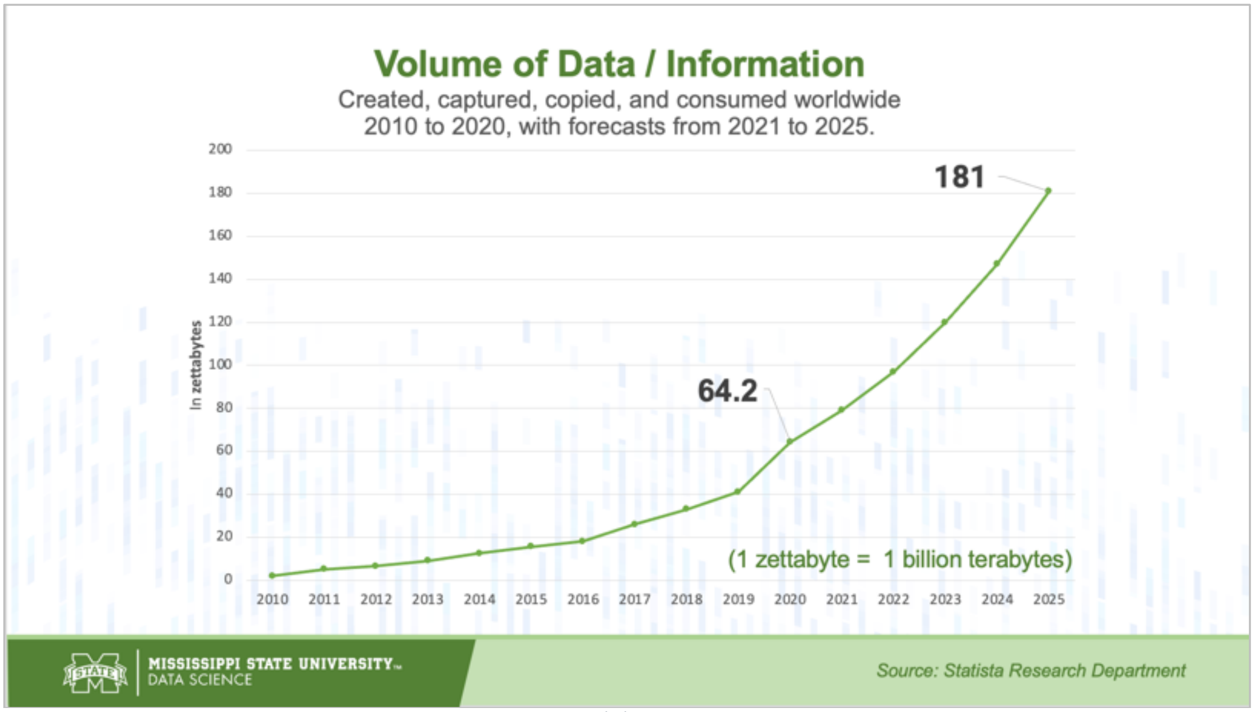
Where Data Meets Innovation
A Framework for Harnessing the Power of Data Science in Agrifood Systems

Dr. Mimmo Parisi
Professor and Executive Director
Data Science Program, Mississippi State University

Slide 1

Good morning and thank you for the opportunity to be here with all of you today. The value and contribution of data to science and technology in general, and to agrifood systems in particular, is not new. The new opportunities and challenges, however, arise from four major trends.



Slide 2

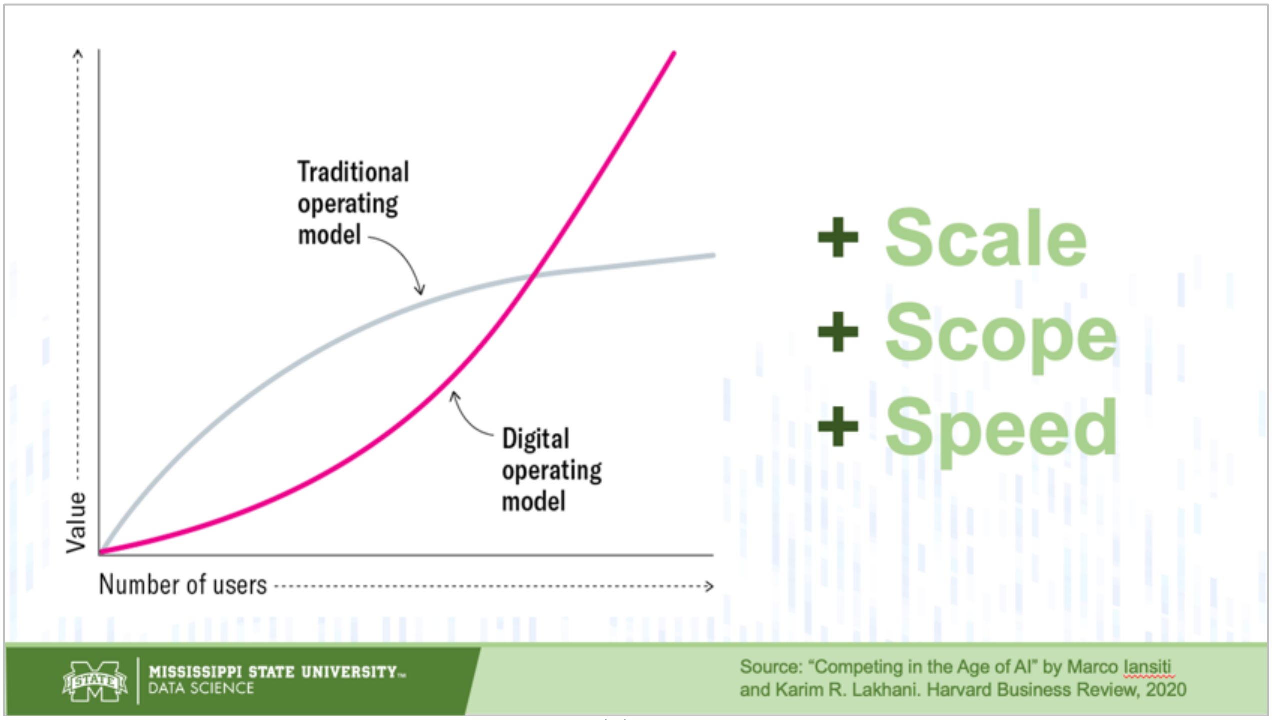
First, the growing volume of machine-generated data is identified with the general term Big Data—data produced by the growing use of digital technologies such as smart phones, sensors, and internet-based applications. Big Data has grown from 2 zettabytes (2 billion terabytes) in 2010 to over 64 zettabytes in 2020 and is projected to grow to 181 zettabytes in 2025 (See slide 2).

Slide 3

Second, the level of data granularity is also increasing, changing how goods and services are produced and delivered (See slide 3). For example, in the health care system, we have shifted from general care to personalized care. In education, we have experienced the same transition where the focus is more on personalized education than general one-size-fits-all approaches. It is no longer about delivering general educational content but rather how we deliver the content based on the learning needs of specific individuals. The same can be said for agriculture. We can imagine moving from a general approach to agriculture (in which data has helped us to make general decisions about planning, planting, harvesting, processing, marketing, consumption, and agricultural trade) to personalized agriculture. Here farmers become both consumers and producers of data.

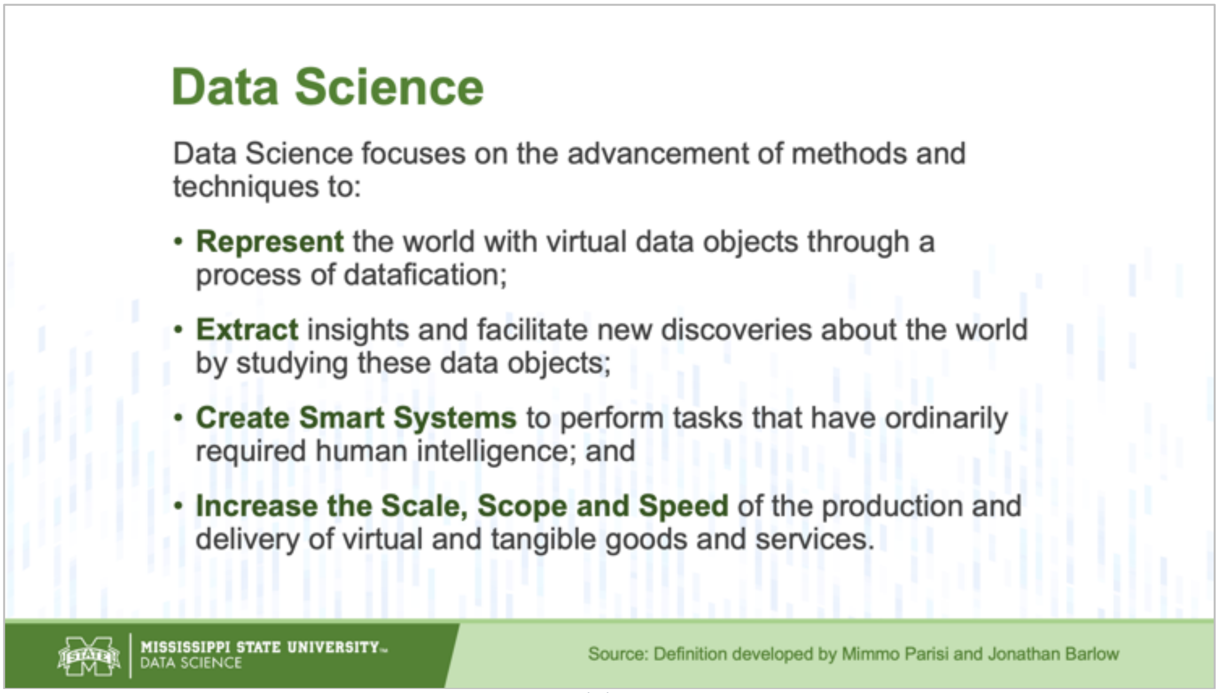
Slide 4

Third, machine learning and AI allow discovery beyond the abilities of human reasoning. For example, AI was used to discover the antibiotic *halicin*. AI is also used to classify, in a very precise way, cancer cells in medical images. AI has also been incorporated into interconnected smart systems that can perform complex functions with an unprecedented degree of independence. Network-connected data, digital technology, and AI are now contributing to the emergence of smart cities and smart agriculture (See Slide 4).



Slide 5

Fourth, digitally operated systems are eliminating the typical constraints to the increase of scale, scope, and speed in the production and delivery of goods and services. Studies show that digitally operated organizations and businesses outperform those that are not digitally operated, bringing even greater value as scale increases (see Slide 5).

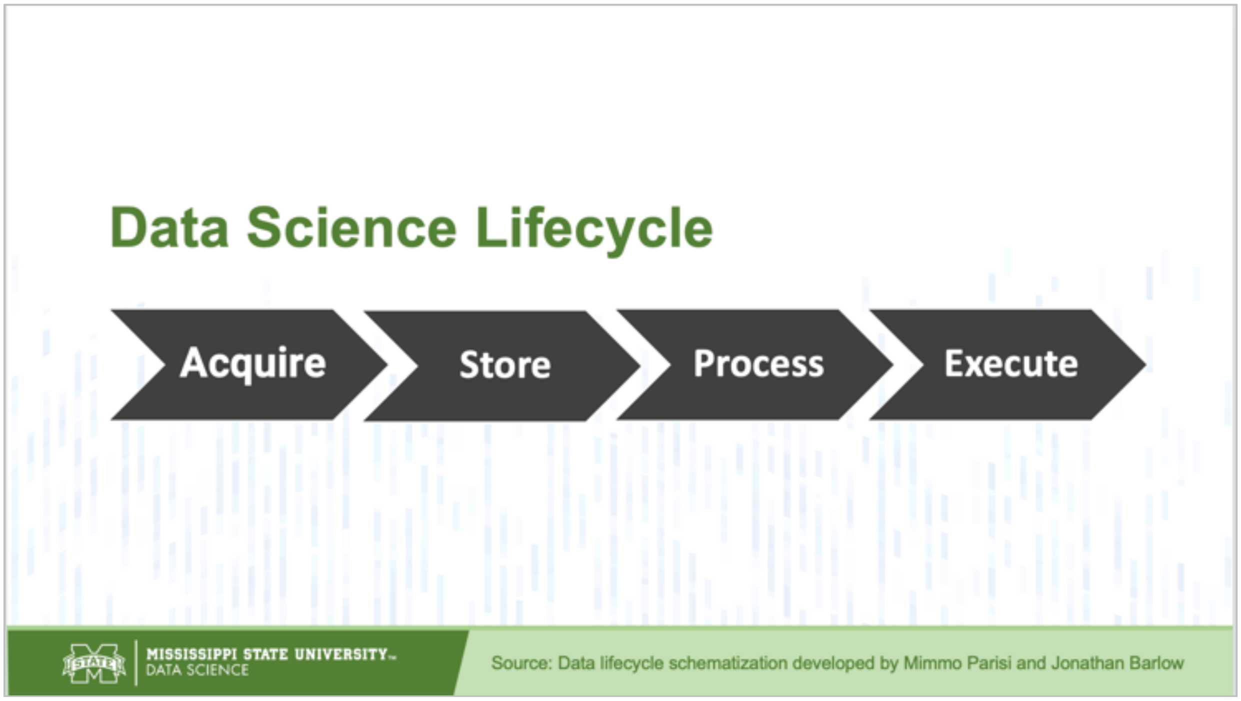


Slide 6

So, what is Data Science?

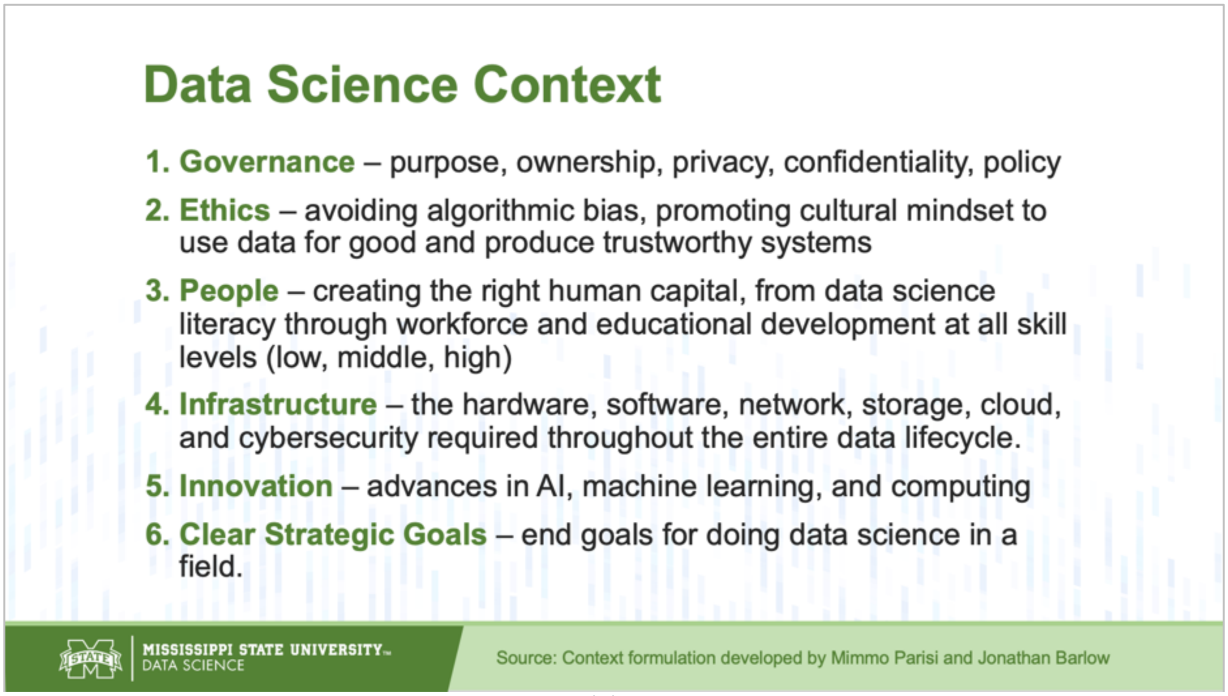
Data Science is the field to advance methods and techniques to improve the use of data for human progress. Specifically, data science focus on the advancement of methods and techniques to:

1. Represent the world with virtual data objects through a process of datafication;
2. Extract insights and facilitate new discovery about the world by studying these data objects;
3. Create smart systems to perform tasks that have ordinarily required human intelligence; and
4. Increase the scale, scope and speed in the production and delivery of virtual and tangible goods and services. (See Slide 6)



Slide 7

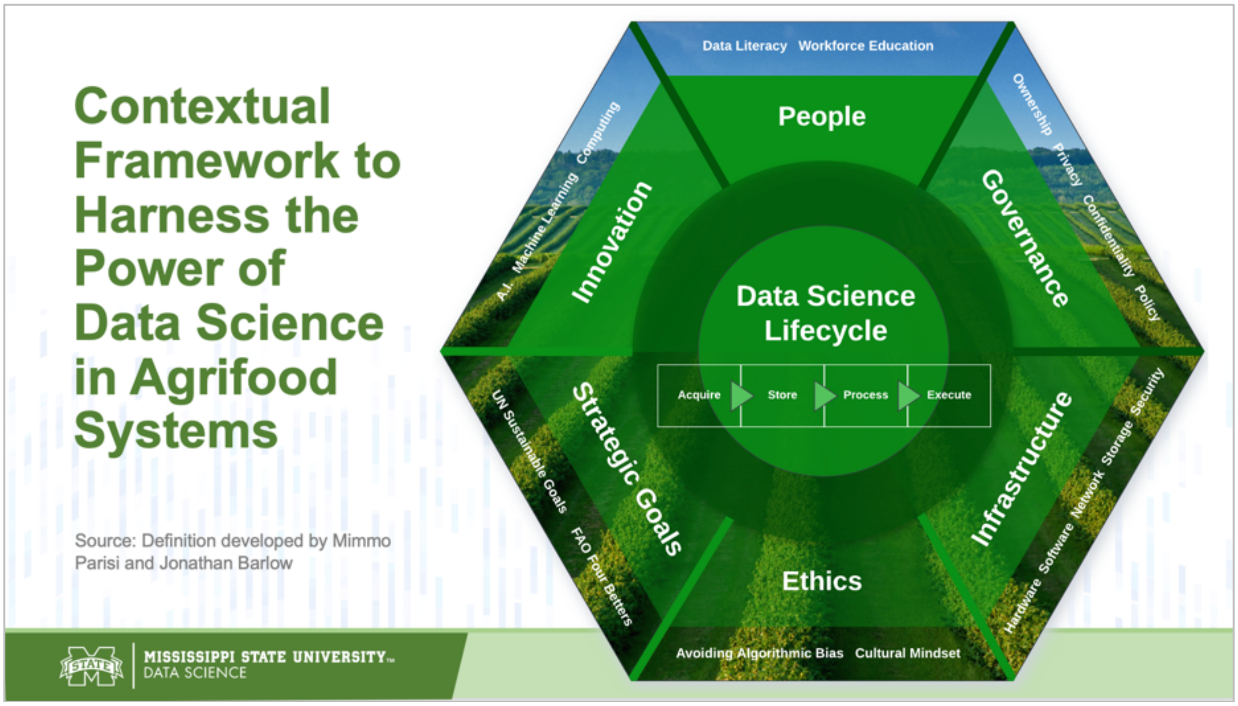
Data Science also focuses on advancing and designing strategies for all aspects of the data lifecycle, which includes Data Acquisition to create, transfer, and clean the data, Data Storage to handle the rapid growth of data, Data Processing to apply the best analytical tools and technologies for the purpose of knowledge creation and discovery, and Data-Driven Execution to inform decisions and automate decision making processes in the production and delivery of goods and services (See Slide 7).



Slide 8

The data lifecycle cannot operate unless it is placed into the context of six general dimensions:

1. Governance – including clearly defining the purpose of the data, and issues of ownership, privacy, confidentiality, and policy.
2. Ethics – including the avoidance of algorithmic bias and the promotion of a cultural mindset to use data for good and produce trustworthy systems
3. People – creating the right human capital, beginning with data science literacy and continuing with workforce and educational development at all skill levels (low, middle, high)
4. Infrastructure – ensuring the supporting hardware, software, network, storage, cloud, and cybersecurity required throughout the entire data lifecycle.
5. Innovation – advances in AI, machine learning, computing, and in all data science technology
6. Clear Strategic Goals – these provide the end goals for data science in a given context. In the context of agriculture, these would be the UN Sustainable Goals and the FAO Four Betters. (See Slide 11)



Slide 9

The lifecycle and these six areas must be integrated into one coherent framework. I firmly believe that if we want to reap the full benefits of Data Science and the use of data in agrifood systems, we need to acknowledge the value of using a common framework of reference to guide the use of data science and data in the most effective way. In so doing, we will also be able to assess in a very uniform and systematic way the "data readiness" of agrifood systems in different ecological and political contexts. We also will be able to develop a plan of action designed to build capacity at all levels for the implementation of a data science framework aimed at achieving the sustainable goals and the four betters. (See Slide 9)

Slide 10

For example, the FAO articulates several program priority areas under “better production” including the theme of innovation for sustainable agriculture production. This goal pursues:

“Sustainable crop, livestock and forestry production systems that are productive, resilient, innovative and competitive, and create integrated entrepreneurial and business opportunities, inclusive of small scale and vulnerable producers, supported through enabling technologies and policies (See Slide 10).”

Slide 11

Let's see how the suggested contextual framework works to apply data science to this strategic better-production goal. Concretely, the model suggests actionable steps in its six contexts:

- People – we must find ways to nurture the development of an agricultural workforce able to collect data, transfer it to computing resources for processing, and act based on intelligent recommendations. Without gathering and using data at the smallest agricultural unit, it is difficult to produce estimates for small-scale farms, connect small farms to markets, and help small farms leverage knowledge, expertise, and technology that increases yields through precision methods.
- Governance – must be in place to ensure clear principles of data ownership, privacy, and confidentiality. Policies that support these values must promote and not become barriers to progress, and ideally, we will find ways to share data, such

as remote sensing data, with smaller farmers as a public good.

- Ethics – this context plays a role both negatively, in avoiding algorithmic bias, and positively in instilling a sense of the importance of a cultural mindset geared toward the use of data for good.
- Infrastructure – must be in place to acquire, store, and transmit data from small producers securely.
- Innovation – to support this work we will need low-cost data collection devices, commercial smart devices, and ways to make computing and internet power available at the local level.
- Strategic Goals – Each of the model's contextual dimensions operates as a way to ensure strategic goals are met, including our present test case – better production especially with regard to inclusion of small-scale farms (See Slide 11).

I have presented this framework, along with an example of its application, with the hope that it will be a useful context for our discussion today.

Thank You