Alain Sandoz, and Léa Stiefel (2022): Trust vs. control: the dilemma between data distribution and centralization. In: Proceedings of the International Conference on Privacy-friendly and Trustworthy Technology for Society – COST Action CA19121 - Network on Privacy-Aware Audio- and Video-Based Applications for Active and Assisted Living

Trust *vs*. control: the dilemma between data distribution and centralization

Alain Sandoz* - Léa Stiefel
University of Neuchâtel, Switzerland – University of Lausanne, Switzerland
alain.sandoz@unine.ch – lea.stiefel@unil.ch

Extended abstract

This paper reports on the motives, design, and implementation of a permissioned peer-to-peer platform for the authorized transmission of sensitive data between organizations of an economic sector, agriculture in Switzerland. The period under consideration spanned five years between 2015 and 2019.

Context of emergence

The peer-to-peer platform emerged indirectly because of an apparent need for greater efficiency in data management. After years of digitization of agricultural policies and of market adaptations, farmers were said to be burdened with data-related administrative tasks (Droz, 2014). They were supplying information to numerous organizations, both public and private, which in return provided subsidies, premiums, or other services related to the farm's needs or production modes. Each organization digitized and recorded sensitive information in its database, with redundancies and inconsistencies among these independent systems. Farmers' data was nevertheless controlled for accuracy by some organizations, which carried the risk of penalties, and farmers were under increasing pressure to make this sprawling system work. In 2015, a group of private actors floated the idea that efficiency would be improved if *all of the sector's data* were centralized *in one unique database* (that the same group would operate). Other parties, such as public administrations or producers' organizations, would be allowed to interoperate with the central database using application programming interfaces

(APIs). For reasons that are documented elsewhere, the proposal prompted a wave of protest and led to the launch of a counter-project for an alternate approach to data management (Stiefel, 2022). The project developed a peer-to-peer platform for organizations and farmers to solve the dilemma between data distribution and centralization, which is the subject of this paper.

The present work is the result of a collaboration between the platform's architect (a computer scientist) who led the project and was its chief strategist, and the ethnographer (a social scientist from the STS field) who followed its developments. The first author worked for a consortium of professional organizations that represented over 50% of Swiss farmers. The second author went behind the scenes of the project and in parallel conducted interviews with organizations and farmers to explore concerns related to the issues of data and data sharing.

Data

Switzerland is a small landlocked country in Western Europe. Apart from marginal forestry and fishery activities, agriculture is the only primary sector in the country, with some 53'000 farms, generally small family businesses. Many other actors make up the value chain uphill and downhill of production itself, and numerous organizations, public, para-public or private, structure the sector, from the import of machinery and fertilizers, down to the processing and distribution of food staples, flanked by professional defense, production control, and regulation. All these actors use digital systems and data to support their activities.

Farmers use digital systems in *production* and for *management*. Systems are developed by the agroindustry and/or software providers. On-farm systems use data to deliver functionality. Sources of data can be machines, sensors, or robots, or remote systems *e.g.*, weather stations. The farmer him/herself supplies data, *e.g.*, dates and places of sowing, types of crop or animal treatment, or quantities of produce and income, or costs, etc., used for resource planning. Data for production or management is often stored remotely, in databases operated by independent (and competing) service providers. These systems also compute, produce, and accumulate data. Production data can be sensitive for the service provider, because it is used to improve the system and has indirect market value. It is also sensitive for the farmer, as it describes modes of production that can be contracted or strictly regulated. Management data is sensitive because it is related to quality and quantity of production, as well as to financial, legal, or fiscal aspects, which, if disclosed, could jeopardize the farmer's position relatively to commercial partners, public regulators, and/or their respective control organizations. The evolution of digital technology *as a service* has brought advantages to farmers as it sometimes simplifies management, and enhances the quality and availability of information. It has also largely benefited database operators by giving them free and unrestrained access to huge amounts of valuable data.

Another type of digital systems that use data from farms are information systems of organizations that interact with farmers both collectively and individually: public administrations supervise the implementation of legislation, notably for subsidies, statistics, and policy-making; producers' organizations define requirements for labels (*e.g.*, organic, integrated, traditional, etc.) and quotas, which, if respected by the farmer, bring a premium on the market; professional organizations compute statistics to define their policies and political lobbying strategies; etc. All of these organizations need data from the farmers who are compelled to supply the data if they want to have a shot at the promised benefit. This data is even more sensitive than the production & management type, because on one side, it is specific to individual farms, but on the other it is collective, homogeneous, and often has a wider coverage than service providers can collect from their market share. Database operators (private or public) will have a precise view of a question, both general and specific to each farm.

To summarize: digital representations of sensitive information on farms are maintained in dispersed databases operated by independent service providers and organizations in the sector (data users). And farmers (data owners) want -in fact, need- to keep control of which actors have access to what data. On the other hand, their livelihood requires from farmers to supply large amounts of data to many organizations, including data that is not sensitive (or that can even be public, such as addresses), that is useful in different ways to different actors, and that is partially redundant and sometimes inconsistent.

Over time, managing the data that a farmer supplies to his/her numerous different "partners" became a problem of its own, and a burden. The problem was more often seen by farmers as the uncontrolled gluttony for their data of all sorts of benevolent organizations. But "data-sharing" between organizations[1] then suddenly popped up as the "solution" to the farmer's problem. The prospect of seeing some powerful actor forging the ability to compare data from his/her farm, compiled from different sources and providing answers to different questions, with all the other farms, became a farmer's digital nightmare.

Designed as a counter-measure to that yet unchecked perspective, authorized transmission between organizations *vs.* centralization by one privileged operator was a radically different approach to the problem of data management in the sector.

How farmers and organizations actually perceived the situation

This article is the result of a multidisciplinary collaboration between its two authors. In January 2018, the PhD student had just started her thesis and was interested in tracking the dynamics of digitization in the Swiss agricultural sector. She had attended a public presentation of the peer-to-peer platform by the architect

[1] A notion that goes well beyond the simplistic view of interoperating databases (Hummel, *et al.*, 2018).

and introduced herself, asking him to ethnographically follow its developments. At their second meeting in May 2018, the terms of their collaboration were set.

The ethnographer would be able to go behind the scenes of the project, and to follow and document all its developments. In return for full access, she would provide the architect with regular feedback on her observations, according to the rules of her discipline and her progressive understanding of digitization, via anonymized reports of her interviews. In addition to the practice of his own discipline and professional experience, the architect would benefit from this informed perspective to drive the project within its socio-technical environment.

The ethnographer conducted her interviews between January 2018 and September 2019 with some 40 actors in the Swiss agricultural sector, farmers (5), but especially representatives of agricultural organizations: agents of public administration (11) and professional defense (2), representatives of control bodies (6), certification bodies (2), professional associations and companies in the animal and dairy sectors (6), IT service providers for agriculture (4), and system operators of these same organizations (7) (in parallel, the architect held project and information meetings with over 50 representatives of public and private organizations, farmers and researchers, covering in particular the Eastern part of Switzerland[2], which he also reported back to the ethnographer).

Her interviews documented a range of concerns and risks perceived by farmers (data-owners) and agricultural organizations (data users) regarding the data centralization project that proposed to collect all data in a single database and, on this basis, to develop smart-farming services (decision support modules). Among its shareholders were the largest agricultural cooperative in Switzerland, both the farmers' main supplier and a major buyer of their products, a European software development company, linked to the cooperative by a German machinery manufacturer, and two important publicly owned, resp. supported, organizations.

For data-owners, the fact that the project was backed by such a conglomerate of powerful private players was a source of concern. The centralized database promised to provide its shareholders with full visibility into what was happening on all farms, and on a daily basis. Combined with its decision-support tools, the database would allow them - the cooperative and its foreign partners - to drive the demand for inputs and the supply of agricultural products, and to influence market and supply prices. The risk of "vertical integration" was great for farmers, who would meanwhile retain the burdens of dept and production risks (such as losses due to weather or disease). Moreover, they would have to pay for access to "services" developed on the basis of *their* data, the quality of which they would be held *liable* by contract, while all the profits would go to the database owners.

In addition, it was unclear how data would flow between partners associated with the centralized database. Without control over the flow of their data, farmers

---

[2] German-speaking, in contrast with the French-speaking Western part (origin of both authors) and with the Italian-speaking Southern part (marginally covered, with approx. 4% of the Swiss population and 2% of farms).

were at risk. If data inadvertently reached a government agency, indicating high nitrogen levels in one field that were compensated in another (which can happen every day on any farm), the farmer could lose subsidies. If data from a government inspection showing a health problem in an animal was inadvertently passed on to a dealer, the farmer, and even neighbors, could be sidelined for fear that disease might spread from the shipment to the slaughterhouse (what actually did happen to an entire village because of a single sick animal).

Finally, the push for smart-farming was problematic for farmers, who saw it primarily as a debt driver. Smart-farming was expensive and of little interest for Switzerland because of its lack of applicability in its mountainous, small scale and tradition-oriented agricultural model. Smart-farming favored industrial methods for export crops and intensive livestock that were incompatible with the quality-driven, environmental, and legal frameworks of Swiss farming.

For data-users, *i.e.*, the Swiss agricultural sector organizations, centralization also posed problems. If farmers were to enter their data into a single database, the organizations would have to "log in" to the database to access the data they needed (previously supplied directly by farmers). Farmers' data is of great importance to the organizations. Public administrations provide subsidies to farmers, compile statistics for the evaluation and development of agricultural policy, and control epizootics on the basis of data provided by farmers. Private organizations base their services on farmers' data, some of which are supported by the regulator, such as improving the genetic profile of animal breeds to ensure their resistance to pathogens. But there were no guarantees that they would actually be allowed to access the data in the centralized database, in contents and formats, and at times necessary to carry out their duties, nor was there any indication of the price to be paid. Centralization promised to jeopardize the autonomous management of the organizations' activities, to the point of threatening their very existence.

The project also foresaw to store all farmers' data in a cloud in Germany, under the control of the European software company partner. This posed a problem of data sovereignty, which was unacceptable to public administrations. It also posed problems as to how to resolve conflicts between farmers and organizations arising from data management, with data residing in the legal realm of a foreign authority.

The promoters promised that organizations could propose functional modules connected to the central database. But it was not clear to these organizations if this openness would be observed in reality beyond the rhetoric. The shareholders could very well act single-handedly, as long as they controlled the APIs. More fundamentally, this single, centralized database would introduce a distortion of competition. Faced with powerful foreign shareholders, who would concentrate all the farmers' data, smaller Swiss organizations wouldn't stand a chance to compete, which would sound their death knell.

Finally, private actors (device and machine manufacturers, service and software providers, etc.) who were not in the consortium and not part of the discussion simply waited for the storm to pass.

How the peer-to-peer platform worked

In Switzerland, each organization is legally and technically independent, and is liable towards the owners of sensitive data it manages on the base of some contract (explicit, or implicit as in the case of public administrations representing the regulator). In particular, data can be accessed online by a farmer only if the latter has been identified and authenticated by the database operator.

If data owned by a farmer were to be *transmitted* from one database to another, this would have to be with that farmer's authorization. Authorizations could be set and revoked at any time (by farmers using a mobile application) and were specific to *i*) one farmer, *ii*) one pair (sender-receiver) of database operators, and *iii*) one *datatype*. As long as an authorization was valid, the (willing) sender could send the famer's data of that type to the (requesting) receiver (using a three-step asynchronous protocol). Authorizations and transmissions were traced so that they could be recovered in case of suspected misconduct (a process that would be supervised by an auditing authority or a judge). Each peer (data user) operated its IT infrastructure under its own legal responsibility, including the platform's component, called *node*, that was its access-point to other peers, and the place where its transmissions were traced. For each sender-receiver pair, the datatypes that could be transmitted, for what purpose and under what conditions, were published on the platform by the operators. How authorizations were managed and how transmission was implemented in each node was transparent and certified (the platform was an open standard). However, what data was actually transmitted, with what values and when, was known only to the three parties involved: the owner, the sender, and the receiver, and transmission was direct between the latter two.

Design rationale and constraints

As mentioned above, the identity of the owner must be determined whenever the data is required for an operation, and access to the data by programs must be controlled. The organizations that operate the database and application servers that run the programs to provide a service to the owner are *de facto* users of the data. These systems are specific to each application domain (cereals, livestock, milk, etc.), and often to the organization itself. They have evolved over long periods of time and are heteroclite and heterogeneous assemblages of technologies. Low-level interactions between the legacy components and the solution would be specific. To simply distribute a software package that would be installed within their legacy infrastructure to implement authorized data transmission between organizations was not a technical option. The solution also needed to be isolated from that

infrastructure, so that a peer could disconnect from the platform without any other loss of functionality other than data transmission to others. A detailed description of how this was envisioned *in general* was fundamental to making an infrastructure acceptable to its future users (be they data owners or data users). This was provided by choosing the standard components (middleware, interface framework, and general-purpose functions) of the platform from free open-source, widely respected, software projects.

However, it was hardly enough to convince operators to adopt the system, and less even, to *share* their data with other organizations. A set of *principles* was established and communicated to the organizations, and then strictly implemented by the project (without any compromise or trade-off, for any reason). This covered the part of the development specific to authorized data-transmission. It was also to be freely distributed as open-source. The platform was *fully* distributed: it had no central component and all roles were completely *symmetrical* (what a peer could, every peer could, with the same constraints). Distribution of the platform ensured freedom of association, equal treatment, and symmetry among peers (Stiefel, Sandoz, 2022). Being neutral with respect to power relations between peers, the platform could enhance trust between operators. *Functionally*, the design was limited to the transmission of data between users (in the above sense), when and only when the data owner authorized it. Data was sent and received, and stored and accessed, only by the peers that used the given data, under separate and mutually unknown contractual conditions they had established with the data owner. Transmission, if authorized, was bilateral and direct between peers. Traces required for a peer to positively prove *correct* behaviour were always left locally and under the control of that peer only. Control data could not be forged (without the collusion of a qualified majority of peers). It could only be removed from a node by its peer, because of the latter's full local control. Consequently, there could be no proof of *misconduct*, only an absence of proof of correct behaviour, that could then lead to the suspicion of a rule violation. These considerations follow the technical line of what is possible or not in a distributed system under the principles stated above. It is not the purpose of this extended abstract to go into the details (nor *e.g.*, to argue why data transmission *was not* realized using blockchain technology[3]), but only to mention that the system design was keen on meeting its principles.

However, this still did not seem sufficient: legal requirements (collective contracts, node certification, general public licensing of the platform) were added to bring operators to adopt a model of action acceptable to the community of farmers (and among the organizations themselves).

---

[3] In particular, blockchain technology orders state changes using a decentralized computation that can delay certain operations in order to achieve consensus, thus potentially threatening participants' control over local operations; whereas in our case, neither consensus, nor ordering among all peers, was required globally to establish a farmer's authorization and to achieve bilateral data transmission among the peers involved. Only the full control of all three participants over any part of these operations was necessary.

Technically, the project faced two problems: 1) *asynchrony* in distributed systems (which is usually overcome by using the master-slave paradigm underlying internet protocols based on APIs); and 2) *matching the different meanings attributed to information* by a sender, a receiver, and a farmer *using digital data* (which is usually overcome by imposing data standards and formats between operators, without asking the data owner's opinion). The first problem was solved based on the properties of communication in distributed systems, which were established at the time of the architect's PhD thesis, early in the 1990s. The latter was trickier because, on one side, farmers (as well as most employees in organizations) are not familiar with the concept of digital data, and, on the other, most IT technicians have no idea of the gap between digital data and the information it is meant to represent. A mechanism called *segmentation* was designed and implemented in the project, and used to bridge the gaps in time and meaning that existed between organizations that would exchange data over the platform. The same mechanism underlay the touch-screen graphical management-app for authorizations used by the farmer. Since organizations would know how to link information they managed for farmers to digital data by using segmentation, it was foreseen that the partners that farmers trusted (*i.e.*, the professional organizations *from* which they were willing to have their data *sent* to others) would help them manage their authorizations by providing guidelines and templates.

The technologies and technical mechanisms used to implement a platform with these characteristics must either be broadly available or represent a small set of specific features that will be made freely available (Sandoz, 2020). To integrate the platform, operators' legacy production infrastructures could not be modified, but only *extended* in cheap and standardized, yet secure, ways, without affecting the mission or function of these systems. The platform architecture, node implementation, and connection of legacy systems to nodes, were based on the Kubernetes (K8s) microservice architecture, respectively the gRPC interface framework and the Hyperledger Fabric distributed ledger. The latter two technologies were available at the time of the project on top of K8s and all three were freely available in OS code. Mastering these technologies was at that time a big effort for any IT operation, especially those of organizations active in agriculture. Connecting a legacy system to its node through gRPC could also be challenging, depending on the legacy system. The project proposed to *lease* certified nodes to organizations as long as would be required, and to help them connect their legacy infrastructure to their node. The collective effort invested in the platform would bring the investment by individual organizations to build and configure their node down to a couple of man-months. At least one operator (operating the public database of five cantons) proceeded to migrate *completely* to K8s *during* the course of the project and was still happy with the move in March 2021. The platform and its technical architecture were also openly described as a modern initiative by an operator who managed the databases of four other cantons.

With this approach, the cost of the platform for any peer was very low, compared to what would have been required to build and maintain APIs to an alternate central database for all agricultural data. And additionally, the risks remained under each organization's control.

Conclusions

The platform was designed to address the concerns of data owners and of data users, and to support mutual trust with its socio-technical architecture (Mazzella, 2016). Farmers demand privacy from the organizations that manage their data. They trust them more or less willingly to provide services in accordance with the information they supply about their farms. They don't necessarily trust those organizations to use their data properly, transparently, and solely for their customers' benefit. An architecture that relies on transparency in how data is collected and used by operators could enhance trust. In a framework with clearly defined rules, control mechanisms and sanctions, and that the user community itself can steer according to changing conditions and needs, possibly under the guidance of external authorities (Hess and Ostrom, 2007), farmers might better accept, and push for, data sharing. Data management might then become more efficient.

The peer-to-peer platform was designed and developed with this in mind. Initiated in early 2018, it went into production in mid-2019 as a first productive prototype with five peer-demonstrators.

References

Droz, Y., Miéville-Ott, V., Jacques-Jouvenot, D., and Lafleur, G. (2014): *Malaise en agriculture. Une approche interdisciplinaire des politiques agricoles France-Québec-Suisse*. Karthala Editions, Paris.

Hess, C., and Ostrom, E. (eds.). (2007): *Understanding Knowledge as a Commons. From Theory to Practice.* The MIT Press, London.

Hummel, P., Braun, M., Augsberg, S., and Dabrock, P. (2018): "Sovereignty and data sharing", ITU Journal: *ICT Discoveries*, Special Issue No. 2, 23 Nov. 2018

Mazzella, F., Sundararajan, A., D'Espous, V., and Möhlmann, M. (2016): "How digital trust powers the sharing economy". *IESE Insight*. Vol. 30. No. 3. 10.15581/002.ART-2887.

Sandoz, A. (2020): *Inter-operating Co-opeting Entities: A Peer-to-Peer Approach to Cooperation between Competitors*. Proceedings of the 10th Int. Conf. on Business Intelligence and Technology. Nice, Oct. 2020.

Stiefel, L. (2022): "Les données du problème. Une plateforme numérique inadaptée à l'agriculture suisse". *Etudes Rurales*, No. 209 / 2022. Editions de l'EHESS. *In press*.

Stiefel, L., and Sandoz, A. (2022): *Critique de la concentration: une analyse des relations de dépendance sur les plateformes numériques*. Proceedings of the XXXIst AIMS Conf. on Strategic Management. Annecy. 31 May – 3 June 2022. *In print*.