



Big Data and Artificial Intelligence applied to Healthcare and Biology (BigArt-HuB)

First InPrint Seminar

Partnership Building towards

Stronger Engagement in International Collaboration


Gisele L Pappa, CS Department, UFMG

UFMG, Belo Horizonte

19-23 November 2018



How PrInt/UFMG changes the ways we collaborate

 Matrix of the Institutional Project of Internationalization

Themes → contemporary challenges

	Sustainability, Risk Management and Governance	New Technologies and Frontiers of Science	Health and Well-being	Human Rights
Aging			*	
Agro- and bio-business		*		
Basic science and its applications		*		
Big data and artificial intelligence		*	*	*
Biotechnology		*		
Borders and migrations	*			
Chronic, emergent and neglected diseases			***	



PrInt/UFMG mobility goals

- To foster and enhance **collaboration** with partner universities worldwide by means of joint innovative research and capacity building of human resources
- To consolidate the **training and experience abroad of UFMG faculty members** with both a junior and a senior profile as visiting professors at partner universities
- To **train PhD students abroad** through internships at partner universities with a **strong focus on cotutelle** (double PhD degrees)
- To **recruit postdoctoral fellows and early-career researchers** with experience abroad to work at UFMG within the PrInt themes and projects
- To **attract internationally renowned visiting professors** with highly recognised experience for short stays (15 days) at UFMG



PrInt/UFMG mobility actions

OUTGOING

- PhD mobility grants for internships abroad (six to twelve months)
- Junior Professor grants for visiting professorships abroad (six months)
- Senior Professor grants for visiting professorships abroad (six months)

INCOMING

- Postdoctoral grants for activities at UFMG (12 months, renewable)
- Early-career researcher grants for activities at UFMG (12 months, renewable)
- Senior Professor grants for international visitors at UFMG (15 days)



UFMG Graduate programs taking part in Big Data and Artificial Intelligence applied to Health and Biology

- Applied Linguistics
- Biochemistry and Immunology
- Chemistry
- Clinical and Toxicological Analysis
- Computer Science
- Economics
- Knowledge Organization and Management
- Physiology and Pharmacology
- Production engineering
- Public Health
- Rehabilitation Sciences
- Statistics
- Tropical Medicine



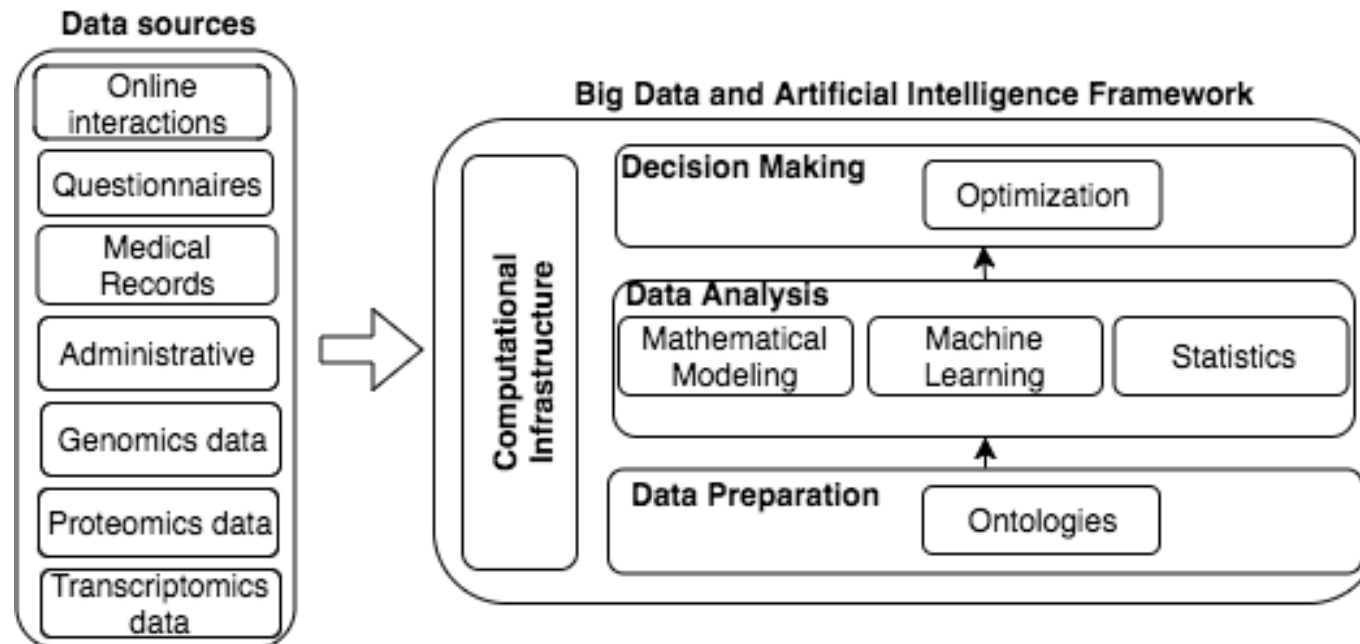
Partner institutions willing to collaborate with UFMG in this project (so far)

- University of Auckland
- University of Lille
- University of Southampton

New collaborations will be most welcome!

What do we mean by Big Data and Artificial Intelligence applied to Health and Biology?

Conceptual framework:





What do we mean by Big Data and Artificial Intelligence applied to Health and Biology?

Existing competencies and expertise are divided into 3 groups :

- Data collection, data understanding and related health/biology problems
- Methods/Techniques related to the BDAI framework phases
- Evaluation of health services and health policies



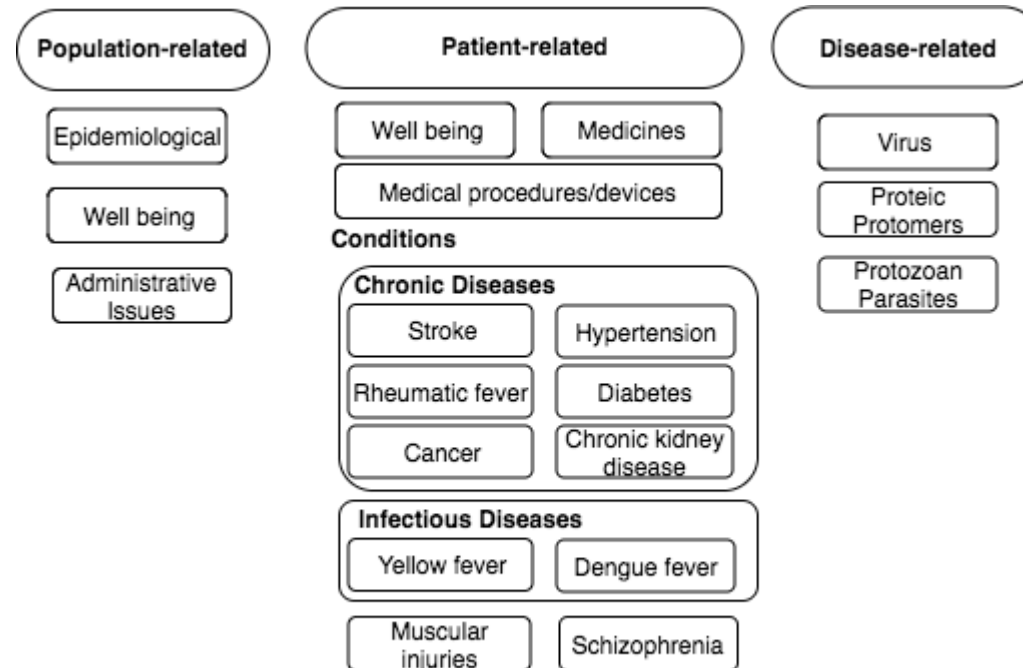
What do we mean by Big Data and Artificial Intelligence applied to Health and Biology?

Existing competencies and expertise are divided into 3 groups :

- Data collection, data understanding and **related health/biology problems**
- Methods/Techniques related to the BDAI framework phases
- Evaluation of health services and health policies

What do we mean by Big Data and Artificial Intelligence applied to Health and Biology?

Problems are organized in 3 main groups:





What do we mean by Big Data and Artificial Intelligence applied to Health and Biology?

Data collection and understanding related to the BDAI framework

1. Data Preparation using ontologies
 - a. Data integration
 - b. Data enrichment
 - c. Data provenance
2. Semantic Graph creation
 - a. Network of objects that instanciate data



What do we mean by Big Data and Artificial Intelligence applied to Health and Biology?

Examples of data types involved in these problems:

1. Medical equipment geolocation
2. EQ-5D-3L questionnaires
3. Transcription of recorded patient - healthcare provider interaction
4. Administrative and epidemiological records from the Brazilian population using the Public Health System (SUS): Real World Evidences
5. Electrocardiogram recording and echocardiogram videos
6. Data from training and other neuromusculoskeletal aspects of athletes
7. RNA, DNA, Genes, Proteins (Antibodies)
8. Nuclear Magnetic Resonance data - image and solution data
9. Thermal analysis data of drugs and medicines.



What do we mean by Big Data and Artificial Intelligence applied to Health and Biology?

Existing competencies and expertise are divided into 3 groups :

- Data collection, data understanding and related health/biology problems
- **Methods/Techniques** related to the BDAI framework phases
- Evaluation of health services and health policies



What do we mean by Big Data and Artificial Intelligence applied to Health and Biology?

Methods/Techniques related to the BDAI framework phases

1. Machine Learning
 - a. Expertise in both classification, regression, clustering and association rule mining
 - b. Deep learning
 - c. Automated Machine Learning (AutoML)
2. Text mining
3. Bayesian statistics



What do we mean by Big Data and Artificial Intelligence applied to Health and Biology?

Methods/Techniques related to the BDAI framework phases

4. Dealing with time series data, geolocated data, videos, high volumes of text, data with more columns than lines (microarray), among others.
5. Optimization/evaluation of health techniques/technologies to help the decision making process using real-world evidences.



What do we mean by Big Data and Artificial Intelligence applied to Health and Biology?

Existing competencies and expertise are divided into 3 groups :

- Data collection, data understanding and related health/biology problems
- Methods/Techniques related to the BDAI framework phases
- Evaluation of health services and health policies



What do we mean by Big Data and Artificial Intelligence applied to Health and Biology?

Evaluation of health services and health policies:

1. Evaluation of health services dealing with chronic kidney disease, cancer, mental health, among others
2. Analysis of the effectiveness and expenditures of health policies
3. Evaluation of technologies for treatment and high cost medicines
4. Evaluation of educational interventions towards patient self-management and empowerment



What do we mean by Big Data and Artificial Intelligence applied to Health and Biology?

Short, medium and long-term goals:

- Develop new statistical/machine learning methods to deal with the intrinsic characteristics of medical data
- Improve the computational infrastructure to store and manipulate large amounts of data
- Develop new technologies for data acquisition and manipulation
- Be able to take advantage of a large amount of data to improve public health, individual health and the Brazilian Public Health System in general
- Develop new collaborations with other institutions with similar and complementary expertises/interests

Research issues and questions

Ongoing projects and related research questions

- **Computer Science**
 - How to choose the best set of data analysis tasks that should be performed in a given dataset?
 - Develop Auto-ML techniques able to generate customized task flows for data analytics in medicine.
- **Knowledge Organization and Management**
 - How to systemize data preparation allowing better reproducibility?
 - Use of semantic annotation techniques based on ontologies

Research issues and questions

Ongoing projects and related research questions

- **Production engineering**

- How to improve planning and operationalization of services in hospitals?
 - Explore optimization techniques for combinatorial problems to better plan hospital schedules.

- **Statistics**

- What is the time until the occurrence of an event of interest, such as the detection of a disease in an individual?
 - Propose (semi-)parametric methods to treat specific characteristics of health data, including recurrent events, longitudinal covariates, among others.

Research issues and questions

Ongoing projects and related research questions

- **Applied Linguistics**

- Can we find discourse Indicators of patient empowerment and self-management efficacy in chronic diseases (diabetes and sickle cell disease)?
 - Use of text mining techniques

- **Rehabilitation Sciences**

- Can we predict the occurrence of musculoskeletal injuries in athletes from biomechanical, physiological and psychological data collected during the competition season?
 - Regression and predictive methods to deal with time-series data.



Research issues and questions

Ongoing projects and related research questions

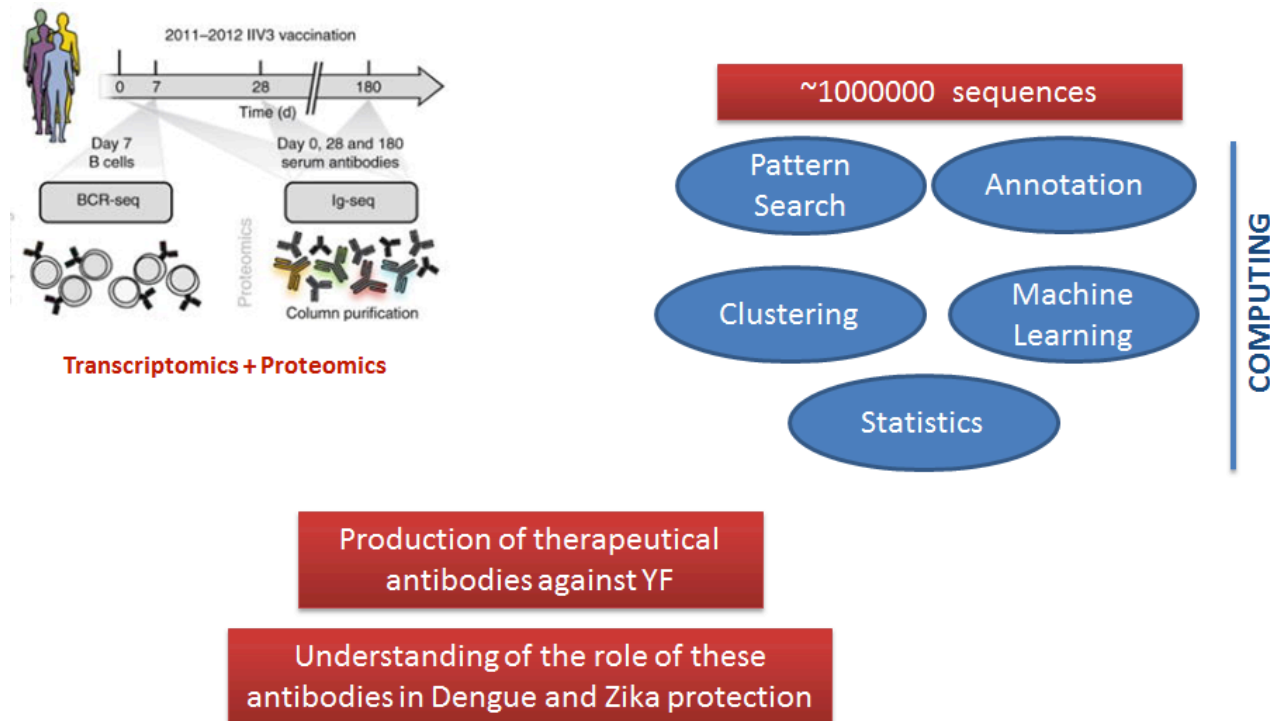
- **Biochemistry and Immunology**

- What are the antibodies activated by the yellow fever vaccine? Can they be used for yellow fever treatment?
 - Statistical analysis of the nucleotides sequences that codify the antibodies.

- **Chemistry**

- What are the kinetic mechanisms for thermal decomposition of medicines, drugs and carbon nanotubes?
 - Methods for non-linear analysis of experimental results

Molecular analysis of antibodies elicited after Yellow Fever vaccination



Research issues and questions

Ongoing projects and related research questions

- **Public Health**

- Can we use administrative and epidemiological data to evaluate the effectiveness and costs of cancer, chronic kidney disease, hospitalization of mental disease, coronary arterial disease, among others?
 - Use integrated administrative and epidemiological data to evaluate the treatment and medicines of high cost



Ongoing and Prospective Collaborations

- Boston University
- Carnegie Mellon University
- CNRS
- George Washington University
- North Carolina State University
- Universidade de Coimbra
- Università degli Studi di Milano
- University of Bremen
- University of Cincinnati
- University of Kent (UK)
- University of Melbourne
- University of Pittsburgh
- University of Southampton
- University of Sydney
- University of Texas, Austin
- Università di Padova
- Vrije Universiteit (Amsterdam)



Concluding remarks for Big Data and Artificial Intelligence applied to Health and Biology

- UFMG has different expertises in the area of Big Data and Artificial Intelligence applied to Health and Biology, mainly divided into 3 groups:
 - Data collection, data understanding and related health/biology problems
 - Methods/Techniques related to the BDAI framework phases
 - Evaluation of health services and health policies
- In the long term, we want to be able to take advantage of a large amount of data to improve public health, individual health and the Brazilian Public Health System in general
- Develop new collaborations with other institutions with similar and complementary expertises



Big Data and Artificial Intelligence applied to Healthcare and Biology (BigArt-HuB)

First InPrint Seminar

Partnership Building towards

Stronger Engagement in International Collaboration

Gisele L Pappa, CS Department, UFMG

UFMG, Belo Horizonte

19-23 November 2018