

Project:
SanDAL

(Grant Agreement number 811017)

*ERA Chair in Mathematical Statistics and Data Science for the
University of Luxembourg*

Funding Scheme: Coordination and Support Action
Theme: WIDESPREAD-03-2017 - ERA Chairs
Date of latest Annex I: 22/05/2018

D2.1 Strategic Research Programme – initial version

PC Organization Name: University of Luxembourg
Deliverable Due Date: 31/10/2019
Deliverable Issue Date: 31/10/2019

Document History <i>(Revisions – Amendments)</i>	
Version and date	Changes
1.0 – 30/10/2019	First version

Dissemination Level		
PU	Public	X
PP	Restricted to other program participants (including the EC Services)	
RE	Restricted to a group specified by the consortium (including the EC Services)	
CO	Confidential, only for members of the consortium (including the EC)	

SanDAL aims to create at the University of Luxembourg (UL) a high-level research group in Mathematical Statistics and Data Science (MS&DS). This will fill an identified gap between UL's current research activity in mathematics (including probability theory but also geometry and mathematical physics), which developed well in recent years, and a number of applications of particular importance both for other academic research fields and for Luxembourg's economic development.

The ERA Chair will be offered a permanent full professorship at UL to ensure a sustainable structural change and improvement of Luxembourg's Research and Innovation landscape. The project will also fund two postdoctoral positions for its full duration, and three PhD positions will be UL-funded. The main scientific directions of the project (to be precisely defined by the future ERA Chair Holder) will be High-Dimensional Data Analysis (including Probabilistic Aspects of High-Dimensional Data and Topological Data Analysis) and New Mathematical Tools for Contemporary Statistics (including Signal Reconstruction and the Mathematics of Sparsity, and Non-Parametric Bayesian Methods).

In addition to the development of a strong and sustainable new direction of research at UL, SanDAL will permit the development of new teaching programs at MSc and PhD levels in MS&DS. It will therefore ensure a steady flow of MS&DS graduates that will be of value not only for mathematics but also for academic and non-academic research in other fields and in other economic areas. SanDAL will build on the existing expertise in mathematics (in particular probability theory, but also discrete geometry and topology) at UL. The contribution of an external Scientific Steering Committee, made of top experts in MS&DS, will be essential to its scientific success, while a National Advisory Board, made of some of the main users of MS&DS at the Luxembourg level, will ensure that it is relevant to the economy and other academic fields.

LEGAL NOTICE

Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use, which might be made, of the following information.

The views expressed in this report are those of the authors and do not necessarily reflect those of the European Commission.

© H2020 SanDAL Project 2019

Reproduction is authorised provided the source is acknowledged



Table of Contents

Introduction	4
Purpose of the Strategic Research Programme (SRP)	4
Strengths and areas for improvement	6
SanDAL research priorities	7
Establishment of the postgraduate training programme	9
Allocation of resources to research priorities	10
Guidelines for strategic publications	10
National and European (EU) Funding Opportunities	11
Dissemination and Exploitation Potential	12
Conclusions and recommendations	13

Introduction

One of the principle responsibilities of the SanDAL project is to produce a Strategic Research Programme and Postgraduate Training (PhD and MSc) Programme for the ERA Chair in Mathematical Statistics and Data Science (MS &DS) at the University of Luxembourg (UL). This document has been prepared as an early draft outlining the aims and specific objectives to assess the needs of the institute and to maximise its overall scientific excellence in the MS&DS research sub-topics. Following the recruitment of the ERA Chair, this document has been prepared and submitted as “D2.1 Strategic Research Programme-initial version”.

In this document, the Strategic Research Programme will also highlight the plan for the SanDAL Postgraduate training programme. The ERA Chair Team will be responsible for developing PhD training in MS&DS within UL’s Doctoral School in Science and Engineering. Currently, the main specialities of the Doctoral School include Mathematics and Applications, Physics and Materials, and Systems and Molecular Biomedicine. With the support of UL’s Engineering Science Research Unit and Mathematics Research Unit, a specialisation and critical mass of PhD candidates will be established in MS&DS.

The Strategic Research Programme serves as an essential component in defining the specific objectives which will target the research priorities and the scientific activities which will be carried out to foster excellent research in MS&DS. It defines the guidelines for achieving UL’s structural change and to maximise the potential and opportunities available to succeed over the next 5 (at least) years.

The contents of the Strategic Research Programme will detail the following items:

- Purpose of the Strategic Research Programme
- Strengths and areas for improvement
- SanDAL research priorities
- Establishment of postgraduate training programme
- Allocation of resources to research priorities
- Guidelines for strategic publications
- National and European (EU) funding opportunities
- Dissemination and Exploitation Potential

Purpose of the Strategic Research Programme (SRP)

The Strategic Research Programme has been created to demonstrate the goals, objectives and strategies which will be undertaken to fulfil the vision of the SanDAL project, which is to achieve scientific excellence in MS&DS at the University of Luxembourg. It will also support the structural changes expected within the Mathematics Research Unit in the UL’s Faculty of Science, Technology and Communication to maximise the academic and professional growth at the institution.

The mission of the SanDAL project will be to bring about necessary structural changes in the UL’s Mathematics Research Unit by:

- a) Boosting its research and training capacity in MS&DS and
- b) Forging tight links and providing MS&DS support services to local stakeholders in the logistics, healthcare, ICT, manufacturing and finance sectors.

The objective of the project is the creation of a team (3 PhD students, 2 postdoctoral fellows) in the area of Mathematical Statistics and Data Science (MS&DS) in Luxembourg. The team will achieve a very high quality of research and publish in the best journals of the field (Annals of Statistics, Probability Theory and Related Fields, Annales de l’Institut Henri Poincaré, Bernoulli, Electronic Journal of Statistics, Mathematical Statistics and Learning, among others).

In addition, they will also participate in a diverse set of dissemination and exploitation activities to promote the research excellence in MS&DS, including presenting their research at international conferences, initiating collaborations at different events and attracting new, talented researchers to the University of Luxembourg, specifically to the Mathematics Research Unit.

The team will invite renowned Statisticians for collaborations, seminars and conferences. This will contribute to bridging the team to other strong research centers in Europe and abroad, widening scientific partnerships.

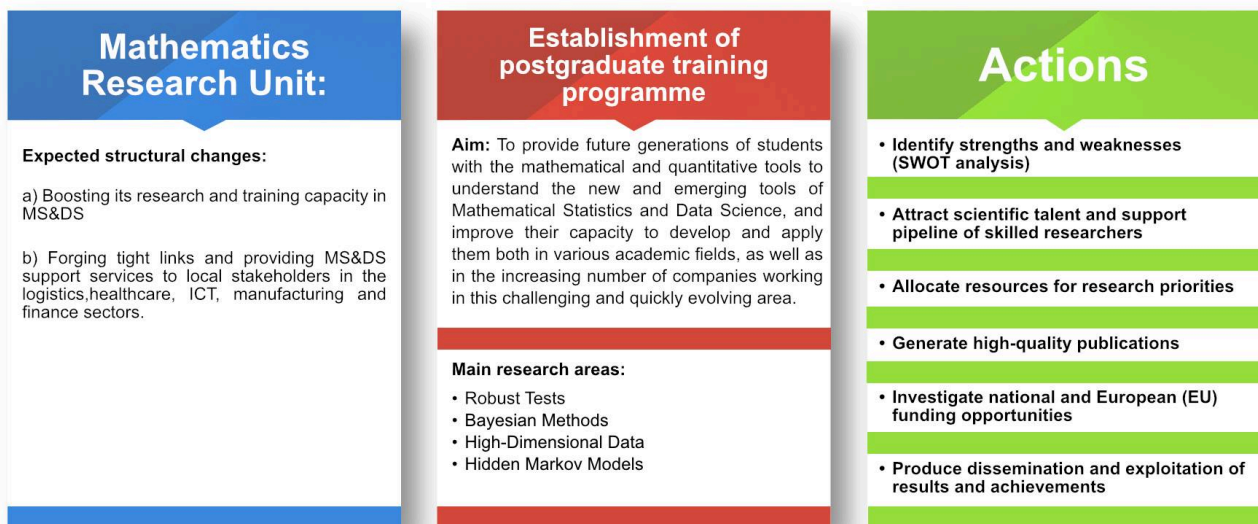
In view of facilitating citations and disseminations of the results, the published papers will be available **in open access on Arxiv, ORBilu and the webpages of the team.**

The team will also support the development of an educational programme for the University of Luxembourg in the field of Statistics and Data Science (Master and PhD levels), encouraging a multidisciplinary approach.

The main components for the SanDAL Strategic Research programme are summarised in the concept figure below:

Strategic Research Programme

Vision: to achieve scientific excellence in MS&DS at the University of Luxembourg



Impact on UL and Luxembourg

The Strategic Research Programme of the SanDAL project will also measure their results with respect to the ERA Chair expected impacts:

- i) Increased attractiveness of the university and of Luxembourg, for internationally excellent and mobile researchers;
- ii) Increased research excellence of the University of Luxembourg in the field of MS&DS;
- iii) Improved capability to compete successfully for internationally competitive research funding;
- iv) Institutional changes within the University of Luxembourg to implement the European Research Area priorities.

One important aspect of the project is the alignment of the research programme with the national strategic research and development priorities, especially the SMART Specialisation Strategies of Luxembourg. These will demonstrate the core objectives to be performed and generate the considerable impacts expected from the project activities. Luxembourg recently undertook a [Strategic Study of the Third Industrial Revolution](#) inspired by the social and economic theories of Jeremy Rifkin. It aims to make Luxembourg's existing economic model more sustainable and interconnected for future generations by working with ICT, energy and transport as part of an intelligent network. The study involved 300 national and international socio-economic stakeholders and resulted in 9 priority measures presented in November 2016. In the context of this study, UL undertook a consultation exercise to pinpoint strategic innovation

niches where MS&DS would create a sizeable multidisciplinary impact. The following concrete examples were identified:

- Complex systems in biophysics (with UL’s [Luxembourg Centre for Systems Biomedicine](#) and UL’s [Physics and Materials Science Research Unit](#))
- Artificial Intelligence with data-driven application (with UL’s [Interdisciplinary Centre for Security, Reliability and Trust](#) and UL’s [Luxembourg Centre for Systems Biomedicine](#))
- Computational finance (with UL’s [Luxembourg School of Finance](#))

More broadly, the Third Industrial Revolution will transform the requirements posed to the Luxembourgish workforce. As automation replaced manual workers with machines during the first two industrial revolutions, the Third Industrial Revolution will see a number of high-qualification jobs superseded by the increasing power of artificial intelligence. This will require the invention of new socio-economic paradigms and already means that highly-qualified staff will need to continuously adapt their skills to the Luxembourgish market and re-educate themselves. This will require strong fundamental bases in STEM subjects, in particular mathematics, which MS&DS can provide, from early-age to on-the-job advanced training and education. Thus, we foresee that the ERA Chair in Mathematical Statistics and Data Science – with its postgraduate training programmes (MSc and PhD) – will have a vital role to play in helping to supply nimble data analytic scientists and engineers that Luxembourg will need.

Bridging the SRP with national and European Research Priorities

To ensure that the ERA Chair supports Luxembourg’s SMART Specialisation and European Research Priorities, the organisation of workshops, featuring regular trainings, will be held to bring awareness of Luxembourg’s SMART Specialisation Priorities as well as EU R&D calls for proposals. These workshops will help ensure the ERA Chair conducts applied research relevant to the needs of local innovative companies. For example, the ERA Chair Team will work closely with several of Luxembourg’s industry clusters - involving public and private sector organisations – namely the [Luxembourg Logistics Cluster](#), [Luxembourg BioHealth Cluster](#), [Luxembourg ICT Cluster Luxembourg](#) and [Materials and Manufacturing Cluster](#). They will also stimulate research funding from public and private sources. More information regarding the strategic plan for funding is illustrated in “National and European (EU) Funding Opportunities” of this document.

Strengths and areas for improvement

Strengths (S) and Weaknesses (W) of UL	Opportunities (O) and Threats (T) faced by UL from External Environment
<p>Strengths</p> <p>S1. Strong research in mathematics, with focus on probability theory, including applied probability (as well as some aspects of geometry)</p> <p>S2. Well-functioning teaching programs in mathematics at the bachelor and master level, and recent expansion of PhD program with creation of a Doctoral School. Provides good opportunity for finding strong students to start PhD.</p> <p>S3. Recent establishment of Computational Sciences priority by UL, with strong potential of expansion.</p> <p>S4. Dynamic academic environment with expanding and improving research in areas connected to Mathematical Statistics and Data Science: biology, medicine, economics, computer science, etc.</p> <p>S5. Economic environment with strong potential ties to data analysis in the finance and insurance areas, as well as “industry 3.0” sector (Amazon, future Google centre, logistics centre at Bettembourg/Dudelange, etc.)</p> <p>S6. Excellent infrastructure, access to HPC clusters</p>	<p>Opportunities</p> <p>O1. Recent substantial investment of the Government to support an HPC cluster</p> <p>O2. Potential for further research capacity building via European Structural Funds.</p> <p>O3. Expected increase in RDI public spending of 2,3%-2,6% of GDP by 2020 creating opportunities for a stronger investment in the S3 priority domains.</p> <p>O4. Strong national investment in the health sector, in discussion the creation of a Medical school.</p> <p>O5. Federation of several innovation fields around the common vision of Luxembourg's S3.</p> <p>O6. Efficient, cooperating and easily reachable support science and innovation institutions: FNR and Luxinnovation.</p> <p>O7. Attractiveness of Luxembourg to researchers (competitive salary level, multicultural, open environment).</p> <p>O8. Inclusion of Luxembourg in the Widening Participation Programme from H2020. A successful Luxembourgish proposal will represent an important opportunity to the science and innovation system in Luxembourg</p>

<p>S7. Experienced research support team for the Faculty (3,5 research facilitators and Quality and process officer).</p> <p>S8. Experienced EURAXESS and International Office to facilitate long/short personnel exchanges</p> <p>S9. Human Resources awarded “HR Excellence in Research” for its strategy to attract high-level researchers and ensure high-quality standards and environment for its researchers</p>	<p>O9. Potential to develop the competences on Mathematical Statistics and Data Science, through the intensification of multidisciplinary collaborations within the Faculty and with other Luxembourgish partners; but also internationally via established and new networks (e.g. IPCEI).</p>
<p>Weaknesses</p> <p>W1. UL is still a young university - lack of international reputation and visibility.</p> <p>W2. Current lack of critical mass in Mathematical Statistics and Data Science to be competitive at international level (both on the academic side and on the private sector).</p> <p>W3. Lack of experience in applied statistics or “real-world problems”</p> <p>W4. Elimination of national funding for the attraction of highly qualified postdoc researchers.</p> <p>W5. Insufficient knowledge and experience about innovation management techniques amongst research staff</p>	<p>Threats</p> <p>T1. National budgetary cuts for the university constitute a threat to current research directions.</p> <p>T2. Insufficient funding line for multidisciplinary research (such as DFG Sonderforschungsbereich) so difficult to compete with other research centres.</p> <p>T3. Threats to Luxembourg’s reputation as a financial centre (36% of national GDP comes from financial sector).</p> <p>T4. Lack of trained researchers and managers experienced in preparing and managing EU projects, leading to difficulties in fully exploiting opportunities in Horizon 2020.</p>

The Strategic Research Programme has been considered by a comprehensive overview of the University of Luxembourg’s current position and high potential within this ERA Chair opportunity. The SWOT analysis (above) details the status at the start of the ERA Chair, including existing resources which denote strong potential, but more importantly, it targets the areas where improvement is necessary to generate progress and ultimately, continued success and programme sustainability. The SanDAL project will work on strengthening its research excellence primarily and then it will be a contributor on different levels towards various desired sectors and create impact on a national, socio-economic scale.

SanDAL research priorities

The Strategic Research Programme has been established based on the research priorities of the SanDAL ERA Chair and the team’s recognised expertise, in which they will foster UL’s research excellence in MS&DS. In order to create robust teaching programmes in this field, the aim is to provide future generations of students with the mathematical and quantitative tools necessary to understand the new and emerging tools of Mathematical Statistics and Data Science, which will lead to their capacity to develop and apply them both in various academic fields and the increasing number of companies working in this challenging and quickly evolving area. It will foster UL’s research excellence in two research sub-topics: High-Dimensional Data Analysis and New Mathematical Tools for Contemporary Statistics. These two scientific areas have been carefully selected to demonstrate the increasing relevance of the field and to highlight their applicability in both academic and industry settings.

Main research areas

In recent years, Data Science has become an area of intense research. It provides mathematical tools that have turned out to be crucial for data analysis in medicine, biology, economy, finance and industry. They not only allow to draw information from data sets but also to train computers for learning complex tasks. Those tools rely on the association of three main ingredients: a statistical model, a statistical procedure and an algorithm. The statistical model describes the nature of the observed data and includes the a priori or pieces of information we have on their probability distribution. The statistical procedure is the strategy that is used to infer information from the data in view of estimating, predicting or learning. This procedure needs to be translated into an algorithm in order to be coded and implemented on a computer.

The **research priorities** of the ERA-Chair Team are to design new statistical procedures that can cope with many different statistical models and result in estimation strategies that possess as many interesting properties as possible, among which optimality and robustness will be key. We shall now describe these properties.

Optimality means that the procedure is expected to achieve the best possible performance with respect to some quality criterion as long as the statistical model is correct. For the problem of estimating a parameter, this means that we wish to achieve the best possible accuracy with respect to the number of data we have at our disposal. This is an important issue in biology and medicine where the data usually lie in a high-dimensional space and the sample size, which corresponds to the number of patients or experiments, is of moderate size only.

We also wish to keep these good estimation properties even in the unfavourable situation where the statistical model is not exact but only approximate, which means that our statistical model only provides a good approximation of the true distribution of the data. This is what we call **robustness**. It also includes situations where the data have been partially contaminated or contain outliers. These two situations can also be seen as a slight misspecification of the statistical model. Robustness turns out to be crucial in machine learning when one usually presumes that the observed data are independent with the same distribution although in many practical situations these assumptions might be debatable.

In view of being more specific, we provide below an account of our research directions.

Bayesian Methods (Yannick BARAUD, ERA Chair holder)

Parametric and non-parametric Bayesian Statistics have produced very popular statistical tools for inference on a parameter of interest. These rely on the choice by the Statistician of a prior distribution on the set of parameters that can be interpreted as a choice of a statistical model for the data. Under suitable assumptions, the Bayes-posterior distribution is proven to concentrate around the parameter of interest with an accuracy that is optimal. Unfortunately, this result only holds when the prior puts enough mass around the target parameter. When that fails to be the case, it is well known that Bayes procedure may perform very poorly, even in the favourable situation where the parameter is close to the support of the prior. In this Bayesian paradigm, a question naturally arises: is it possible to propose an alternative to the classical Bayes posterior that will concentrate around the target parameter not only when the prior put enough mass around it but also when it is close enough to a parameter for which this is the case?

High-Dimensional Data (Yannick BARAUD, ERA Chair holder, and Juntong CHEN, ERA Chair PhD candidate)

In many problems arising from medicine, biology and economy, one observes n data that lie in a linear space the dimension p of which is much larger than n . This situation is typical of the regression problem where one tries to predict a response variable y of interest on the basis of the observation of p real-valued predictors. In view of solving this problem, sparsity is commonly assumed which means, in the above regression model, that only a small number of predictors (smaller than n) actually plays a role in the prediction of the value of y , the other predictors being totally irrelevant. In the Gaussian linear problem, the celebrated Lasso procedure (and its extensions) provides a solution to this problem. However, it heavily relies on the assumption that the errors are Gaussian (or sub-Gaussian). It excludes heavy-tailed error distribution and provides no theoretical guarantee that the procedure would perform well in the presence of contamination or outliers. The situation is even worse in the context of logistic regression, a very popular model that is used when the response variable y is binary, i.e. only takes the values 0 or 1. It is well known that even if the set of influential predictors were known, the popular and celebrated maximum likelihood would not necessarily exist. For such high-dimensional problems, can we propose alternative procedures that could remedy these problems and furthermore address robustness issues?

Hidden Markov Models (Alexandre Lecestre, ERA Chair PhD candidate)

Hidden Markov Models are graphical models that are used to predict change points in the distribution of data. They are commonly used in genetics for the analysis of DNA sequences, in weather forecasting and in speech recognition, among other examples. In such a model, the data are assumed to be generated sequentially and independently according to the same distribution as long as some hidden variable (that is not observed) remains in the same state. When this random variable jumps into another state, this distribution changes. The distribution of the hidden random

variable is assumed to follow a Markov chain on a finite set and the problem is to estimate its transition probabilities and the distributions generating the data which are associated to each state of the Markov chain. This problem is usually solved by using the maximum likelihood estimator (when the number of states is known). Unfortunately, this approach leads to a non-robust estimator and a natural problem would be to propose a robust surrogate.

Robust Tests (Yannick BARAUD, ERA Chair holder)

Consider the general problem of estimating the distribution P of an observed random variable X with values in E and let M be a model for P , that is, a family of candidate probabilities that either contain P , or at least provides a good approximation of P with respect to some distance d . The construction of a robust estimation procedure with respect to the distance d means that we want to find a mapping from E to M that allows to associate to our observation X a point in M which is as close as possible to P (for our distance d) not only when P belongs to M but also when it is close enough to it. When M consists of two distinct elements, this problem boils down to finding a *robust test*. The problem is solved when d is the Hellinger distance but remains open for many other distances. A natural question is how one can extend the results obtained for the Hellinger distance to other distances.

As we see in these examples, very challenging problems in the area of Mathematical Statistics are currently solved by means of statistical procedures that suffer from a lack of robustness. This is precisely what we would like to remedy.

One of our objectives is to translate the procedures we propose into a code that could be used by **non-statisticians** for their research, especially for the problem of robust classification and estimation in the logistic model. We aim at designing **an R-package that would be freely available online**. To do so, we need to **hire an engineer** that could code and test the procedure. We plan to apply to the national **OPEN and CORE FNR Programmes** in view of supporting this project; the former will have a more fundamental focus, while the latter will include collaborations with experts in fields of application (e.g. biomedicine, social sciences, economy, etc.).

Establishment of the postgraduate training programme

The SanDAL project will establish a strong post-graduate training programme for PhD candidates and MSc students. The ERA Chair will develop PhD training in MS&DS within UL's Doctoral School in Science and Engineering. Currently, the main specialities of the Doctoral School include Mathematics and Applications, Physics and Materials, and Systems and Molecular Biomedicine. With the support of UL's Mathematics Research Unit, a specialisation and critical mass of PhD candidates will be established in MS&DS. The curricula of the taught courses and scientific objectives of the PhD projects will take into account the ERA Chair Holder's Strategic Research Programme.

The main objectives of the postgraduate training programme are the following:

- Propose to PhD and Postdoctoral Students courses that are introductory to Statistics. The course will give an account of statistical modelling as well as the main statistical procedures for estimation and testing (procedures based on the likelihood, on the empirical measure, and so on). We shall describe the properties of these tools and explain how they should be used, and their results interpreted. The aim of this course is to provide enough statistical background to non-specialist of Statistics for tackling commonly met statistical problems.
- Propose advanced courses of Statistics the aims of which are to describe the mathematical tools that lie at the foundation of the discipline (concentration inequalities, VC-complexity, metric dimension, etc.) and how those tools are used for statistical learning and inference in high-dimension.
- Encourage and promote interdisciplinarity and intersectorality by offering internships between the Mathematics Research Unit and research units in other disciplines, as well as with potential private sector partners.

The SanDAL ERA Chair will provide courses in Mathematical Statistics within the Master in Mathematics and contribute to an educational collaboration between mathematicians and computer/computational scientists for the training of

students in Data Science. This collaboration will also include professionals from the industry in view of meeting the market needs and to increase employability. In addition, the postgraduate training programme also includes the participation of the PhD students to reading courses, seminars and International Conferences, giving them the opportunity to present their work to an international audience, to receive valuable feedback, and to become aware of and stay up-to-date on the recent developments in the field.

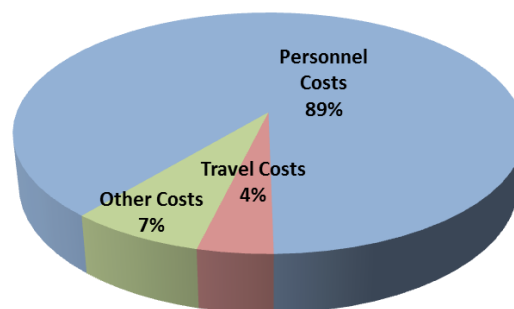
In view of attracting students, we aim to: organize and participate in meetings on Data Science targeting potential students and presenting UL's progress and achievements in MS&DS; set up **exchange programmes** with other universities; apply to **European Erasmus Programmes, individual fellowships**, and for national **FNR PRIDE** funding (PhD block-grants for consortia of at least 8 PIs).

Allocation of resources to research priorities

The current status of the SanDAL project is largely in line with the expectations highlighted in the Grant Agreement. One small deviation is as follows: at the arrival of the ERA Chair holder (M8), two excellent PhD candidates were immediately hired (starting in M9) and an announcement for post-doctoral researchers was published online. This represents a slight delay the hiring of the experienced researchers (first one expected to arrive no later than 01/09/2020) but significantly accelerates the PhD training axis of the project in the above-mentioned SanDAL research priorities, effectively launching both directions (High-Dimensional Data Analysis and New Mathematical Tools for Contemporary Statistics; see details above).

In the initial version of this document, the budget remains to still be primarily allocated to personnel costs: the ERA Chair holder and (at least) two experienced researchers will be paid by the SanDAL budget. In continuing the formation of the ERA Chair research team, the SanDAL project foresees that it will continue to follow the budget distribution upon the recruitment of new members in the Mathematics Research Unit.

The early forecast depicting a general division breakdown is expected to remain as in the proposal:



Any redistributions during the progression of SanDAL will be considered in the framework of the SRP and accounted for in newer versions of this document.

Guidelines for strategic publications

The scientific community will play a crucial role in demonstrating the scientific achievements of the SanDAL project at the University of Luxembourg. Over the course of the ERA Chair, the research team and the faculty overall, will leverage the research excellence of the institution, the Mathematics Research Unit and the ERA Chair Team by producing high-quality research and submitting publications to the best journals of the field. These publications will foster the training and learning component of the project in the field of MS&DS.

During the project, the ERA Chair Team and collaborators aim to publish 16+ papers in peer-reviewed journals. UL is extremely active in promoting and disseminating their research results and SanDAL will continue to demonstrate its work in this manner. In fact, scientific publications enhance the exploitation potential of the project results by being presented at well-known conferences. As a result, the team will also attend and present their work at international conferences, which will further initiate collaborations and attract new talented researchers to the University of Luxembourg.

The SanDAL project aims to submit their work to a variety of international, peer-reviewed journals (in green open access at least, and gold/platinum open access where available) relevant to the field of MS&DS, including:

- Annals of Statistics
- Journal of Statistical Planning and Inference
- Journal of the Royal Statistical Society: Series B
- Probability Theory and Related Fields
- Annales de l'Institut Henri Poincaré
- Bernoulli Journal
- Journal of the American Statistical Association
- Annals of Applied Probability
- Statistical Methodology
- Electronic Journal of Statistics
- Mathematical Statistics and Learning

National and European (EU) Funding Opportunities

The SanDAL project will enhance the UL's scientific network and capabilities in Mathematical Statistics and Data Science by exploring European and national funding programmes throughout the duration of the ERA Chair. In addition to promoting the SanDAL project's achievements in MS&DS, the ERA Chair will also generate a contribution in research to social, cultural and economic development. Elevating their research excellence in MS&DS will further facilitate their participation in R&D funding schemes. As a result, exploring and potentially applying for these opportunities will ensure the sustainability of SanDAL during and after the project.

Thus, the SanDAL strategic research programme has identified four (4) specific objectives for preparing competitive proposals:

- To create new or widen networks for future collaborations in data science
- To raise the profile of the University of Luxembourg in the field of MS&DS and to create new research opportunities for the institution, especially for the Mathematics Research Unit
- To disseminate and exploit the ongoing achievements of the SanDAL ERA Chair
- To support research and innovation on a European level

During the initial stage of the project, the Strategic Research Programme has envisioned that once established, the ERA Chair Team will be active in applying for funding to a range of EU R&D programmes. These include relevant priorities in Horizon 2020 (Societal Challenge "Health", Societal Challenge "Smart, green and integrated transport", LEIT "Information and Communication Technologies", and MSCA "Innovative Training Networks").

Via the project's workshops with local stakeholders, the ERA Chair Team will have opportunities to provide "contract research services" relating to MS&DS to private organisations. A non-exhaustive list of such private organisations includes: Cargolux and CFL Multimodal (logistics); ITTM and Laboratoires Réunis (healthcare); Amazon and ExaMotive (ICT); Dupont and Goodyear (manufacturing); AXA and LALUX Assurance (finance).

FNR Luxembourg has regular calls for funding for individual PhD candidates (AFR), doctoral training block grants of 8+ PhD candidates (PRIDE), postdoc and junior researchers (ATTRACT), thematic research programmes (CORE) and international cooperation (INTER). Mathematics currently has two ongoing PRIDE doctoral training units and is working on a follow-up proposal for the 2021 call. The ERA Chair Holder will be included in future PRIDE applications from mathematics, which will result in one additional PhD position, or even two. Furthermore, given the potential for application of the ERA Chair holder's expertise in other fields (e.g. biomedicine, social sciences, computational engineering, etc.), he will explore the possibility of joining other PRIDE consortia.

Luxinnovation assists in applying for Ministry of Economy subsidies (typically between 20-40%) for innovative companies in Luxembourg to help cover the costs of their research, development and innovation activities. The subsidy is increased by 10% if the company undertakes the work with the support of a public research body such as UL.

It is also expected that the ERA Chair Team will produce high-level researchers able to apply for an ERC grant, providing funding for the research directions after the end of the SanDAL project. However, other funding sources could be also found within the following strategic four-year plan of the University of Luxembourg (2022-2025). Once the work programmes for Horizon Europe are finalised, the ERA Chair team will continue to investigate eligibility for submitting applications.

Lastly, the students in the Mathematics Research Unit will also be encouraged to attend training and info days for increasing their knowledge of EU funding opportunities and be given advice for improving their skills in proposal preparation and appropriate funding schemes.

To align itself with Luxembourg SMART Specialisation and European Research Priorities, several performance indicators with respect to national and European funding opportunities have been established for the project:

- Workshops will be attended and/or organized to prepare EU R&D proposals: 6+
- Proposals submitted for EU funding (e.g. H2020): 6+
- Proposals submitted to European Research Council: 1+
- Number of international projects on a European level involving the ERA Chair: 2+

Dissemination and Exploitation Potential

The dissemination and exploitation measures which will be performed during the SanDAL project are essential to strengthening the University of Luxembourg’s research expertise in Mathematical Statistics and Data Science.

The potential dissemination and exploitation objectives are:

- To raise public awareness about the project, its expected results and progress within defined target groups using effective communication means and tools;
- To exchange experience with other groups and projects working in the field of MS&DS and its related scientific domains for multidisciplinary training and learning;
- To pave the way for successful exploitation of the project outcomes, through attracting new talent, increasing quality of postgraduate studies and research excellence in MS&DS

Through its set of activities, the overall aim is to widen their network and strengthen collaboration during and post ERA-Chair and to enhance the visibility of their research achievements to members of the scientific community, industry/stakeholders and public/societal actors.

The SanDAL strategic research programme will focus on attracting the following target audiences:

1. Academic, including researchers (PhD and Postdocs), students (master level), scientific staff
2. Public and governmental bodies
3. Industry, including start-ups
4. International organisations
5. General public

In general, the Strategic Research Programme encompasses the University of Luxembourg’s plan for managing the SanDAL Dissemination and Exploitation activities during the project:

Project Result	Activity	Target Audience	Indicator and Target
----------------	----------	-----------------	----------------------

Project leaflet and poster	Distribute during international conferences (e.g. ICIP), training workshops and outreach events	Scientists, engineers and general public	400+ leaflets distributed, 30+ posters displayed
Press conferences	Press conferences at the start, midterm and end of the project	Regional and national news media (radio, TV and print newspapers)	3 conferences, 8+ journalists
Project news	Publish project news releases and distribute through broader scientific news channels e.g. Cordis wire and Alpha Galileo	Scientists, engineers and general public	3+ press releases
Project website	Publish project summary, regular news and event updates on website	Scientists, engineers and general public	4000+ visitors
Short project film	Publish project film on Youtube and project website.	Scientists, engineers and general public	1000+ hits
Training workshops	Training workshops on science and innovation issues concerning MS&DS as well as preparing grant applications	Scientists and engineers	8+ workshops, 150+ participants
Summer and/or Winter Schools	Summer and/or Winter schools on science and innovation issues concerning MS&DS	PhD and MSc students	4 schools, 80+ participants
International conferences at UL	International conferences covering MS&DS and hosted at UL	Scientists and engineers	International conferences in Years 3 and 5 (e.g. conference series run by Institute of Mathematical Statistics), 200+ participants
Open days	Tours of UL's facilities. Also, scientific demonstrations and engaging lectures	General public and young students	1 Open day / Year, 400+ visitors
School visits	ERA Chair Team sends "school ambassadors" to local schools to promote higher education, research and innovation	Secondary school students	1-2 schools/year
Science festivals	young researchers to support a major science festival organised by the Luxembourg National Research Fund (FNR)	General public and young students	1+/year
Scientific results concerning MS&DS	Publish results in international peer-reviewed journals (e.g. International Journal of Mathematics)	Scientific research community	16+ journal papers
	Present results at international scientific conferences (e.g. AISTATS)	Scientific research community	16+ conferences

Conclusions and recommendations

This is the initial version of the Strategic Research Programme. It provides a comprehensive overview of the planning stages and envisioned actions concerning the ERA Chair SanDAL project in this early stage of the project. These include details pertaining to the MS&DS research priorities, allocation of resources, strengths and areas for improvement, and the specific objectives developed for: i) Postgraduate training of PhD and MSc students; ii) Guidelines for high-level publications; iii) National and European Funding Opportunities and iv) The dissemination and exploitation plan. Over

the course of the project, this document will be regularly reviewed and continually updated as necessary (M28 and M46) based on the progress of SanDAL's research objectives and project status.

This initial document will serve as a basis to discuss with SanDAL's National Advisory Board (NAB) and Scientific Steering Committee (SSC), whose feedback will be instrumental to the success of the project and will be included in the updates to the Strategic Research Programme.