and breeders aim-

Figure 4: A researcher showing how data on the colour and consistency of the cassava pulp are collected: that is, by peeling off the cortex to uncover the pulp and pressing on it to evaluate its firmness.

ing to test crop varieties in multi-site evaluations and under different environmental conditions and management practices.

4 Making Diversity Matter: The Crop Ontology

The above-mentioned challenges with the IPGRI trait descriptors point to the long-standing general problem of instituting global standards for local, situated procedures characterized by biological, cultural and scientific diversity—resulting in extensive semantic diversity in the classifications and descriptions used within naming practices (Bowker 2006). This problem needs to be confronted by any taxonomic effort, but is particularly pernicious in the case of crops with multiple users, audiences, goals, and high levels of biological and environmental variability. I shall now consider an attempt to improve upon IPGRI descriptors through the intelligent use of data technologies (and particularly computational ontologies, which support complex relations among entities that make it easier to define traits contextually compared to traditional taxonomies)[13] combined with extensive community engagement. This is the effort carried out by the Crop Ontology system of plant data semantics, which aims to collect and link data in ways that are scientifically significant *and* valuable to local communities in ways that manage multiple forms of diversity (as discussed in section 2) and consequential challenges for data collection in the field (as described in section 3).

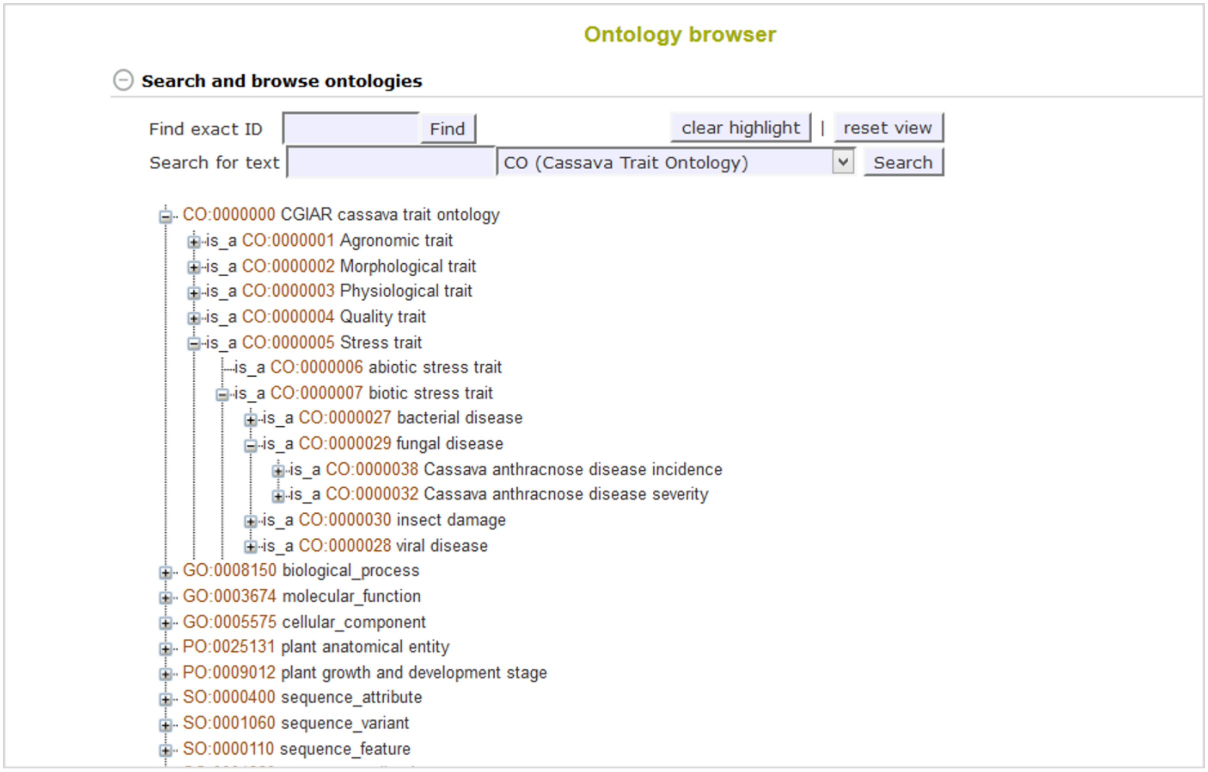
The Crop Ontology is a digital infrastructure developed over the last decade by an international network initiated by CGIAR and coordinated by researchers in Montpellier. It was initially part of the multi-partner Next Generation Challenge Programme (GCP) set up to address the standardization needs of the newly introduced digital fieldbooks and the Integrated Breeding platform devised to boost a new generation of breeding using genetic markers. Broadly, the Crop Ontology aims to facilitate the sharing of information between plant scientists working in laboratories, experimental farms and plant stations in Europe, the United States, South America, Sub-Saharan Africa and various Eastern countries (e.g., Malaysia and Thailand). Its approach is to standardize the representation of crop traits and trait-attributes across crop species and databases, while paying attention to the peculiarities of different crops and related systems of knowledge, and thereby brokering the epistemic cultures involved in the collection and evaluation of the resulting data (Shrestha et al. 2012; Leonelli 2022). The Crop Ontology may more accurately be described as a federation of partly overlapping semantic systems, since it encompasses several data infrastructures, each of which is finely tailored towards data on a specific crop—thus recognising the differences in biology, methods and cultures surrounding each species. As one of its users put it to me, the Crop Ontology is acting as a 'regulatory body' (R_12_B) over a multitude of organism-dependent standards, which involves the difficult task of calibrating the general approach and structure of the data system to the needs and characteristics of each crop. These efforts started with reference to crop-specific traits for chickpea (*Cicer arietinum*), maize (*Zea mays*), potato (*Solanum tuberosum*), rice (*Oryza sativa*), sorghum (*Sorghum* spp.) and wheat (*Triticum* spp.), with banana (*Musa* spp.) and maize added by 2010. From 2016 it has been expanded to cassava, yam, and other crops whose prevalent consumption is in the Global South. In what follows, I consider the development of the Crop Ontology in relation to cassava and in collaboration with CassavaBase, an 'open' database used to store and share data from field trials carried out in West Africa, South America and Southeast Asia (Fernandez-Pozo et al. 2015).[14] The screenshots of CassavaBase and the Crop Ontology used for my analysis below pertain to the 2020 versions, which have of course been subject to updates and improvements ever since.

The Crop Ontology functions by associating a descriptive nomenclature for plant traits, along the lines set out by IPGRI descriptors, with a unique identifier; and by embedding each term into a complex series of relations to other terms, so that for instance when looking for a specific cassava disease as a trait, it becomes immediately clear that the disease is fungal and that it is a form of biotic stress (figure 5). A tool such as the Crop Ontology is designed to support the digitalisation of the phenotyping workflow: in other words, it enables researchers to collect data and metadata directly from the field via a standardised interface on their mobile phones. Work on the app layout and trait file is continuously updated in response to feedback from the field, resulting in user-friendly fieldbook that can be easily utilized by data collectors

---

[13] For a discussion of the peculiarities of computational ontologies as systems for data classification, see Leonelli (2012, 2016) and Franz and Sterner (2018).

[14] CassavaBase is not the only international database developed in the last fifteen years that is devoted to cassava data. Among other relevant databases are the Cassava Genome Hub (https://www.cassavagenome.org/) and the Cassava Online Archive (http://cassava.psc.riken.jp/), whose history and relations to the Crop Ontology and CassavaBase I don't have the scope to discuss here.



CO:0000000 CGIAR cassava trait ontology
- is_a CO:0000001 Agronomic trait
- is_a CO:0000002 Morphological trait
- is_a CO:0000003 Physiological trait
- is_a CO:0000004 Quality trait
- is_a CO:0000005 Stress trait
  - is_a CO:0000006 abiotic stress trait
  - is_a CO:0000007 biotic stress trait
    - is_a CO:0000027 bacterial disease
    - is_a CO:0000029 fungal disease
      - is_a CO:0000038 Cassava anthracnose disease incidence
      - is_a CO:0000032 Cassava anthracnose disease severity
    - is_a CO:0000030 insect damage
    - is_a CO:0000028 viral disease

GO:0008150 biological_process
GO:0003674 molecular_function
GO:0005575 cellular_component
PO:0025131 plant anatomical entity
PO:0009012 plant growth and development stage
SO:0000400 sequence_attribute
SO:0001060 sequence_variant
SO:0000110 sequence_feature

Figure 5: A snapshot of the Crop Ontology structure as it appeared to users accessing it from the CassavaBase interface in 2020.

on the field. In this sense alone, the Crop Ontology represents an obvious improvement over labour-intensive manual processes such as those required by IPGRI descriptors.[15]

What makes the Crop Ontology into a radical improvement over descriptors, however, is its ability to record and track the the diversity of tools, terminologies and variables used to describe widely diverse crop traits in different parts of the world, as well as to the differences in expectations, goals and working conditions among researchers, technicians and breeders involved in efforts of data collection. Computational ontologies more generally have proven to be versatile and flexible tools not only to capture data pertaining to biological entities, but also and most importantly contextual information about the provenance and history of such data and related entities—a crucial advantage of such a relational, digital system over traditional list-based taxonomies (Leonelli 2016). A key insight guiding this work and providing a way to field its complexity is the idea that tracking the history of particular datasets is just as significant as being able to access that dataset, since knowledge of the history of the data is essential to their interpretation and re-use (see also Leonelli 2020). As one of the database curators put it to me, "If you get an accession, you should trace its history, get its attributes, in which trials it has been used and its performance in the trials at every level. Quality, agrobiotics, stresses. All information should be linked to accession identifier". To understand how this works, we need to have a slightly more detailed look at the ways Crop Ontology terms are related to each other.

What the Crop Ontology proposes to do is to link the naming and quantification of traits with information about the methods and skills used, so that future users of the resulting data can reconstruct the conditions under which such assessments took place (Shrestha et al. 2010). To

---

[15] My assessment of the scientific value of the Crop Ontology is not focused on the validity and adequacy of specific terms within it, which is not for me to adjudicate, but rather on the robustness of their procedures and general approach to trait naming, which I find philosophically significant and generative as I argue below.

this aim, Crop Ontology terms—the plant traits in question—are organised in acyclic graphs by differentiating between *trait definition* (whose origins are described through the relation "derived_from"), *measurement method* (indicated through the relation "method_of") and *scale* (indicated through the relation "scale_of" and describing "how the trait observation is expressed"; (Pietragalla et al. 2022, 15).[16] This approach gives the Crop Ontology an ability to deal with variation in approaches and skills applied to estimates and measures (Shrestha et al. 2012). We have seen how some field observation cannot be easily recorded, as there may be different methods to do—in this case, the specificity of the measurement approach is signaled on the ontology. Consider again the nuances of color assigned to a given trait, such as the multiple shades of yellow color of the cassava root pulp. Within the Crop Ontology is it well-recognized that such nuances are significant since they are used to assess the taste, quality and market value of cassava roots for local consumption. To capture them, the ontology does not simply assign the value "yellow" to the trait "root surface color", but rather complements this assessment with information about the methods and scale of measurement used by data collectors to discern that color, including a specification of whether the root was evaluated in the field of in storage conditions (figures 6, 7 and 8).

In the words of one of the Crop Ontology developers: "Ontologies need to capture everything people are doing, all the methods, there is no wrong or right way" (PI_17_A). The system thus tries to encompass and order as much contextual information as possible, thereby providing structured access to data about the biological, scientific as well as cultural diversity characterizing the crops in question, including the ways in which such diversity may change over time. In this way, the Crop Ontology is in a position to effectively document diversity and make it into an object of study for its users, rather than being forced to capture such rich information through a fixed, context-independent and narrowly focused set of terms as in the case of IPGRI descriptors.

Cultural diversity is particularly treasured, with Crop Ontology developers placing strong emphasis on regular dialogue among stakeholders beyond plant science, particularly breeders and farmers. This was something that Crop Ontology director Elizabeth Arnaud fought hard for since the start of her work on the system in 2008; and it remains an exception rather than the rule for plant data collection, given the general tendency in this domain to privilege efficient and highly standardized communication over attention to local customs and cultural perspectives on crops.[17] In the case of cassava, this has been facilitated by close collaboration with CassavaBase developers Afolabi Agbona and colleagues, some of whom are Nigerian themselves and locally based at IITA. As the curator in charge, Agbona has played a fundamental role in updating the Cassava Ontology—efforts bolstered by IITA contributions to the broader BreedBase platform, as well as by international contributors from various universities (including Cornell), corsortia (such as the Quinoa Phenotyping Consortium) and private companies (including PepsiCo). Many key terms have been chosen and calibrated through open communication with breeders, which was achieved through yearly farm visits by the database developers and their collaborators, as well as Cassava breeders' meetings and training sessions. Discussions at such meetings

---

[16] Scale is a particularly interesting and innovative variable considered within the Crop Ontology, since it explicitly documents the different units of measurements that data collectors use in reporting information about plant traits. As specified by the Crop Ontology guidelines: "when the observation is expressed by a quantitative value, the CO scale described the unit. Alternatively, when the observation is expressed by predefined categories, the CO scale describes the possible values and their meaning" (Pietragalla et al. 2022, 15). The Crop Ontology admits of several scale classes, including date/timestamps, duration, nominal, numerical, ordinal, text-based and code-based (for exceptionally complex traits: Pietragalla et al. 2022, 16).

[17] For details on the history of the Crop Ontology and its relation to broader, transnational plant research efforts, see Leonelli (2022).

Figure 6: Researchers collecting data on the colour and consistency of the cassava root pulp—and the skills used to discern and record those—through the CassavaBase app.

include debate over which traits are viewed as more valuable and for which purposes, and consideration of practical problems with specific trait names and assessment methods. For instance, a debate over "what is shoot weight and how should it be measured" was resolved in favor of measuring the shoot "without stem", i.e., to avoid including any part of the stem into the weight of the shoot. This decision benefitted from the feedback of breeders who had strong views on the implications for how field workers cut the shoot from the stem when preparing it for data collection—and how shoots and stems would be measured in future field trials. Another example is the insistence from breeders that traits specific to Nigerian ways of cooking cassava be inserted into the ontology, given the significance of such traits for the local economy, even if these traits are not normally considered as having biological significance. Indeed, CassavaBase developers were involved in a broader initiative to develop a gender-sensitive approach to breeding strategies and evaluations, which included a gender-sensitive protocol for collecting trait preferences (highly significant given marked differences in trait preference between female and male breeders, with the former paying more attention to traits facilitating the processing of crops as food for human and animal consumption; see also Ashby and Polar 2019). This fuelled a rich discussion over which plant traits had a direct impact on food preparations, with "gari content" and "fufu content" identified as terms describing the quantity of gari and fufu (two staple Nigerian foods) that could be extracted from a given variety (figure 9). Even the assessment of how many roots could be marketable, and where, ended up being recorded as a "plant trait" within the Cassava Ontology (figure 10). Last but not least, though I do not have the scope to expand on this point here, interactions with cassava breeders at the local level

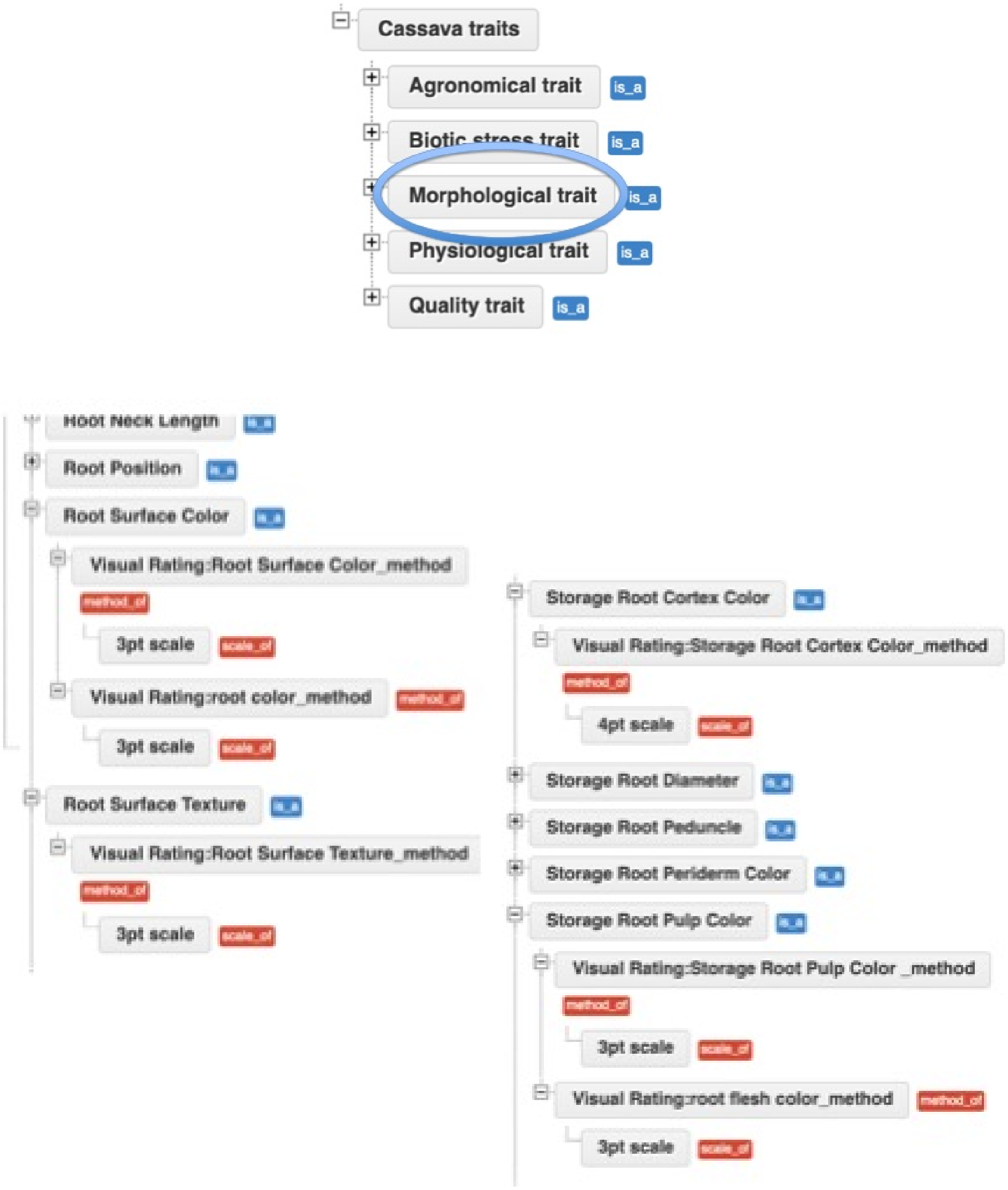
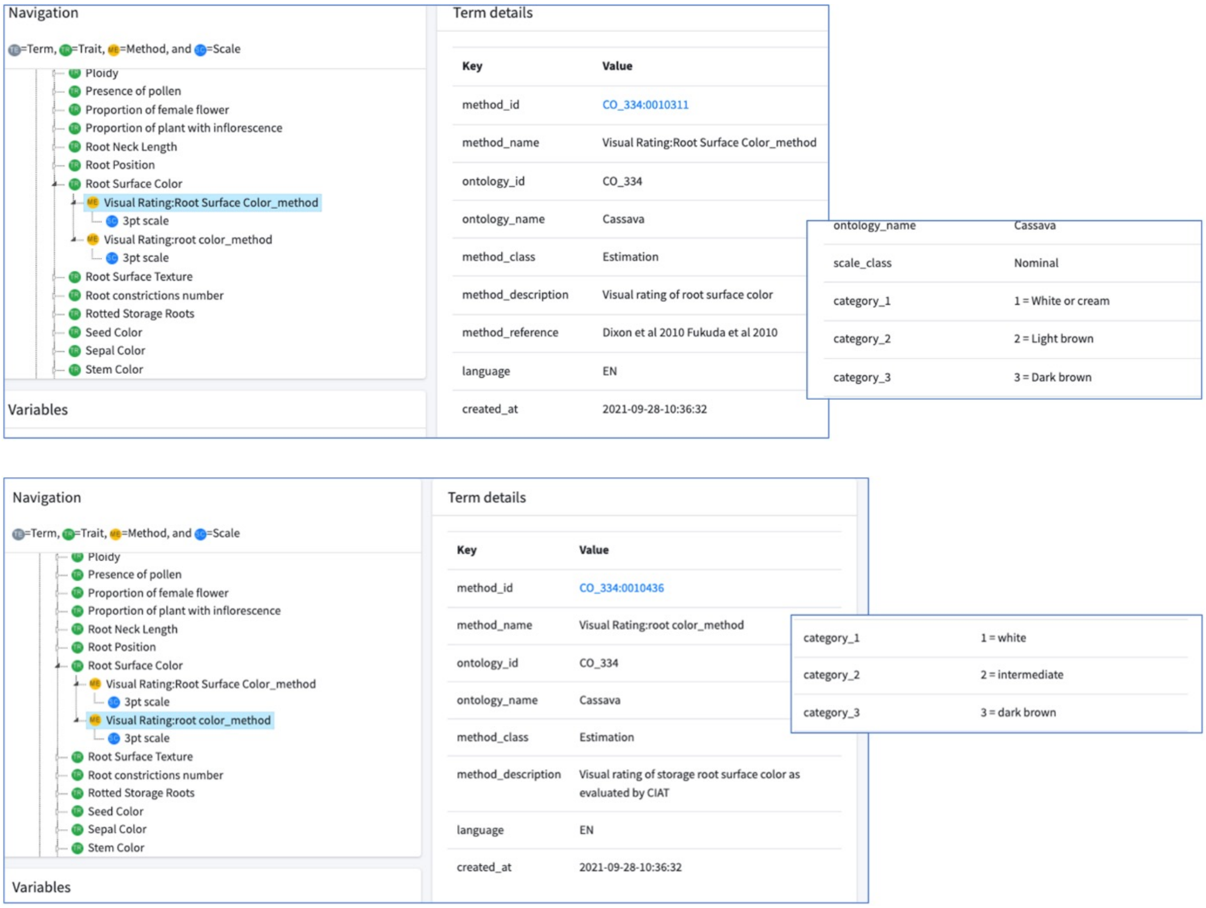


Figure 7: Extracts from the Crop Ontology, again accessed through the CassavaBase interface in 2020, dedicated to the morphological trait "root surface color", where the method of data collection is specified as "visual rating" (with further specification when clicking on that term) and the scale is specified as 3pt. This is usefully differentiated from the trait "storage root surface color", thus recognising that the characteristics of the root change depending on whether data are collected in the field or once the plant has been brought into a warehouse for storage.



Key	Value
method_id	CO_334:0010311
method_name	Visual Rating:Root Surface Color_method
ontology_id	CO_334
ontology_name	Cassava
method_class	Estimation
method_description	Visual rating of root surface color
method_reference	Dixon et al 2010 Fukuda et al 2010
language	EN
created_at	2021-09-28-10:36:32

ontology_name	Cassava
scale_class	Nominal
category_1	1 = White or cream
category_2	2 = Light brown
category_3	3 = Dark brown

Key	Value
method_id	CO_334:0010436
method_name	Visual Rating:root color_method
ontology_id	CO_334
ontology_name	Cassava
method_class	Estimation
method_description	Visual rating of storage root surface color as evaluated by CIAT
language	EN
created_at	2021-09-28-10:36:32

category_1	1 = white
category_2	2 = intermediate
category_3	3 = dark brown

Figure 8: Extracts from the Crop Ontology, now in its 2022 interface, showing the difference between two documented methods to measure root surface colour—one utilized by researchers IITA and one utilized by researchers at its sister institutes in Colombia, CIAT (International Centre for Tropical Agriculture).

helped to calibrate the English language terms used for these descriptors with the terms used in local languages, thus recognising linguistic diversity as another significant variable shaping plant descriptions from user to user.

While these local interactions helped to capture the biological and cultural diversity of cassava plants, Crop Ontology developers also worked with colleagues and stakeholders around the world to keep abreast of new developments in data technologies and standards, thus learning from the scientific and semantic diversity in approaches to plant data linkage. First there is the above-mentioned “regulatory” work required to align the structures and terms of the various ontologies developed in relation to each crop. Then there are the efforts to regularly cross-validate Crop Ontology terms with several other data platforms around the world, to avoid duplication and foster interoperability across and even beyond the plant data ecosystem (for instance, when considering broader challenges such as the relations between animal and plant breeding, or the impact of specific agricultural strategies on specific local ecologies and the human communities therein). For instance, the Crop Ontology helps to coordinate the Cassava Ontology with other international initiatives on cassava and other tubers, through participation in venues such as the Planteome.org project and the Agricultural Communities of Practice (Arnaud et al. 2020). The recent effort by CGIAR to construct an overarching data platform for the myriad forms of data collected at its institutes around the world (what they call ‘One CGIAR’) is but one example of the ambition to link data infrastructures with each other to guarantee easy access while at

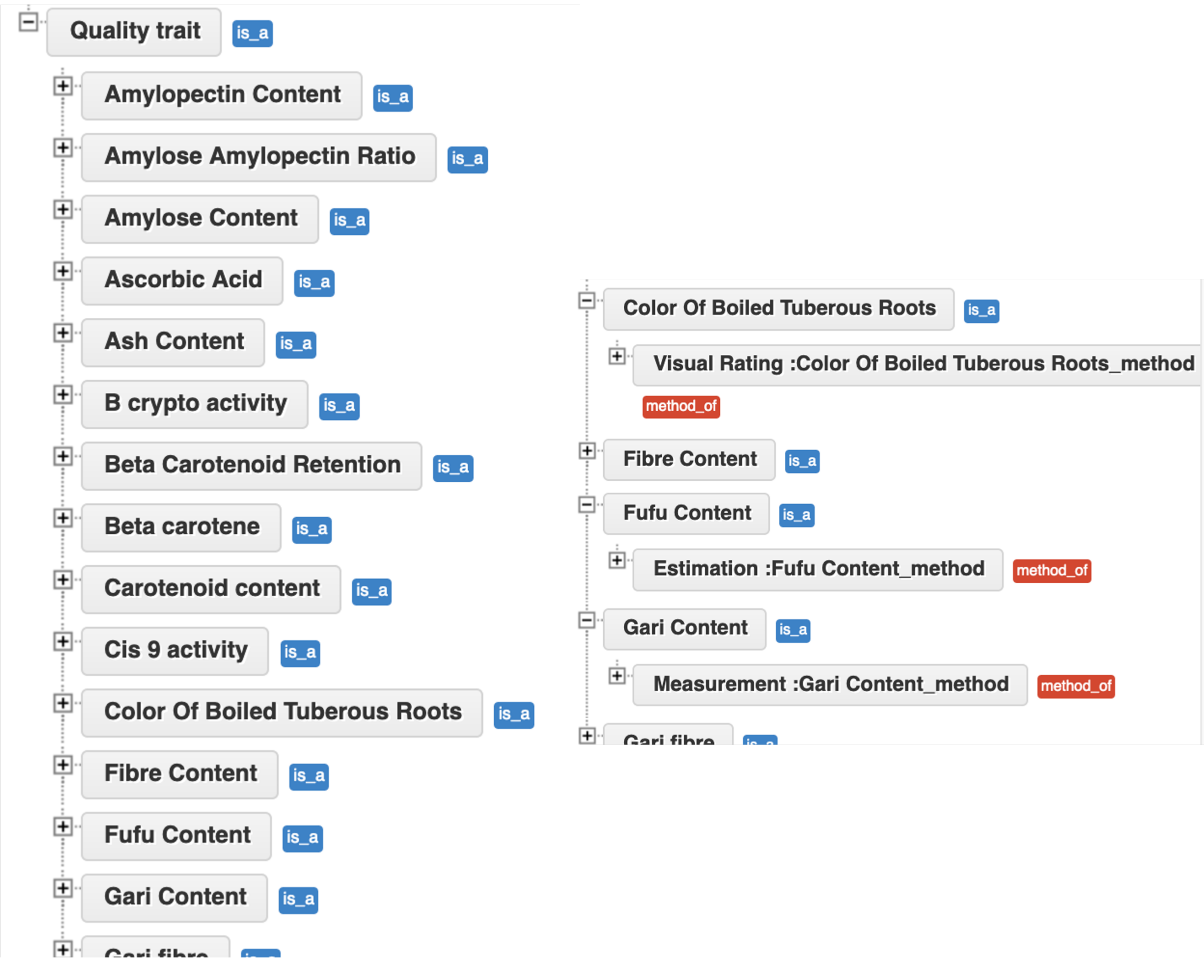


Figure 9: This part of the Crop Ontology, accessed in 2020, reports an assessment of the extent to which a given root is judged to be usable to prepare gari or fufu.

the same time retaining the system-specific, local elements that make such data actually legible and usable. This work is never-ending as research continues, and plants themselves evolve and change. As an informant put it to me, "at no point can the ontology be static": there are always traits to add and things to adjust, with the ontology responding to the challenges posed by the geographical and temporal specificities of biological entities and related methods of data collection.

Another consequence of the Crop Ontology's attention to various forms of diversity is the significant expansion of audiences for this kind of technical data work. The Crop Ontology makes itself accountable to a broad ensemble of stakeholders, going well beyond plant and data science to include those with an interest in the consumption, trading and conservation of plants. This is illustrated by an infographic produced in 2017 by the Research Data Alliance working group on agrosemantics, which included Crop Ontology developers (figure 11). The infographic identifies data semantics as a way to "reconcile points of view and data", including those of biologists, farmers and breeders but also nutritionists, chefs, food manufacturers, traders, information managers and even sociologists.

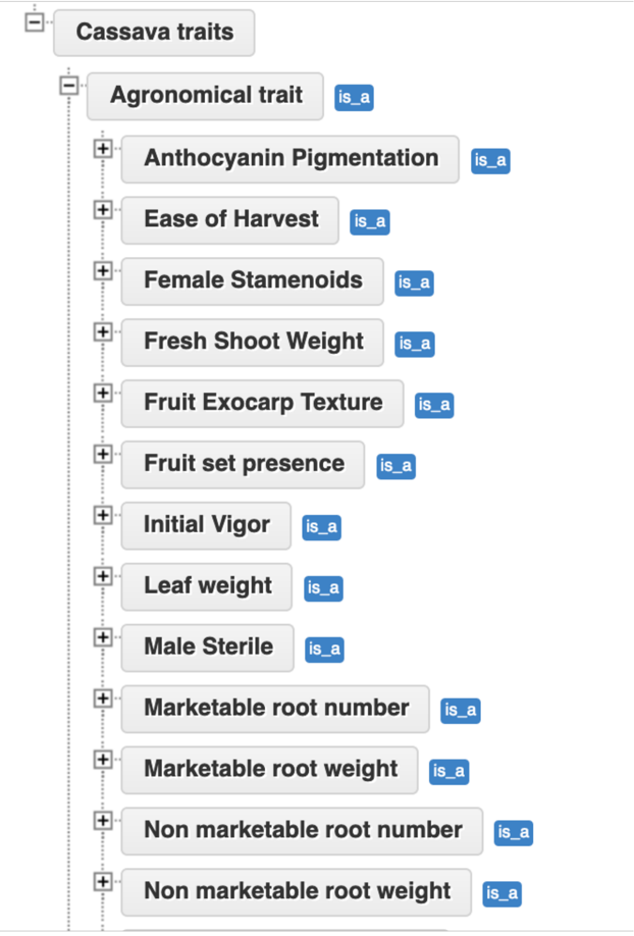


Figure 10: The inclusion of "marketable / non-marketable root number / weight" as agronomic traits within the Cassava Ontology, 2020 version.

5 Process-Sensitive Naming for Plant Data Linkage

> *It is, I think, a counterproductive waste of energy, from the point of view of user-sensitive agricultural development, to strive to design and promote the technology or the variety that is supposed to work or to be desirable under all (or almost all) circumstances. In the first place, farmer resistance and/or apathy will be considerable, if not overwhelming. Secondly, the ramifications could backfire in terms of loss of diversity—both genetic and cultural.*
>
> –Virginia Nazarea (1988, 70)

The naming practices of the Crop Ontology, as described in the previous section, do not relate in straightforward ways to traditional taxonomic practices. Here is a case where reliance on a universal approach to identifying and labelling traits has repeatedly proved problematic, and the attempt to articulate semantic differences is generating new ways to develop and communicate biological knowledge. As I interpret it, the key conceptual move underpinning the semantic work done by the Crop Ontology lies in shifting classificatory focus from *biological products* to *environmental processes*. In other words, the Crop Ontology is moving away from describing plant traits as they exist at the moment of measurement, which typically does not include attempts to document how that moment fits the plants' broader lifecycle and environment (including humans). It aims instead to capture the interactions between plants and environment that give rise to specific traits at particular times, including processes of cultivation and data collection required to contextualise the data produced when evaluating and measuring plant

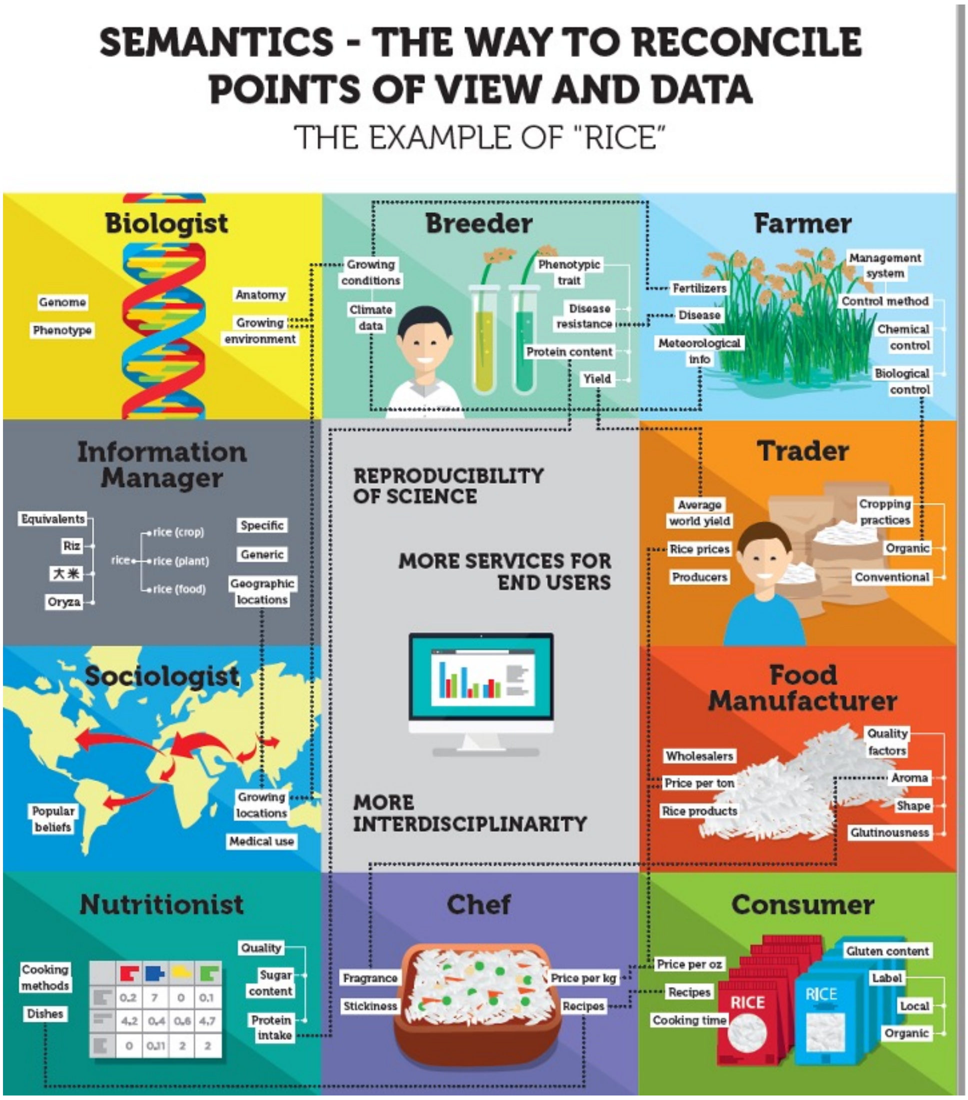


Figure 11: Illustration of the key stakeholders affected—and brought together—by plant trait semantic systems such as the Crop Ontology. CC-BY. Produced by the Research Data Alliance (RDA) Agrisemantics Working Group, 2017.

growth and responsiveness to stressors. Thus, the naming practices of the Crop Ontology focus on documenting the relation between traits and their surroundings, including their history and intersections with human intervention, rather than the traits in and of themselves.

This brings me to articulate what I take to be a key component of the Crop Ontology's success: its emphasis on what I call *process-sensitive naming*. This does not simply mean that those involved in naming practices take account of environmental processes when making decisions on how to label a given specimen: this is of course what taxonomists already do and what much of their expertise consists of. Rather, it means opening up the decision-making system underpinning naming practices, by providing as much information as possible to its users about why a given name has been chosen, by whom and under which circumstances—this effectively helping to 'coordinate dissent' (Sterner, Witteveen and Franz 2020). This undoubtedly places new demands on both data collectors and data users: data collectors need to think carefully about how they annotate information about their methods and the local ecosystem in which the trait has developed; data users need to take such information into account when evaluating the significance of data, particularly in comparison to other datasets; and collectors and users alike have to negotiate the semantic diversity reflected within data infrastructures. Given these difficulties, major efforts continue to be invested, by Crop Ontology developers and many other curators working with biological data, in structuring metadata in ways that are as user-friendly as possible, so as to not to put off researchers who would rather just access data points without having to worry about their provenance and the various ways in which they may have been grouped (Ćwiek-Kupczyńska 2016; Boumans and Leonelli 2020). The pay-off for such efforts, however, is significant. Such a system emphasises the plasticity of plant traits by giving a prominent place to the specificity and temporality of environmental interactions, above and beyond (though of course in connection to) gene-centred accounts of biological structure and function. This is a crucial conceptual move given the urgent need to better understand the long-term impact of environmental stressors (and particularly those related to human interventions over the last century) on the metabolism, development, ecology and evolution of life forms on this planet (Landecker under review). It also opens the way towards a relational approach to data, whereby data points cannot be considered separately from specific situations of inquiry—including ever-changing problem agendas, stakes and communities of reference (Leonelli 2016). This in turn can increase the accountability and trustworthiness of data systems used to document biodiversity (Franz and Sterner 2018).[18]

Process-sensitive naming extends trait naming expertise firmly beyond the scientific domain of taxonomy. By explicitly focusing on the relation between valuing practices, data practices and naming practices, this approach proposes to pay attention to the variability of token specimens and their prospective uses, as well as to the multiplicity and broader implications of processes of data design, collection, management and interpretation. No single forms of knowledge or unique epistemic culture can satisfy this requirement: rather, this approach is grounded on transdisciplinary exchange across a wide variety of knowledge systems and related communities, with the opportunity for each relevant perspective to voice its insights. The classificatory emphasis on environmental processes thus fosters the incorporation of traits and naming criteria pertaining to forms of expertise not often considered within scientific taxonomies, such as the expertise of breeders, farmers and other professions relating to the management of land and agriculture, as well as indigenous knowledge—a move conducive to what David Ludwig (2016) calls ontological self-determination. For instance, we have seen how the process-sensitive naming system developed by the Crop Ontology takes into account, rather than side-stepping, the central role played by know-how and sensory knowledge in identifying and describing phenomic traits. Curators are aware of the difficulties intrinsic to using symbols and quantification tools—such as language and numbers—to capture and classify observations on biological organisms, particularly when those are performed in the field as discussed in section 3. To address this, they developed a semantic system that includes both qualitative and quantitative variables, including terms specifically dedicated to measurement and valuation practices used by breeders and researchers conducting field trials, with the explicit aim of capturing tactile traits such the consistency, taste and color of the cassava pulp.[19] In turn, such data help investigating what may explain large phenotypic differences in plants which have the same genetic make-up. What biological and environmental mechanisms and patterns underpin the observed resistance of some plants, but not others, to being dried and ground as flour?

This long-standing line of questioning raises deeper issues with the research priorities of

---

[18] Process-sensitive naming is thus an integral part of what John Dupré and I called *process epistemology*, an approach to scientific epistemology intended to "highlight the dynamic relation between the objects produced through research practices and the phenomena that these objects are used to study" (Dupré and Leonelli 2022).

[19] This reflects a broader attention to what Miller (2019) calls 'sensory ethnobotany', where qualitative information about sensory experience of plant characteristics is placed at the centre of knowledge-making practices.

contemporary biology. Paramount among those issues is the study of infra-species diversity, especially—but not only—in relation to crop science where species-level analysis does not help explain the phenotypic differences in environmental responses, agronomic uses and morphology as documented within the Crop Ontology. This in turns means opening the Pandora's box of what constitutes biologically meaningful difference among organisms, thereby challenging what Staffan Müller-Wille has identified as a key driver for Linnaean taxonomy: that is the creation of "horizontal equivalence" among species, which enables to create and count associations and, to some extent, "eliminate difference" between token organisms (Müller-Wille 2017). This focus on similarity, with its related emphasis on the typicality of species rather than traits (see also Witteveen 2015, 2018), has been immensely successful in creating a level playing field to exchange information about groups of organisms. Its usefulness is obvious when the goal of biological analysis is to understand the evolutionary history of organisms. As Müller-Wille points out, however, the focus on similarity is less useful when organisms are being classified for other purposes, such as for instance their agronomic utility and ecological role, where the ability to recognise and exploit differences among individual organisms is paramount.

Process-sensitive naming responds to these requirements by attempting to articulate and document differences among plant varieties and methods of data collection, sharing and analysis, paying particular attention to the semantic diversity associated to different settings and uses of crop data—including the history of such settings and uses, and thereby the shifts in motivations and goals underpinning conceptual and methodological approaches to measuring traits. While part of this effort unavoidably lies in translating and standardising the language used to depict differences (not least by translating the different languages used by the various stakeholder communities into English), Crop Ontology developers are trying to avoid simplistic systems of equivalence or translation between classificatory schemes. Easy translations often imply significant loss of local knowledge, as Catherine Kendig has discussed at length in relation to the use of synonyms in lichen naming practices. As she points out and I discussed in relation to Crop Ontology labels, naming practices for organismal traits need to capture non-linguistic entities and rely on extensive know-how and culturally specific cues to achieve that goal, which are easily lost in translation. In her words, "assigning a name is one goal of a naming system but names and naming practices may also encode meaning and value that outstrips a simple description of the name as a convenient or pragmatically useful label. […] Synonyms are therefore not synonymous in all contexts but are delimited perspectivally: they are dependent upon who it is that is comparing the names as potential synonyms; what ontological commitments they hold; and for what reason is the comparison sought" (Kendig 2020, 8).

6 Conclusion: Naming Plant Traits in the Era of Big Data

> *Lack of flexibility in the way we name the units of biodiversity leads to unwarranted assumptions of ontology.*
>
> –Alessandro Minelli (2018, 31)

The invitation to challenge contemporary construals of biodiversity and consider the multiple (and culturally and historically situated) goals served by classificatory systems is far from new. It clearly reflects the approach to plant knowledge fostered by ethnobotanists over the last half century, and their calls to broaden the remit of plant science beyond Western and academic perspectives (Nazarea 1998; Ludwig 2016). It also mirrors long-held concerns within biology and philosophy alike about considering species—as exemplified by type specimens—as key units of analysis and research in biology, thereby disregarding other ways of classifying and generalising over biological variation (Witteveen 2018; Minelli 2019; Ankeny and Leonelli 2020). Within this paper, my intent was to explore some implications of such calls for current developments in data-intensive plant science. This domain aims to collect, link and model large and heterogeneous data from multiple sources to acquire a better understanding of biodiversity *as well as* how diverse plant characteristics can be harnessed for sustainable agricultural development and, more generally, planetary health. As my discussion of the difficulties in collecting data on plant traits illustrates, data work in this domain is far from being just an exercise in quantification. Quantification is certainly an essential step towards making plant traits amenable to large scale computational analysis. But it is necessarily underpinned by qualitative, value-laden, and historically situated efforts to identify suitable labels for data clusters, such as trait descriptors. And in turn, the irreducible multiplicity of naming practices means that any effort to mobilise and link plant data needs to include strategies for managing semantic diversity.

In their study of data practices used to document biodiversity, Devictor and Bensaud-Vincent (2016) argued that the ever-expansive datafication of the environment has been accompanied by a shift away from the study of ecological interactions among organisms, with biologists focusing instead on environmental monitoring in the hope to quantify the ongoing loss of biodiversity. This shift away from efforts to understand local ecologies across multiple scales and levels of biological organisation has been accompanied by a highly selective datafication process, where only organisms perceived as charismatic, economically valuable, genealogically significant or 'rare' (exotic) tend to be documented within data infrastructures. As Bowker clearly stressed already two decades ago, in the data world "things which cannot be described easily and well get ignored, and so receive an ever-decreasing amount of attention" (Bowker 2000, 650)—a trend which has only intensified with the acceleration of efforts to construct global linkage tools and AI algorithms to manage and analyse plant data. Thus existing data collections become performative (Bowker 2000, 675): they make the world in their image, by feeding and training models and algorithms that are then relied upon as representations of nature. How semantic diversity is handled within such a system is far from a mere technical issue. The choice of descriptors for crop traits determines which traits are *not* documented and datafied, which forms of expertise are *not* consulted and voiced, what parts of the ecological and social context are *not* considered as relevant towards understanding agrodiversity and its role within agronomy and agricultural policies. And indeed, within most contemporary big data systems the types of data and data sources used to train machine learning algorithms and other computational tools are severely limited and the incorporation of multiple data sources into global data infrastructures is fraught by political, social and technical constraints (Williamson and Leonelli 2022). Adopting a product-focused naming system, where plant traits are taken at face value and decontextualized to the point of total disconnection from their local environment, is an easy way to speed up data collection as well as the dissemination of data—since, as I argued in previous work, data travel faster when unburdened by their history, including information about their provenance (Leonelli 2016). However, this approach is producing a highly skewed data landscape, within which much knowledge around plants has been obscured or ignored altogether; and where elements from local knowledges are included, they are immediately decontextualized and commodified in ways that do not adequately acknowledge and reward contributors—thus producing novel, digitalised forms of bioprospecting (Hayden 2005).

This is why process-sensitive naming constitutes a significant development within the overarching project of a data-intensive biology powered by AI. Its focus on the methods, goals and scales of data collection can be used to explicitly challenge the focus on inter-species comparison, as well as uniformity and similarity among traits, which have long served traditional taxonomy but left the semantic diversity exemplified by grey nomenclatures behind. Process-sensitive naming systems can be deployed to foster users' ability to understand the environmental context of plant traits, thereby supporting investigation of infra-species differences and well-situated comparisons which take account of differences in data provenance. This in turn may help to refocus plant science on the study of agrodiversity and the role it may play in boosting the long-term sustainability of food systems across different geographies and cultures, thereby moving away from conceptions of agricultural development that see high-yield monocultures as the main route to food security.

Whether process-sensitive naming as exemplified by the Crop Ontology will actually be deployed in these ways in the long term, thereby affecting the framing of crop science as a conduit to agricultural development, remains to be seen. The effort required to engage in process-sensitive naming within the current scientific and agricultural landscape cannot be underestimated. It involves challenging the structure and inclusivity of dominant taxonomies and data infrastructures, exhorting them to forge data systems, chains of evidence and conceptual tools that explicitly bridge between epistemic cultures to better understand biodiversity (Franz and Sterner 2018), thus taking account of local knowledge and uses of the plants in question (Kendig 2020) as well as environmental factors of relevance to the traits being described, such as soil and climate. And it involves a strong and practical commitment to engaging multiple perspectives—which in turn demands exercising what could be conceptualised as science diplomacy, with developers brokering exchanges between various data users and continuously mediating between crop-specific, local databases and international initiatives in plant data management, many of which are led by US- and EU-based researchers or by international agencies such as FAO; and negotiating the tensions arising from attempts to link locally acquired digital information into global networks, and the related effort to regulate the transfer of information about plant genetic materials, such as germ plasm, across national borders. Whether the Crop Ontology may continue to engage multiple sources of data and knowledge about crops in these ways, especially given the general tendency to conceptualise efficient data systems as systems that impose standards from above, remains an open question and will depend on factors well beyond the convictions and preferences of those who develop data infrastructures (Leonelli 2022).

Moreover, one could argue that moving to process-sensitive forms of naming and related data work will not do much to challenge the overarching regime of contemporary global food production and its commitment to capitalist extraction of natural resources from marginalised communities and mass production of consumables with little regard for long-term implications. Examples of inclusive and sustainable technologies and approaches being assimilated into settler capitalism abound; think only of how the ecosystem services framework for environmental economics has been used to turn conservationist critiques into neoliberal commodities (Rosset and Altieri 2017). In a similar fashion, process-sensitive data systems can be used by international agrotech to help appropriate local and indigenous knowledge of crops, turning it into IP-protected novel varieties that can be sold back to cassava farmers as supposedly optimized for specific local growing conditions and markets (Miles 2019).[20] These are not issues that can be solved solely at the level of scientific innovation, and there are clearly severe limits to the social and scientific benefits attached to specific approaches to data collection and interpretation in plant research. Nevertheless, I hope to have shown that the process-sensitive approach to naming can, if responsibly implemented, do some justice to the multiple forms of diversity relevant to the study of plant traits, and that this sensitivity could, at least in principle, support the development of more robust, reliable and sustainable knowledge of and interventions on plant life. At a moment of fast transformation for the global system of food production and en-

---

[20] I thank the referees that rightly encouraged me to highlight this crucial issue; see also Williamson and Leonelli (2022b).

vironmental monitoring, due to climate change as well as the accelerated digitalisation efforts associated to the COVID-19 pandemic, innovative data practices are an important component of an urgently needed shift in how humans relate to the plant world.

Acknowledgments

This paper has taken several years and many different forms before settling into a definitive shape. I have strongly benefitted from feedback from audiences in São Paolo, Prague, Paris, Exeter, and Halle (Germany). Special thanks are due to: Elisabeth Arnaud, who not only took the time to discuss her work with me at length and to help me start my research on cassava semantics, but also provided helpful comments and suggestions on the draft manuscript; Peter Kulakow, Agbona Afola, Paterne Agre and Durodola Ouavade for their generous help in guiding me through their research, and all the IITA staff that welcomed me to their campus in 2017; Rachel Ankeny, Helen Anne Curry, John Dupre, John Krige, Hannah Landecker, Staffan Müller-Wille, Nick Smirnoff, David Studholme, Ozlem Yilmaz and Hugh Williamson for many food, plant and data-related conversations; and colleagues at the Wissenschaftskolleg zu Berlin for providing me with the support, nurture and inspiration needed to write the final draft. This research was funded by the European Research Council grants n° 335925 "The Epistemology of Data-Intensive Science" and n° 101001145 "A Philosophy of Open Science for Diverse Research Environments"; the Australian Research Council Discovery Project (DP160102989) "Organisms and Us: How Living Things Help Us to Understand Our World"; the Alan Turing Institute Project "From Local Fields to Global Indicators" (EPSRC grant EP/N510129/1); and indeed, the Wissenschaftskolleg zu Berlin.

Literature cited

Albuquerque, U. P., D. Ludwig, I. S. Feitosa, *et al.* 2021. "Integrating Traditional Ecological Knowledge Into Academic Research at Local and Global Scales." *Reg Environ Change* 21 (45). https://doi.org/10.1007/s10113-021-01774-2.

Ankeny, R. A. and S. Leonelli. 2020. *Model Organisms.* Elements Series, Cambridge University Press.

Arnaud, E. et al. 2020. "The Ontologies Community of Practice of the CGIAR Platform on Big Data in Agriculture: Improving Practices in Developing and Applying Semantics for Data Annotation." *Patterns* 1 (7): 100105.

Ashby, J. A. and V. Polar. 2019. "The Implication of Gender Relations for Modern Approaches to Crop Improvement and Plant Breeding." In *Gender, Agriculture and Agrarian Transformations: Changing Relations in Africa, Latin America and Asia*, edited by C. E. Sachs. London: Routledge.

Atlin, G. N., J. E. Cairns, and B. Das. 2017. "Rapid Breeding and Varietal Replacement Are Critical to Adaptation of Cropping Systems in the Developing World to Climate Change." *Global Food Security*. 12: 31–37. http://dx.doi.org/10.1016/j.gfs.2017.01.008.

Benjamin, R. 2009. "A Lab of Their Own: Genomic Sovereignty as Postcolonial Science Policy." *Policy and Society* 28:341–55.

Biodiversity International. 2007. *Developing Crop Descriptor Lists: Guidelines for Developers. Biodiversity Technical Bulletin No. 13.* https://cgspace.cgiar.org/handle/10568/74489. Last Accessed January 2022.

Biodiversity International. 2021. "Descriptors." https://www.bioversityinternational.org/e-library/publications/descriptors. Last accessed September 9, 2021.

Bonneuil, C. 2019. “Seeing Nature as a ‘Universal Store of Genes’: How Biological Diversity Became ‘Genetic Resources’, 1890–1940.” *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 75:1–14.

Boumans, M., and S. Leonelli. 2020. “From Dirty Data to Tidy Facts: Practices of Clustering in Plant Phenomics and Business Cycles.” In *Data Journeys in the Sciences*, edited by S. Leonelli, and N. Tempini. Springer.

Bowker, G. 2000. “Biodiversity Datadiversity.” *Social Studies of Science* 30:5, 643–683.

Brickell et al. 2016. “Nomenclature for Cultivated Plants.” https://www.ishs.org/sites/default/files/static/sh_18_Sample_chapters.pdf.

Brush, S. B. 2004. *Farmers’ Bounty: Locating Crop Diversity in the Contemporary World.* New Haven: Yale University Press.

Conix, S. 2019. “Taxonomy and Conservation Science: Interdependent and Value-Laden.” *History and Philosophy of the Life Sciences* 41 (2): 15.

Coppens, F., N. Wuyts, D. Inzé, and S. Dhondt. 2017. “Unlocking the Potential of Plant Phenotyping Data Through Integration and Data-Driven Approaches.” *Current Opinion in Systems Biology.* 4: 58–63. https://doi.org/10.1016/j.coisb.2017.07.002.

Crop Ontology Community of Practice. 2021. Crop Ontology (website). http://www.cropontology.org. Accessed January 2022.

Curry, H. A. 2017. “Breeding Uniformity and Banking Diversity: The Genescapes of Industrial Agriculture, 1935–1970.” *Global Environment* 10 (1): 83–113.

Curry, H. A. 2019. “Why Save a Seed?” *Osiris* 110 (2): 337–40.

Curry, H. and S. Leonelli. Under review. “Describing Crops in the CGIAR Era.” In *Research as Development: Historical Perspectives Agricultural Science and International Aid in the CGIAR Era*, edited by H. Curry and T. Lorek. Cambridge University Press.

Ćwiek-Kupczyńska, Hanna, Thomas Altmann, Daniel Arend, Elizabeth Arnaud, Dijun Chen, Guillaume Cornut, Fabio Fiorani, et al. 2016. “Measures for Interoperability of Phenotypic Data: Minimum Information Requirements and Formatting.” *Plant Methods* 12 (1): 44. https://doi.org/10.1186/s13007-016-0144-4.

Devictor, Vincent, and Bernadette Bensaude-Vincent. 2016. “From Ecological Records to Big Data: The Invention of Global Biodiversity.” *History and Philosophy of the Life Sciences* 38 (4): 1–23. https://doi.org/10.1007/s40656-016-0113-2.

De Queiroz, K. 1988. “Systematics and the Darwinian Revolution.” *Philosophy of Science* 55:238–259.

Dupré, J. and S. Leonelli. 2022. “Process Epistemology in the COVID Era: Rethinking the Research Process to Avoid Dangerous Forms of Reification.” *European Journal for the Philosophy of Science* 12:20 https://doi.org/10.1007/s13194-022-00450-4.

Farnham T. J. 2007. *Saving Nature’s Legacy: Origins of the Idea of Biological Diversity.* New Haven: Yale University Press.

Fernandez-Pozo, N., Naama Menda, Jeremy D. Edwards, Surya Saha, Isaak Y. Tecle, Susan R. Strickler, Aureliano Bombarely, et al. 2015. “The Sol Genomics Network (SGN)—From Genotype to Phenotype to Breeding.” *Nucleic Acids Res.* 43 (Database issue): D1036-41. https://doi.org/10.1093/nar/gku1195.

Franz, Nico M., and Beckett W. Sterner. 2018. “To Increase Trust, Change the Social Design Behind Aggregated Biodiversity Data.” *Database* 2018:1–12. https://doi.org/10.1093/Database/bax100.

Franz, N. M., and D. Thau. 2010. “Biological Taxonomy and Ontology Development: Scope and Limitations.” *Biodiversity Informatics* 7 (1): 45–66.

Frison, C. 2018. *Redesigning the Global Seed Commons: Law and Policy for Agrobiodiversity and Food Security*. Routledge.

Fullilove, C. 2017. *The Profit of the Earth*. Chicago University Press.

Garnett S. T., and L. Christidis. 2017. "Taxonomy Anarchy Hampers Conservation." *Nature* 546 (7656): 25–7.

Gotor, E., A. Alercia, V. Ramanatha Rao, J. Watts, and F. Caracciolo. 2008. "The Scientific Information Activity of Bioversity International: The Descriptor Lists." *Genetic Resources and Crop Evolution* 55:757–772.

Haber, M. H. 2019. "Species in the Age of Discordance." *Philosophy, Theory, and Practice in Biology* 11 (12).

Harfouche, Antoine L., Daniel A. Jacobson, David Kainer, Jonathon C. Romero, Antoine H. Harfouche, Giuseppe Scarascia Mugnozza, Menachem Moshelion, et al. 2019. "Accelerating Climate Resilient Plant Breeding by Applying Next-Generation Artificial Intelligence." *Trends in Biotechnology*. 37 (11): 1217–1235. https://doi.org/10.1016/j.tibtech.2019.05.007.

Harwood, J. 2012. *Europe's Green Revolution and Others Since: The Rise and Fall of Peasant-Friendly Plant Breeding*. London: Routledge.

Hayden, C. 2005. *When Nature Goes Public: The Making and Unmaking of Bioprospecting in Mexico*. Princeton: Princeton University Press.

Henkhaus, Natalie, Madelaine Bartlett, David Gang, Rebecca Grumet, Ingrid Jordon-Thaden, Argelia Lorence, Eric Lyons, et al. 2020. "Plant Science Decadal Vision 2020–2030: Reimagining the Potential of Plants for a Healthy and Sustainable Future." *Plant Direct* 4:1–24. https://doi.org/10.1002/pld3.252.

Kendig, C. 2020. "Ontology and Values Anchor Indigenous and Grey Nomenclatures: A Case Study in Lichen Naming Practices Among the Samí, Sherpa, Scots, and Okanagan." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences.* 84:101340 https://doi.org/10.1016/j.shpsc.2020.101340.

Kendig, C., and J. Witteveen. 2020. "The History and Philosophy of Taxonomy as an Information Science." *History and Philosophy of the Life Sciences* 42 (40). https://doi.org/10.1007/s40656-020-00337-8.

Kloppenburg, J. R. 2004. *First the Seed*: *the Political Economy of Plant Biotechnology,* 2nd ed. University of Wisconsin Press.

Leonelli, S. 2016. *Data-Centric Biology: A Philosophical Study*. Chicago, IL: University of Chicago Press.

Leonelli, S. 2020. "Learning from Data Journeys." In *Data Journeys in the Sciences*, edited by Sabina Leonelli and Niccolò Tempini. Springer.

Leonelli, S. 2022. "How Data Cross Borders: Globalising Plant Knowledge Through Transnational Data Management and Its Epistemic Economy." *Knowledge Flows in a Global Age: A Transnational Approach*, edited by J. Krige. Chicago, IL: University of Chicago Press.

Ludwig, D. 2016. "Overlapping Ontologies and Indigenous Knowledge. From Integration to Ontological Self-Determination." *Studies in History and Philosophy of Science Part A*, 59:36-45. https://doi.org/10.1016/j.shpsa.2016.06.002.

Miller, T. L. 2019. *Plant Kin: A Multispecies Ethnography in Indigenous Brazil.* Austin: University of Texas Press.

Miles, C. 2019. "The Combination Will Tell the Truth: On Precision Agriculture and Algorithmic Rationality." *Big Data & Society* January–June 2019: 1.

Minelli, A. 2019. "The Galaxy of Non-Linnaean Nomenclature." *History and Philosophy of the Life Sciences* 41 (31). https://doi.org/10.1007/s40656-019-0271-0.

Müller-Wille, S. 2017. "Names and Numbers: 'Data' in Classical Natural History, 1758–1859." *Osiris* 32 (1): 109–28. https://doi.org/10.1086/693560.

Murphy, D. 2007. *Plant Breeding and Biotechnology: Societal Context and the Future of Agriculture*. Cambridge: Cambridge University Press.

Nazarea, Virginia D. 1998. *Cultural Memory and Biodiversity*. University of Arizona Press.

NextGen Cassava. 2022. CassavaBase (website). https://www.cassavabase.org/.

Pietragalla, J., L. Valette, R. Shrestha, M.-A. Laporte, T. Hazekamp, and E. Arnaud. 2022. "Guidelines for Creating Crop-Specific Ontology to Annotate Phenotypic Data: Version 2.1." Alliance Bioversity International and CIAT. 38. https://hdl.handle.net/10568/110906.

Rajagopal, Prabhakar, Thomas Vattakaven, and Balasubramanian Dhandapani. 2017. "Databasing Crops Plants From the People's Biodiversity Register of India." *Biodiversity Information Science and Standards* 1: e19820. https://doi.org/10.3897/tdwgproceedings.1.19820.

Rosset, P. M., and M. A. Altieri. 2017. *Agroecology: Science and Politics*. Rugby: Practical Action Publishing.

Saraiva, T. 2013. "Breeding Europe: Crop Diversity, Gene Banks, and Commoners." In *Cosmopolitan Commons: Sharing Resources and Risks Across Borders*, edited by N. Disco and E. Kranakis, 185–212. Cambridge MA: MIT Press.

Scott, J. 1999. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed.* Yale University Press.

Shrestha, Rosemary, Elizabeth Arnaud, Ramil Mauleon, Martin Senger, Guy F. Davenport, David Hancock, and Norman Morrison. 2010. "Multifunctional Crop Trait Ontology for Breeders' Data: Field Book, Annotation, Data Discovery and Semantic Enrichment of the Literature." *AoB Plants* 2010:1–11. https://doi.org/10.1093/Aobpla/plq008.

Shrestha, Rosemary, Luca Matteis, Milko Skofic, Arllet Portugal, Graham McLaren, Glenn Hyman, and Elizabeth Arnaud. 2012. "Bridging the Phenotypic and Genetic Data Useful for Integrated Breeding Through a Data Annotation Using the Crop Ontology Developed by the Crop Communities of Practice." *Frontiers in Physiology* 3 (August): 1–10. https://doi.org/10.3389/Fphys.2012.00326.

Sterner B., J. Witteveen, and N. M. Franz. 2020. "Coordinating Dissent as an Alternative to Coordinating Consensus: Insights From Systematics From Bio-Ontologies." *History and Philosophy of the Life Sciences* 42 (8): 1–25.

Striphas, Ted. 2015. "Algorithmic Culture." *European Journal of Cultural Studies* 18 (4–5): 395–412. https://doi.org/10.1177/1367549415577392.

Taylor, M. 2018. "Climate-Smart Agriculture: What Is It Good for?" *Journal of Peasant Studies*. 45 (1): 89–107 https://doi.org/10.1080/03066150.2017.1312355.

Tardieu, F., L. Cabrera-Bosquet, T. Pridmore, and M. Bennett. 2017. "Plant Phenomics: From Sensors to Knowledge." *Current Biology*. 27 (15): R770–R783 https://doi.org/10.1016/j.cub.2017.05.055).

Trappes, R. Forthcoming. "Individual Differences, Uniqueness, and Individuality in Behavioural Ecology." *Studies in the History and the Philosophy of Science*.

Walter, Achim, Frank Liebisch, and Andreas Hund. 2015. "Plant Phenotyping: From Bean Weighing to Image Analysis." *Plant Methods* 11 (14). https://doi.org/10.1186/s13007-015-0056-8.

Williamson, Hugh F., Julia Brettschneider, Mario Caccamo, Robert P. Davey, Carole Goble, Paul J. Kersey, Sean May, et al. 2021. “Data Management Challenges for Artificial Intelligence in Plant and Agricultural Research.” *F1000Research* 10:324. https://doi.org/10.12688/f1000research.52204.1.

Williamson, H., and S. Leonelli, eds. 2022. *Towards Responsible Plant Data Linkage: Data Challenges for Agricultural Research and Development*. Springer Open Access.

Williamson, H., and S. Leonelli. 2022b. “Accelerating Agriculture: Data-Intensive Plant Breeding and the Use of Genetic Gain as an Indicator for Agricultural Research and Development.” *Studies in the History and Philosophy of Science*.

Witteveen, Joeri. 2015. “Naming and Contingency: The Type Method of Biological Taxonomy.” *Biology and Philosophy* 30 (4): 569–86. https://doi.org/10.1007/s10539-014-9459-6.

Witteveen, Joeri. 2018. “Typological Thinking: Then and Now.” *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 330 (3): 123–31. https://doi.org/10.1002/jez.b.22796.

ISSN 2475-3025