

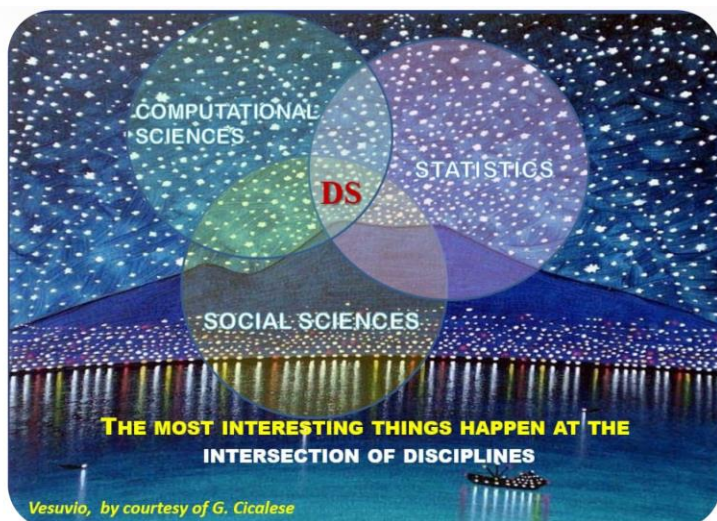
DSSR 2024

Introduction by Carlo Lauro

Data Science & Social Research

4th international conference

Naples (Italy) on **25th-27th March 2024**



<https://dssr2024.unina.it>

Dear colleagues, ladies and gentlemen,
on behalf of the **Association of Professors Emeritus Federiciani**, which I am honoured to chair, I welcome you to the **4th edition of the international conference on Data Science & Social Research (DSSR)** which began its journey in 2016 in this prestigious seat of the University, founded by Frederick II 800 years ago, the first public university in the world.

The aim of this edition is to take stock of the path taken so far in the DSSR conferences, but with an eye to the future.

The idea of this conference was born in 2015 while I was wondering what Data Science was in relation to **Statistics**, in which I was born in this same university, to **data analysis** that saw me follow in the footsteps of masters such as Vittorio Amato, Jean Paul Benzécri, Chikio Hayashi, and finally with respect to **Computational Statistics**, I contributed to the growth of the ISI in the 1990s, as President of the International Association for Statistical Computing and its European section. I soon became convinced that in my approach to research I had navigated the main stream of what is now called Data Science.

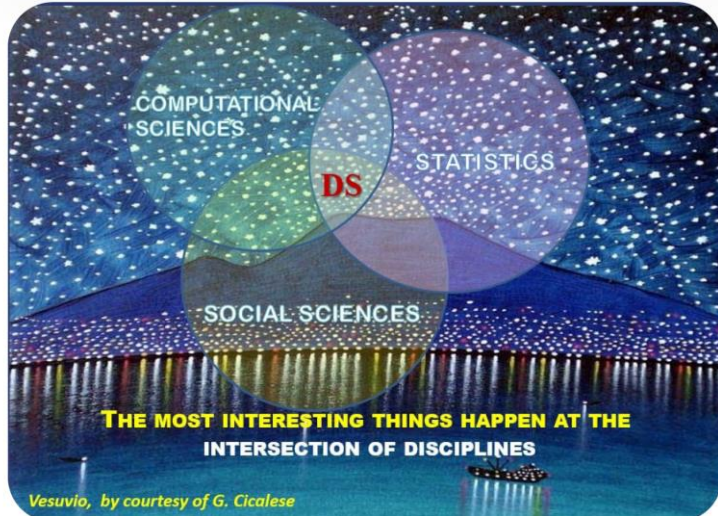


DSSR 2024

Data Science & Social Research

4th international conference

Naples (Italy) on **25th-27th March 2024**



<https://dssr2024.unina.it>

In fact, **in the 5 principles of data analysis enunciated in 1973 by Benzécri**: (1) "statistics is not probability..."; (2) "the models should follow the data.. "; (3) "simultaneously process the information relating to the greater number of dimensions for a complete representation of phenomena ", (4)/(5) regard use of the computer to process the data "for the analysis of complex phenomena the computer is indispensable" and even "use the computer implies the abandonment of all the techniques designed before of computing ".

They anticipate **data-driven** approaches, the role of the **big data**, and the **paradigm shift of classical statistics** that inspire Data Science.

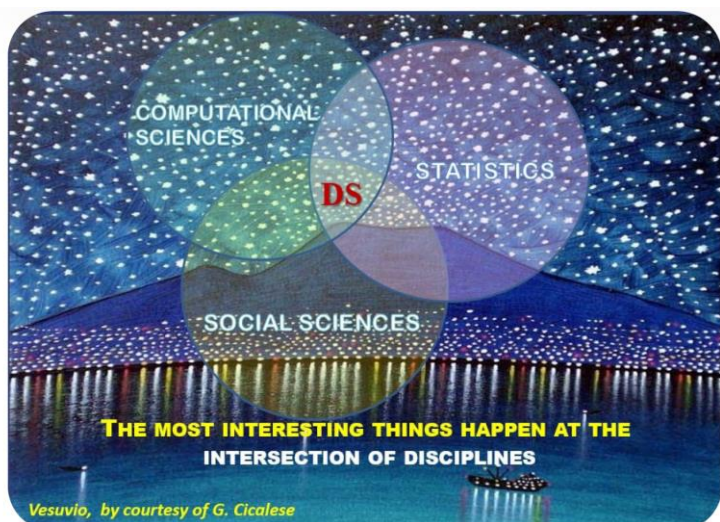
Hayashi then proposed, for the first time, the term Data Science on the occasion of a historic conference held in Kobe (1997) in which I participated as President of IFCS, the International Federation of Classification Societies. Finally, it should be noted that **the Mission of the IASC** ("to link traditional statistical methodology, modern computer methodology and the knowledge of domain in order to convert data into information and new knowledge"), was practically overlapping with that of Data Science as proposed in the following years by numerous authors.



DSSR 2024

Data Science & Social Research

4th international conference



<https://dssr2024.unina.it>

After this first conclusion, I tried to answer the following **questions**: **Is there a Data Science?** If yes, then **what is** Data Science? And **what does Data Science mean** in “data revolution era”: new methods for old content or old methods for new content? What about new **professions**? What are the **challenges for Statistics**?

To this end, I decided to **analyze 80 definitions of Data Science and 70 of data scientists** found on the web over a span of 18 years, with content analysis techniques, lexical factor analysis conducted with correspondence analysis and a subsequent cluster analysis were of great help in finding answers to the previous questions.

The results obtained, for the sake of brevity, referring here to the map of the first 2 axes of the lexical factor analysis on definitions, were immediately clear, leading to some first **interesting evidences**:

i) Existence of 3 main groups of words referring, from right to left, to **epistemological aspects** of definitions, to **methodological aspects** related to statistics and computational sciences and to the **applications** of data science in different domains of knowledge mainly to social sciences, business science and health care data science (second factorial plane).

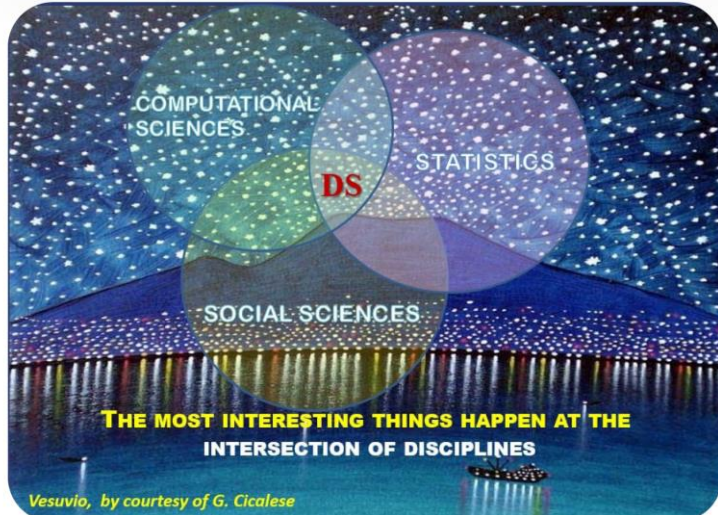


DSSR 2024

Data Science & Social Research

4th international conference

Naples (Italy) on **25th-27th March 2024**



<https://dssr2024.unina.it>

ii) On the basis of these different applications characterized by different types, data sources and methods ^{si} concluded that we could not speak of a data science but rather of **data sciences as a function of the reference domain**.

iii) The first axis (from right to left) opposing **the older definitions with the more recent ones** in correspondence with the epistemological definitions and those referring to the applications of Data Science, passing through methodological developments evidentiare the **evolution of DS**

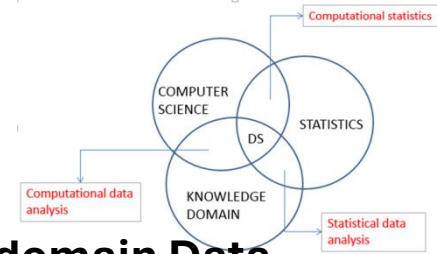
iv) The previous results were **confirmed by the annotation of the authors' profession** (academics, technicians, professionals) who were distributed along the first factorial axis following the different types of definitions.

Starting from these results, I tried to propose a first **definition of Data Science taking into account the logic, the language and the domain of knowledge of reference**, well depicted in the logo of this conference. And it was precisely in this latter perspective that with colleagues from the department of social sciences we decided to deepen the **relationship between Data Science and Social Research in the context of the first DSSRR**.

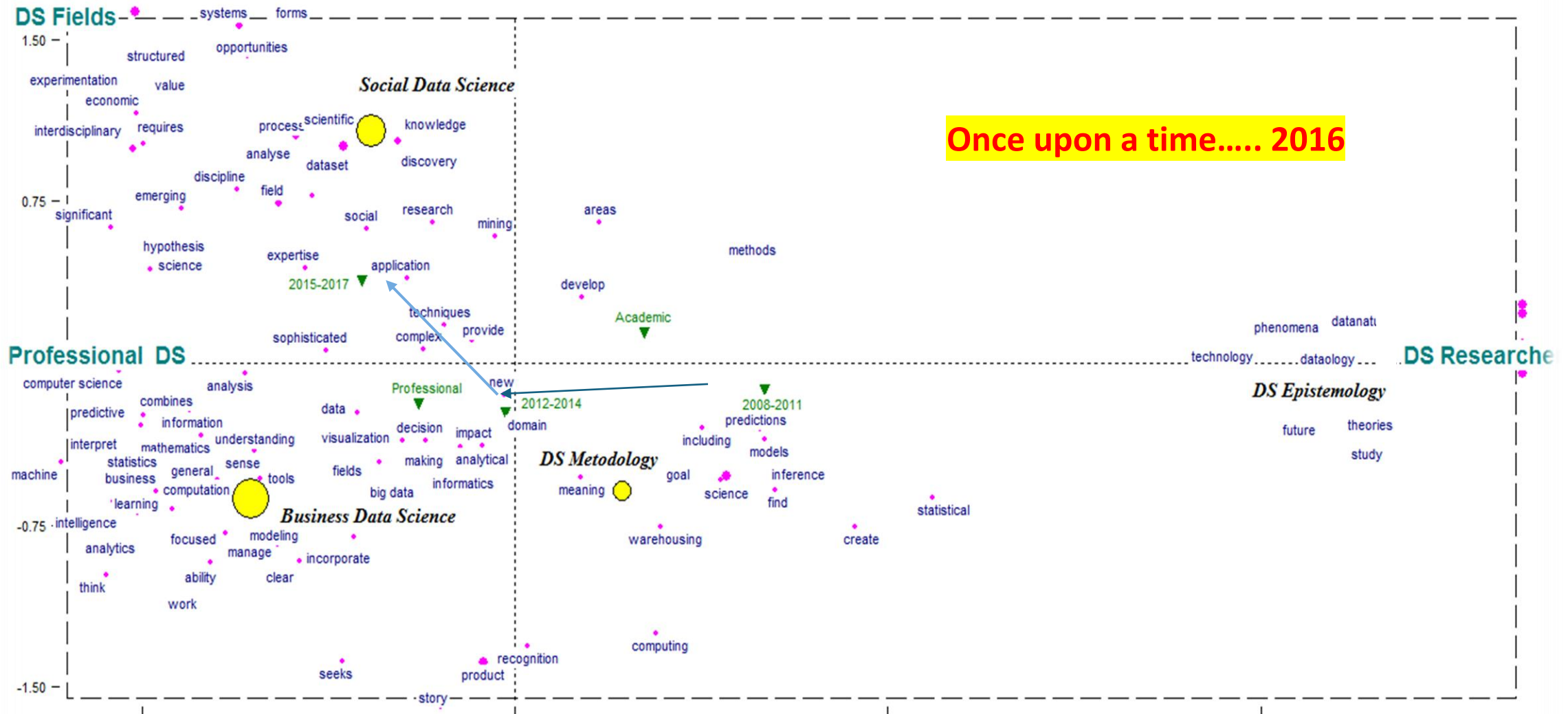


“Let’s talk about Data science”

A Lexical Correspondence analysis of 80 DS definitions



1st axe: opposition of Research and Professional DS. 2nd axe: opposition of domain Data Sciences .A typology according 4 Clusters: Epistemology DS, Methodology DS, Social DS, Business DS



DSSR 2024

Data Science & Social Research

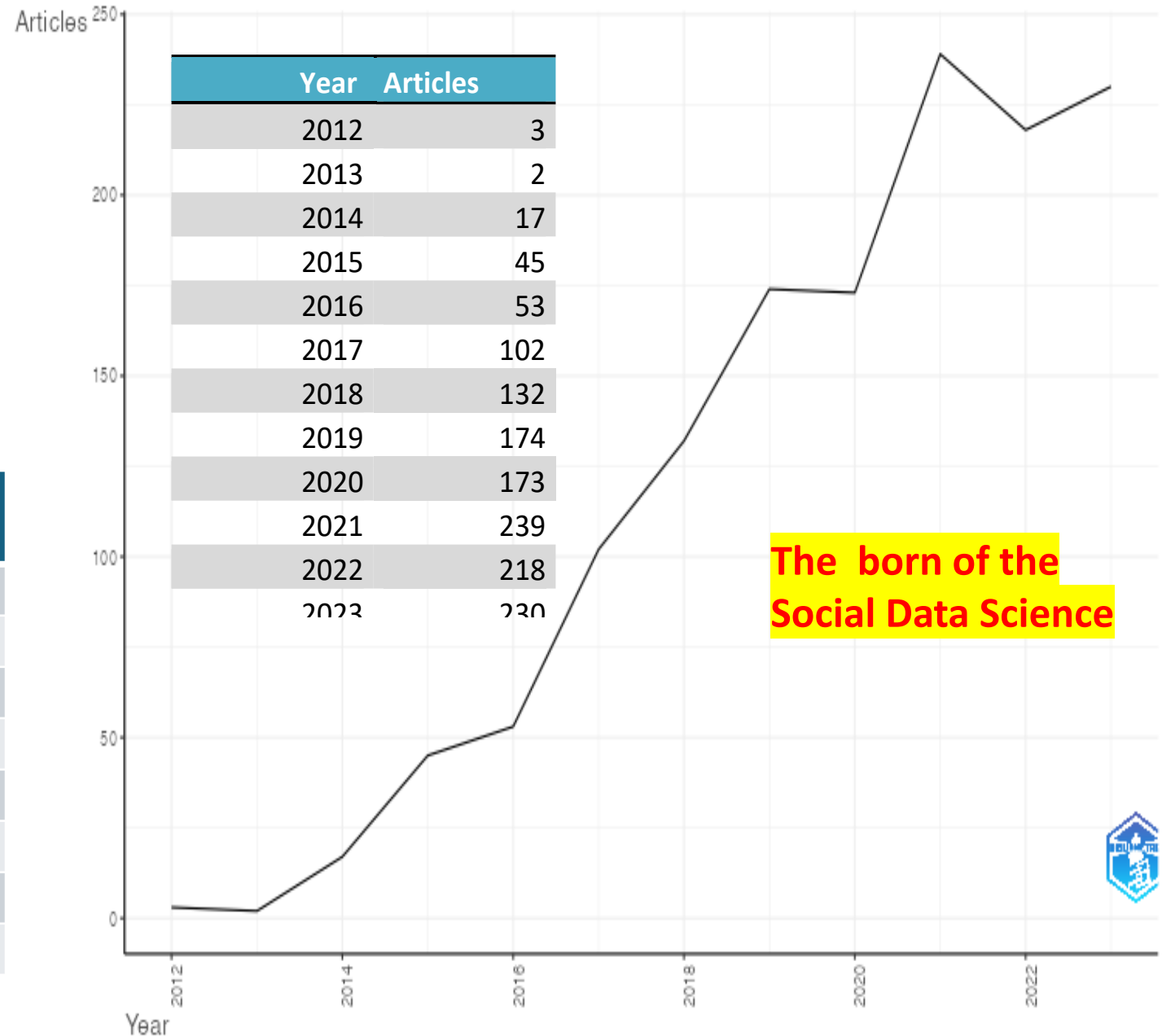
4th international conference

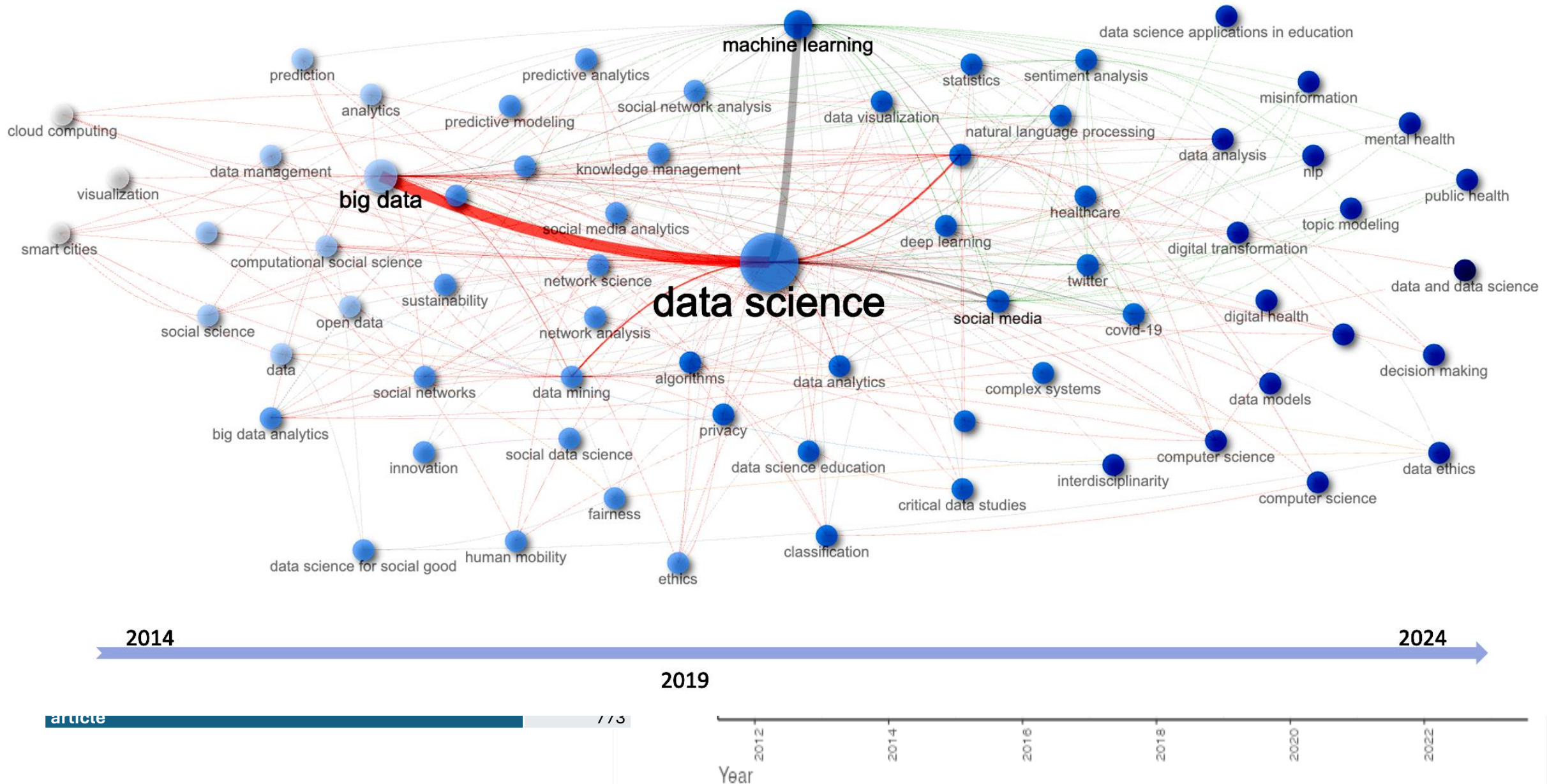
Naples (Italy) on **25th-27th March 2024**

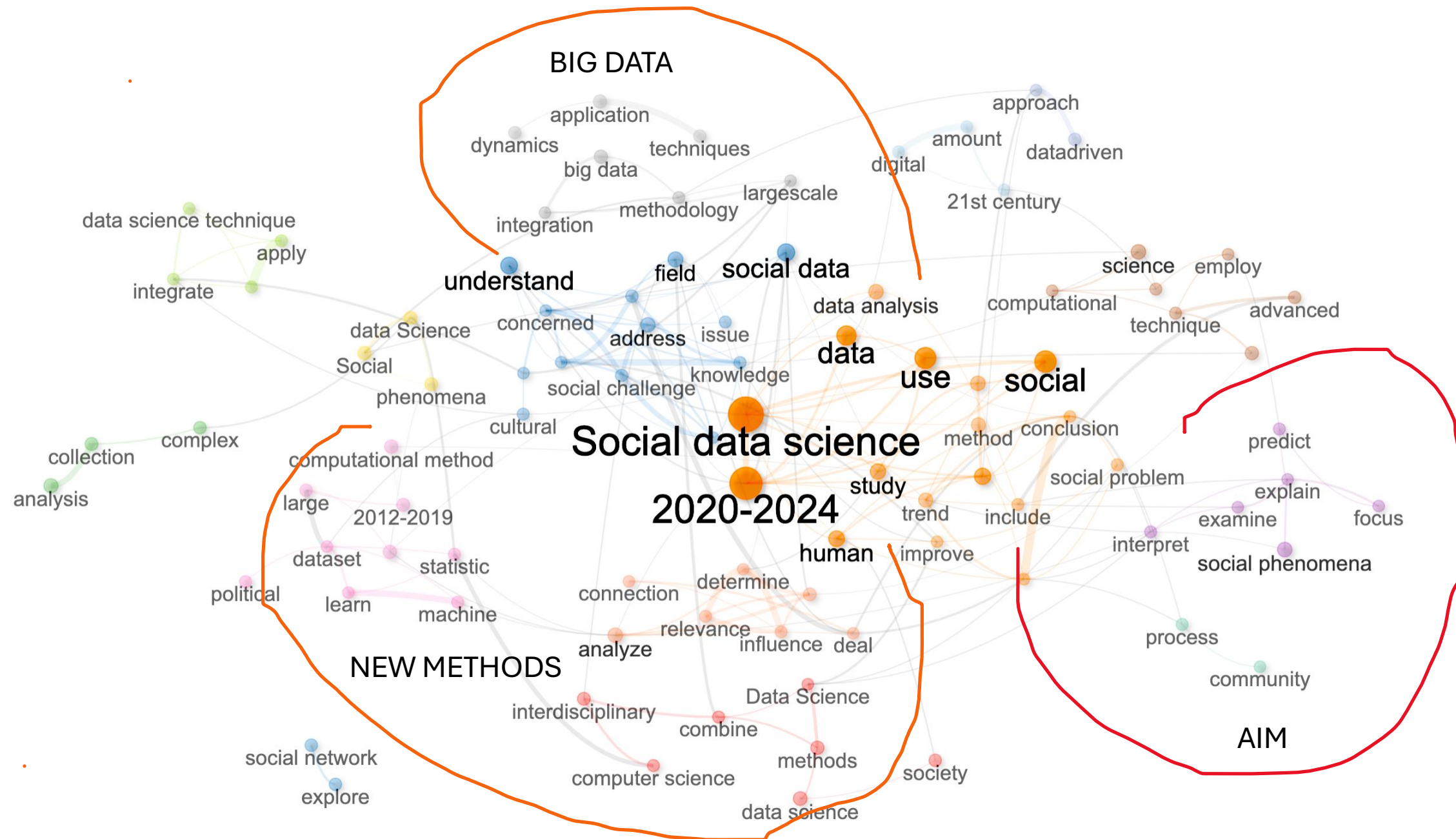
What happens later.....

SDG key words in Web of Science Documents 2012:2024	
Sources (Journals, Books, etc)	967
Documents	1409
Annual Growth Rate %	17,6
Average citations per doc	13,39
References	63655
Keywords Plus (ID)	1927
Authors	5001
article	773

Annual Scientific Production





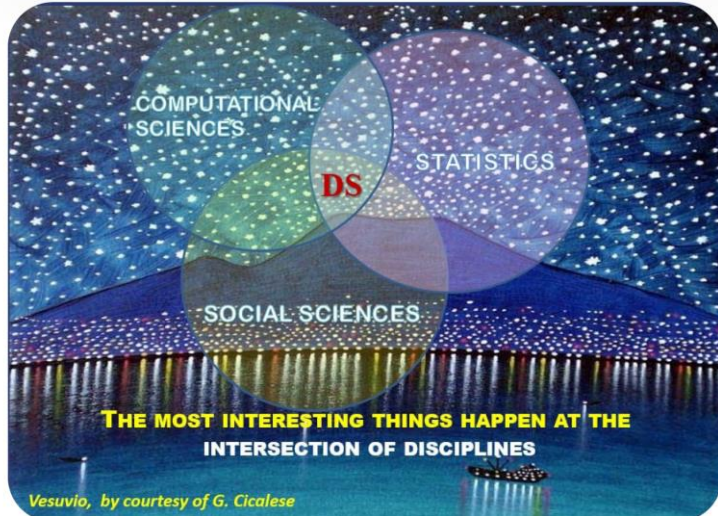


DSSR 2024

Data Science & Social Research

4th international conference

Naples (Italy) on **25th-27th March 2024**



<https://dssr2024.unina.it>

Data Science, represented by a circular triangle, is an interdisciplinary approach, designed to address the emerging challenges presented by the **Information Society**. It is mainly based on methods and algorithms of **Statistics and Computational Sciences** (including AI), suitably integrated by the Knowledge of a specific domain, the Social Sciences in our case. These three areas represent in particular the **logic, language and object** of what we refer to as **Social Data Science**.

Intersection by couple of scientific area are respectively defined as **Computational Statistics, Algorithmic Data Analysis and Statistical Data Analysis**. Accordingly the **SDG employs and/or develops appropriate methodologies and algorithms** with the aim of discovering new knowledge, making predictions also for future scenarios and supporting decision-making processes in an increasingly complex landscape.



The Data Science curvilinear triangle

a DS definition by Carlo Lauro
Some CS tools:

Data extraction and cleaning; Data Warehousing; Optimization and numerical algorithms; Matrix algebra Simulation; High Performance Computing; Cloud Computing R; Hadoop; Python; SAS; Rapid Miner; Tableau;; Visualization ; Data Mining; A. I.; Machine Learning ANN; Deep Learning;.....

Algorithmic Data Analysis

CDA ->

Data = Algorithm + Accuracy

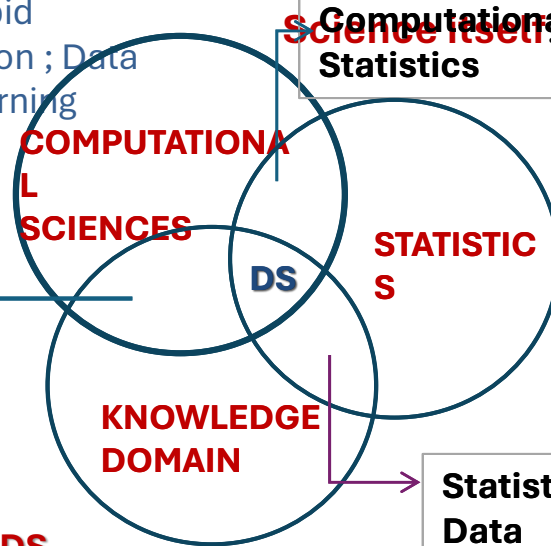
The role of Knowledge in DS

Data Science = Knowledge based or “Intelligent” Computational Statistics

Data Science = “Intelligent” Computational or Statistical Data Analysis Breiman

asspts and/or develops appropriate methodologies for purposes of knowledge discovery, prediction and decision-making in the face of an increasingly complex reality often characterized by large amounts of data (big data) of various types (numeric, ordinal, categorical, symbolic, texts, images, data streams, multi-way, structured , networks , functions, etc.), coming from disparate sources (experiments,surveys, official data, media, sensors, transactions, open data . etc.). According to the knowledge domain we have a Social DS , Healthcare DS, Business DS, Environmental

Data Science (DS) is an interdisciplinary approach to meet the challenges of the Information Society, based on the methods of **Computational Science (CS)** and **Statistics** supplemented by **Knowledge of different domains**. **Statistics represents the logic of the Data Science, whereas Computational Science is its language. Knowledge of various domains of interest is the prerequisite of Data Science itself.**



Computational Statistics

Some Stat tools:

Exploratory and inferential methods ; Probability, Density estimation; Regression; Time series; Causal Models and SEM; Bayesian models; Factorial analysis and PCA; Cluster analysis; Classification; SNA

Some Knowledge representation

tools: Metadata; Logical rules; Hierarchical ruleAlgebrical or geometrical constraints; Probability models; Network; Ontologies

Knowledge Graphs:

SDA -> Data = Model + Error
 (The 2 cultures,

DSSR 2024

Data Science & Social Research

4th international conference

Naples (Italy) on **25th-27th March 2024**



<https://dssr2024.unina.it>

The complexity of the **SDG** often arises from the presence of **large volumes of different types of data** (numerical, ordinal, nominal, symbolic, text, images, multidimensional data, social networks, knowledge graphs, etc.) **from various sources** (surveys, official data, open data, social media, web, sensory data, transactions, etc.).

Aim of the IV DSSR24 international conference is to describe the opportunities that the "data revolution" offers to social research, through statistics and computational sciences in the framework of Data Science. The availability of new data offers great opportunities for social research to enrich knowledge – and face the new challenges of the future – on some key areas of research such as sustainable development, social inequalities, public health, communication and future studies.

