INRAe

MIMS

SCIENTIFIC NETWORK MIMS

2022-2023

Coordination
Mohamed Hanafi
StatSC
mohamed.hanafi@oniris-nantes.fr
Jean Michel Roger
UMR ITAP
jean-michel.roger@inrae.fr

INRAE units involved
StatSC
BIA
QuaPA
SPO
LBE
ITAP
MaIAGE
CSGA Centre des Sciences du Goût et de l'Alimentation
UNH Unité Nutrition Humaine
PhAN
LABERCA
Micalis
Prose
BioForA
LBLGC
AGAP Institut
SELMET

Partnerships
Faculté des Sciences, Paris
INRIA
University of Genève
University of Toulouse
ANSES
CNAM
University of Paris-Saclay
University of Montpellier
ADLIN
French Wine and vine Institut

Cross Methodological Insights for Multi-source Data Integration

Context and challenges

In biology, as in other scientific fields, the integration of multi-source data is more relevant than ever. Indeed, the data collected are increasingly complex and their volume is growing, due to the development of analytical platforms, imaging techniques, the rise of omics data, etc.

This context has stimulated the search for new methods allowing the joint analysis of several data sets (structured data, multi-block, multi-channel) in many fields, such as:

- Machine Learning, where several approaches are considered for the processing of multi-source data (matrix factorisation, probabilistic approach).

- Chemometrics, where different methods are proposed to establish a chemical mapping of samples using several analytical techniques (generalisations of canonical analysis, NIPALS algorithm and tensor decompositions)

- Bioinformatics, where integrative methodological approaches allow the most complete picture possible of the dynamics of molecular systems to be drawn.

In order to contribute to meeting the challenge of analysing and exploiting these multi-source data from an exploratory, but also predictive perspective, it is essential to bring together different viewpoints, practices and paradigms in order to reconcile these different approaches. It is also necessary to encourage collaboration between "method generators" and "data generators" in the various application fields.

This is the challenge that the MIMS consortium proposes to take up, by bringing together an interdisciplinary community working on approaches to the analysis and integration of multi-source data.

Logo MIMS

Métaprogramme
DIGIT-BIO

digitbio@inrae.fr
www.inrae.fr/digitbio/

Goals

MIMS is a multidisciplinary consortium gathering more than 60 researchers, whose objective is to examine the analysis and exploitation of multi-source data, both in an exploratory and predictive perspective.

This consortium brings together multidisciplinary skills: information processing, biological sciences and analytics. The implementation of this multi-disciplinarity and its management will be based on the sharing of data, practices and methods between the partners, with the aim of formalising a scientific project to meet a common challenge: the opti-mal analysis of multi-source data for exploratory and predictive purposes.

Research units involved and partnerships

INRAE scientific division	INRAE research units	Expertises
Sciences for food, bioproducts and waste engineering	USC StatSC	Sensometry, Chemometrics, Statistics, Multispectral imaging
	BIA	Chemometrics, computer science
	QuaPA	Volatolomics, MRI Chemometrics, Data Analysis, Image Analysis, System & Data Management
	SPO	Chemometrics
	LBE	Biostatistics, machine learning
Mathematics, computer and data sciences, digital technologies	ITAP	Chemometrics
	MAIAGE	Mathematical statistics, applied statistics, bioinformatics
Human nutrition and food safety	CSGA Centre des Sciences du Goût et de l'Alimentation	Chemometrics
	UNH Unité Nutrition Humaine	Bioinformatics, metabolomics, chemometrics
	PhAN	Perinatal nutrition and metabolic diseases, Bioinformatics, Data analysis, metagenomics and metabolomics
	LABERCA	Metabolomics, Chemometrics, Expology, Epidemiology
Microbiology and the food chain	Micalis	Biologist, Microbiota, Data Analysis
	Prose	
Ecology and biodiversity of forest, grassland and freshwater environments	BioForA	Quantitative Genetics, Modelling
	LBLGC	Physiology
Plant biology and breeding	AGAP Institut	Quantitative genetics, Genomics, Biochemistry, Evolutionary genetics, Selection, Ecophysiologist, Biostatistics, Bioinformatics

Animal physiology and livestock systems	SELMET	Biometrics, Chemometrics, Machine Learning, Agronomy

External partners		Expertises
Faculté des Sciences, Paris	Centre Boreli	Unsupervised learning, Statistics, Graph networks, Bioinformatics
INRIA	Équipe projet LORIA	Knowledge Discovery, Life Sciences
University of Genève	Sciences Analytiques	Metabolomics, Chemometrics
University of Toulouse	Institut de mathématique de Toulouse	Statistics, Multi-omics data analysis and integration
ANSES	Laboratoire de Ploufragan-Plouzané	Statistics, multi-block methods Epidemiology
CNAM	EPN6 - Mathématiques et Statistique	Analysis of complex heterogeneous data, Clusterwise methods, High dimensional classification
University of Paris-Saclay	Signaux et Statistique	Multi-block data analysis, tensor analysis (high dimensional), Structural equation models
University of Montpellier	Institut Montpelliérain Alexander Grothendieck	Supervised component models, classification
ADLIN	ADLIN	Finance, Strategy, Multi-omics, Bioinformatics, Transcriptomics, Visualisation
French Wine and vine Institut	IFV	Chemometrics, Analytical Chemistry

References

- Bersanelli, M, Mosca E, Remondini D. et al. Methods for the integration of multi-omics data: mathematical aspects. BMC Bioinformatics 17, S15 (2016). https://doi.org/10.1186/s12859-015-0857-9
- **Boccard J, Schvartz D, Codesido S, Hanafi M, Gagnebin Y, Ponte B, Jourdan F, Rudaz S. (2021).** Gaining insights into metabolic networks using chemometrics and bioinformatics: chronic kidney disease as a clinical model. *Frontiers in Molecular Biosciences* 8, 682559.
- **Boutalbi R, Labiod L, Nadif M. (2021):** Implicit consensus clustering from multiple graphs. D*ata Min. Knowl. Discov.* 35(6): 2313-2340
- **Hanafi M, Kiers H. A. L. (2006).** Analysis of K sets of data, with differential emphasis on agreement between and within sets. *Computational Statistics and Data Analysis.* (51), 3, 1491-1508.
- Eicher T, Kinnebrew G, Patt A, Spencer K, Ying K, Ma Q, Machiraju R, Mathé AEA. Metabolomics and Multi-Omics Integration: A Survey of Computational Methods and Resources. Metabolites. 2020 May 15;10(5):202. doi: 10.3390/metabo10050202. PMID: 32429287; PMCID: PMC7281435.

Métaprogramme
DIGIT-BIO

digitbio@inrae.fr
www.inrae.fr/digitbio/