



**Ministry of Science
and Technology**



Bar-Ilan University



**Fondation
France-Israël**

Save the date - WORKSHOP

NEW WAY OF THINKING IN DATA SCIENCE:

The Symbolic Data Analysis Paradigm for Big and Complex Data.

From units to classes and classes considered as new units.

RESEARCH – INDUSTRY

BAR ILAN UNIVERSITY (Ramat-Gan)

15 January 2019

**Building 206 – Nanotechnology Complex
Room C-50**

Free of charge for registered participants, the number of places being limited.

Registration: MaorT@most.gov.il

(Name, Institution, department, position email and mobile number requested).

In all domains of human activity, we are more and more faced with the problem of understanding and extracting knowledge from standard, big and complex data, often multi-sources (with mixture of numerical, textual, image, social networks data).

Data Science, considered as a science by itself, is in general terms, the extraction of knowledge from data.

Symbolic Data Analysis (SDA) gives a new way of thinking in Data Science by extending the standard input to a set of classes of individual entities. Symbolic Data Analysis (SDA) is an emerging area of Data Science based on aggregating individual level data into group-based summarized by symbols, and then developing Data Science methods to analyze them. It is ideal for increasing the explanatory power of machine learning in the case of standard, large and complex datasets, and has immense potential to become a standard methodology in the near future.

Several books and numerous papers has already been published: Bock and Diday (Springer, 2000), Billard and Diday (JASA (2003), Wiley books (2006, 2019)), Diday and Noirhomme-Fraiture (Wiley 2008), Afonso, Diday, Toque (Technip, 2018), etc.

Over the world, teams are developing this domain in America (USA, Brazil, Costa Rica), in Asia (China, Japan, Korea, Taiwan), in Europe (France, Italy, Spain, Portugal, Poland, Slovenia), in Australia (University of New South Wales, Sydney). They are developing theory and practice in a great variety of domains. SDA has been recently applied in health, genetics, cybersecurity, socio-demography, epidemiology, economy, finance, trade, structural health monitoring (bridges, railways, nuclear power plot, concrete structural material, etc.), text mining, agriculture, education.

This workshop aims to introduce the field (at the crossroad of statistics, mathematics and computer science) to academic researchers and industrials in all domains where data are obtained and need to be analyzed for understanding them and so improving decisions. Several open theoretical SDA questions and open data applications will be presented.

Its aim is also to create interest on this topic in Israel, leading to national France-Israel and international cooperation.

The workshop speakers are:

- **Edwin Diday** (Paris-Dauphine University) is the founder of the SDA domain and has been awarded by the French Academy of Science (Montyon prize). He will present an “*Overview on SDA and Future and recent advances for ranking complex objects singularity and improving the explanatory power of Machine Learning*”
- **Lynne Billard** (University of Georgia, USA) has served as president of the American Statistical Association, and the International Biometric Society, one of a handful of people to have led both organizations. She will present: “*The SDA Statistical Foundation*”.
- **Richard Emilion** (Orleans University, France). He is a professor of applied mathematics and has a long experience on the “*Probabilistic foundation of SDA*” that he will present.
- **Oldemar Rodriguez** (San José University, Costa Rica). He has a long collaboration with Stanford University (USA). He has applied SDA to Bank Security and sold his company to a big American company. He manages now a new company. With his team he developed the SDA methodology and a popular R Symbolic Data Analysis software called R-SDA. He will present: “*Advances in SDA and several applications including Bank Security and the R-SDA package*”.
- **Filipe Afonso** has a PhD in Computer Science from Paris-Dauphine University. He has long experience in industrial SDA applications and has developed the SDA industrial software called SYR with many industrial applications. SYR is complementary to R-SDA. He will present “*Industrial, socio-demographic and medical applications of SDA and the SYR software*”.
- **Myriam Touati** Engineer at CEREMADE Paris-Dauphine, now at Jerusalem, from where she continues her cooperation. She was one of the pioneers of the academic SDA software SODAS that she will present with numerous analyzed symbolic data samples.

After those talks we will organize a training of SDA software on your laptop illustrated by simple examples.

SCIENTIFIC PROGRAM

Tuesday, January 15, 2019

8:00 – 8:30 Registration

8:30 – 9:00 Opening Greetings:

Dr. David Harari, co-Chairman of the High Council Israel-France for Science & Technology

Dr. Eli Even, Head of Research Authority, Bar Ilan University

Dr. Muriel Haïm, President of the France-Israel Foundation

Prof. Alexander Bligh, Chief Scientist, Ministry of Science and Technology

SESSION 1: THEORETICAL AND PRACTICAL FOUNDATION OF SYMBOLIC DATA

9:00 – 9:45 E. Diday (Paris-Dauphine University, France)

Overview on the theory and practice of SDA (illustrated by examples in: socio demography, insurance companies, nuclear power plant, epidemiology).

9:45 – 10:30 L. Billard (UGA University, USA)

Statistical foundation and practice of SDA (illustrated by examples in: medicine, biology and economy)

10h30 – 11h Coffee break

11:00 – 11:45 R. Emilion (Orléans University, France),

Probabilistic foundation of SDA. Likelihood on symbols illustrated by applications in renewable energy.

11:45 – 12:30 O. Rodriguez (Costa-Rica University), F. Afonso (Paris-Dauphine University, France)

Basic tools in SDA: symbolic extension of dissimilarities, clustering, factorial analysis, shrinkage regression, decision trees, and reduction methods for symbolic variables

12h30 – 13h15 Lunch

SESSION 2: INDUSTRIAL APPLICATIONS AND SOFTWARE

13:15 – 14:00 F. Afonso (Paris-Dauphine University, France), O. Rodriguez (Costa-Rica University),

SDA tools and Industrial applications in cybersecurity, health, structural health monitoring, finance and marketing.

14:00 – 14:45 Rodriguez (Costa-Rica University)

The RSDA software and Industrial applications in bank security ...

14:45 – 15:30 F. Afonso, M. Touati (Paris-Dauphine University, France)

The SYR software through an application in official statistics

The SODAS software and Symbolic Data Bases in Education, Health, Countries Development, Marketing, Road Safety, Sports, Hobbies...

15:30– 16:00 ROUND TABLE open direction of research and cooperation.

Example: in France we have huge databases due to the social security management, from which the effect of the environmental conditions on disease can be studied.