Big data: a cultural and technical challenge for which agriculture has strong assets

C. Huyghe, M. Mambrini-Doudet, O. Hologne
INRA, France

P. Bergeret
Directeur du CIHEAM Montpellier et Président d'EURAGRI

EURAGRI

No doubt that big data is a technical challenge. During the Euragri workshop organized in Paris on March 9th 2016 on how to tackle this challenge in agriculture, the extent to which it is accompanied with cultural transformation became more and more obvious. The workshop highlighted also that the agriculture sector has a number of assets to benefit from the opportunities given by big data. Reversely agriculture could provide an interesting application field for specialists of big data.

For big data, the agriculture sector has numerous assets

Agricultural sector has a huge potential for big data specialized companies, both because of the large amount and diversity of data and because of the added value expected from data processing. Indeed, compared to data processed in other sectors, agricultural data have a large spectrum of variety, multiple processes of validity and mobilize a rich vocabulary. Ontology is then a critical issue. This is linked to the multifunctionality and the complex but well defined systems of actors operating in the agriculture sector.

These features are particularly advantageous for a big data approach. However one has to keep in mind that the data per se have no value. It is the global analysis that is generating value. They have to be searchable, accessible, interoperable and reusable among scientific disciplines and economic sectors to provide useful information and generate added value.

- ➔ *Those particularities (data heterogeneity * variety of assets) should attract data scientists. Agriculture is an outstanding playground and an amazing field for demonstration of the potential offered by big data exploitation.*

- ➔ *Additionnally, knowledge discovery, farming activities of the Agri/food business sector would strongly benefit from "big data" tools and methods to discover new patterns and tackle the challenges of a sustainable agriculture*

One key step is the information system interoperability and the subsequent possible data mining. It is usually addressed through ontological approaches that also define semantic links between the concepts. Ontologies can be seen as conceptual models that contribute to a better exchange of information across disciplines and sectors. Again, as the agri-food system has the peculiarity of a large diversity of actors and a huge diversity of vocabulary, a special attention must be given to this issue. It implies to define reference ontologies as shared standards but it is a tricky issue because the first definitions proposed by an actor can be a way to impose his vision and drive the usages. Public organizations have a key role to play in the definition of the standards, while ensuring the permanent link with all the stakeholders, including farmers and farmers' organizations.

- ➔ *Analyzing data patterns and their associations helps to reintegrate farmers into the project*

- ➔ *Analyzing the co-evolution of the concepts and vocabulary may reveal orientations.*

Big data is transforming the research process, because data to be investigated do not necessarily arise from an experimental procedure and are not necessarily provided by the traditional actors. Hypotheses are generated, which cannot be validated because of the tremendous scale tackled. As a consequence, they have to be re-internalized in the research laboratories to set new research questions and new research approaches.
It happens for agriculture sciences as well as for other sciences: the academic frontiers are blurring. More interestingly, new types of research products must be anticipated.

➔ *Expressing the fundamental differences between the research process and the value creation process, may help to operate the transformation while warranting the scientific quality and deontology.*

One of the first questions raised is data property. The property per se can be treated by searching for agreement and common rules among the partners before undertaking the project. Intellectual property related to data appears to be weak. Big data may also be considered as public goods and be handled to avoid destructive competition. Shifting from data property to data use and production of added value must be investigated. This also leads to the critical issue of the perimeter for investigating big data issue. This is essential in agriculture. Indeed, should we consider agriculture and food industry together? How to consider production and economic performances, environmental issues, social aspects, all of them being at the core of multifunctionnality of agriculture? Thus, it seems that the question behind property is, on one side, how can we infer the future value of the data? and on the other side, how do we foresee the increase of business value with the data?

➔ *Strategies are required to build prospects and in the meantime, attention has to be paid to the costs of data treatment and warehousing, which usually remain obscure or unconsidered.*

E-Infrastructures are engines linking data, actors and knowledge production

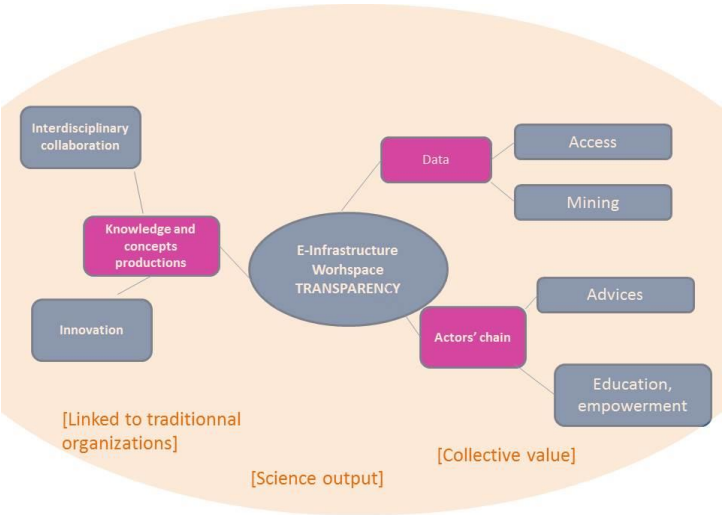
The data science projects funded by public bodies support e-infrastructures which are also workplaces, through the creation of virtual research environments.

They are shaped to ensure

- Regarding the data
- Access
- Mining
- Communication among devices

- Regarding the actors chain
- Advice
- Education and empowerment

- Regarding knowledge transformative effect
- Interdisciplinary collaboration
- Innovation



They ensure:

- transparency and negotiation for property,
- scientific output,
- collective value output,
- link with the regular organizations (academy, public bodies, professional organization).

➔ *Sustainability of such e-infrastructures at the end of the funded projects is not clear and is generally marginally considered in their business model. It is becoming an increasing challenge*

Between big data and decision making, there is a concept space politically shaped

"Who pays, who masters, which country and which people will be able to deal with this challenge?" are remnant questions going along with big data. But it has to be reminded that it is quite unusual in scientific workshops to raise such deep concepts on guiding principles and ethical concerns. It shows the unique status of big data, as seen from agricultural sciences.
Beside data management, big data vehicles the wish of disposing of actionable insights to democratize decision making and work with prediction rather than with intuition and guess.

Regarding the tremendous potentials and challenges of the big data and the inescapable requirement of permeability of present frontiers, the academic organizations should couple scientific and strategic postures as such:

- Focus on applications in line with the main mission of the organization and look for enhancing sustainability,
- Develop a strategic partnership mixing traditional and new partners (e.g. along the actors chain, new players related to the technological issues, from both public and private sectors),
- Favor (open)-innovation while warranting the feed-back to the traditional organizations, the quality of the scientific output and the public and private valuation,
- Facilitate community building and emergence of new skills related to big data, and reinforce education and training, from research side (data science) to users (farmers and industries).