Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

BUILDING GLOBAL RESEARCH CAPACITY IN PUBLIC HEALTH: THE CASE OF A SCIENCE GATEWAY FOR PHYSICAL ACTIVITY LIFELONG MODELLING AND SIMULATION

Anastasia Anagnostou
Simon J. E. Taylor
Derek Groen
Diana Suleimenova

Modelling & Simulation Group
Department of Computer Science
Brunel University London
Uxbridge, UB8 3PH, UK

Nana Anokye

Health Economics Research Group
Department of Clinical Sciences
Brunel University London
Uxbridge, UB8 3PH, UK

Riccardo Bruno
Roberto Barbera

Department of Physics and Astronomy
University of Catania
Piazza Università, 2
Catania, 95131, ITALY

ABSTRACT

Physical inactivity is a major risk factor for non-communicable disease and has a negative impact on quality of life in both high and low- and middle- income countries (LMICs). Increasing levels of physical activity is recognized as a strategic pathway to achieving the UN's 2030 Sustainable Development Goals. Research can support policy makers in evaluating strategies for achieving this goal. Barriers limit the capacity of researchers in LMICs. We discuss how global research capacity might be developed in public health by supporting collaboration via Open Science approaches and technologies such as Science Gateways and Open Access Repositories. The paper reports on how we are contributing to research capacity building in Ghana using a Science Gateway for our PALMS (Physical Activity Lifelong Modelling & Simulation) agent-based micro-simulation that we developed in the UK, and how we use an Open Access Repository to share the outputs of the research.

1 INTRODUCTION

Physical inactivity is a global public health problem. According to the World Health Organization (WHO) (WHO 2019), worldwide there are approximately 3.2 million deaths each year attributed to insufficient physical activity. Moreover, physical inactivity is recognized as one of the major factors for increasing the risk of dying from non-communicable diseases (NCDs). NCDs have been characterized as a "global epidemic" affecting both high and low- and middle- income countries (LMICs), with the latter carrying the largest burden (NCD Alliance 2019). In 2018, NCDs were the cause for 41 million deaths or 71% of all deaths worldwide, 15 million of which were premature (between the ages of 30 and 69 years). Of these premature deaths, 85% occurred in LMICs (WHO 2018a, WHO 2018b). Reducing premature deaths from NCDs by one-third by 2030 is one of the UN's 2030 Sustainable Development Goals and increasing levels of physical activity is recognized as a strategic pathway in achieving these goals (UN 2019). It is therefore imperative that governments take immediate action supporting increase of physical activity.

Research can support policy makers by providing evidence and tools for better-informed decisions. However, there are still many barriers preventing researchers in LMICs from reaching their full potential and capacity. For example, Chan et al. (2005) pointed out the inequality in producing research outputs where only ten per cent of global research in healthcare is undertaken in LMICs. Efforts to build (or strengthen) research capacity globally focus on improving researchers' skills, their access to research information and resources, and supporting researchers in playing a more regular and effective role in policy-making (DFID 2010). WHO (2014) suggests seven principles for good practice in research capacity building, one of which is to enable networking, collaboration, communication and sharing of experiences. Open Science principles closely align to these needs as they strive to make the outputs of research more visible and accessible (FOSTER 2019). Our Energising Scientific Endeavour through Science Gateways and e-Infrastructures in Africa (Sci-GaIA) (www.sci-gaia.eu) project worked towards methods and processes for Open Science in Africa. One of the outputs of the project was the recommendation that African Open Science be supported by the use of Science Gateways (web-based systems facilitating access to scientific applications) and Open Access Repositories through common e-Infrastructure services. These recommendations are being adopted in Africa (for example by the West and Central Research and Education Network (WACREN) that represents eleven countries in the region).

How might we contribute to the growth of physical inactivity research in LMICs? In the UK, we have conducted a major study to develop evidence-based techniques that can help decision makers assess the effectiveness of physical activity interventions. This brought together health assessment audits across a range of health conditions influenced by physical activity and resulted in the PALMS (Physical Activity Lifelong Modelling & Simulation) agent-based micro-simulation. PALMS is used to assess the cost-effectiveness of physical activity interventions by predicting the lifelong physical activity behavior of a population over time. PALMS can therefore provide an evidence base to assist policy makers in evaluating physical activity intervention strategies. It may be that PALMS could be an effective tool in assisting decision makers in LMICs. However, PALMS can be difficult to use by healthcare researchers. The question is how to make PALMS easily accessible to LMIC researchers and contribute to strengthening local research capacity?

In this context, this paper presents our experiences in developing a Science Gateway for PALMS that is being used by scientists and policy makers in Ghana. The paper is structured as follows. In the next section we give a brief overview of Open Science. We then describe the research problem and our Physical Activity Lifelong Modelling & Simulation (PALMS) agent-based micro-simulation. We then discuss how we used our Open Science approaches to create a Science Gateway for PALMS. We conclude the paper with reflections on our experiences and outline our future work.

2 OPEN SCIENCE

Open Science can be defined as "… the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods." (www.fosteropenscience.eu). There is a wide range of related topics: open access, open research data, open research protocols and notebooks, open access to research materials, open source software, citizen science, open peer review, open collaboration, etc. Digital technologies and infrastructures such as open access repositories, science gateways and e-Infrastructures (cyberinfrastructures) play a major role in realizing Open Science. Many key benefits have been cited (OECD 2015). These include more efficient research practices by reducing duplication and improving reuse, better transparency by more effective communication of scientific processes and methods, improving knowledge transfer from research to innovation as well as knowledge spillovers to citizens, and more effective coordination of action on global challenges and sustainable development goals.

At the heart of scientific progress is the published paper. A high quality scientific paper is typically published in a high quality journal, usually behind some kind of paywall. To make these works open access, publishers typically provide a Gold open access policy that makes an article free to download on the basis that the author pays some Article Publishing Charge (APC). However, there is an alternative. As part of Open Science and Open Access agreements, when a paper gets accepted authors deposit the last version of the work on acceptance typically in some Open Access Repository. This might be the final article, a post-print or a pre-print depending on the agreement with a publisher. The self-archived article might also be subject to an embargo period set by the journal publishers. Around 80% of publishers allow self-archiving (www.sherpa.ac.uk/romeo/index.php). This Green open access policy allows far great access at no cost to the author. However, this in turn requires access to some kind of Open Access Repository.

If we assume that open access is possible to the majority of scientific journal publications on this basis, then the journal paper becomes the “front end” to the outputs of the research discussed in the paper. The following approach to “capturing” research outputs might then be used (Taylor et al. 2017).

- Adopt good Open Data and Reusability practices that encourage independent verification and/or standardized reporting checklists.
- Consider making your data, results, software, etc. openly accessible (and trackable) by submitting works to Open Access Repositories that support the use of Digital Object Identifiers (DOIs).
- Use Creative Commons licenses to specify how work should be shared and used.
- Use a Researcher Registry such as ORCID to uniquely identify a scientist and link this to the scientist’s works via DOIs.
- Ensure that both DOIs and ORCIDs are used when publishing or in social media to correctly identify the scientist and the scientist’s works so that these can be tracked through scientometrics and altmetrics.
- Consider deploying simulations via a Science Gateway or similar portal-based approach to enable the widest possible access to your work.

Examples of how these might be used can be seen in (Taylor et al. 2017). In the next section we describe PALMS.

3 PHYSICAL ACTIVITY LIFELONG MODELLING & SIMULATION

PALMS is an agent-based micro-simulation that predicts the lifelong physical activity behavior of individuals of a population and its effect on their quality of life. The simulation can be used to provide evidence of the cost-effectiveness of physical activity interventions. Using an individual’s characteristics it assesses how their physical activity and relative risk of various health conditions change over time. Health conditions currently include cardiovascular disease (CVD), diabetes, depression and musculoskeletal incidents (MSI). PALMS can also output information on the related costs of these conditions for a population.

3.1 PALMS Conceptualization and Design

Figure 1 shows the overall PALMS process. Each individual of the population has characteristics that have an impact on physical activity and health conditions risks. These characteristics can be invariants such as ethnicity, gender, Type 1 Diabetes, etc. or variants such as age, cholesterol ratio, body mass index (BMI), systolic blood pressure, etc. The initial population is sampled from the Health Survey for England (HSE 2012). We performed data cleansing to exclude missing data (e.g. records that did not include relevant characteristics data). Table 1 shows the profile of the UK population used in PALMS. Bootstraping (random sampling with replacement) can be used when larger populations are required. Life expectancy data comes from UK lifetables (UK ONS 2018). Apart from the baseline life expectancy, cardiovascular incidents fatality is modelled as well. Data on stroke and coronary heart disease (CHD) mortality is taken from Smolina et al. (2012) and the British Heart Foundation (BHF 2012) respectively.

Well established risk estimation algorithms are embedded in PALMS for CVD (QRISK2) (Hippisley-Cox et al. 2008) and Type 2 Diabetes (QDiabetes) (Hippisley-Cox et al. 2017) risks. Incidence data from literature is used to estimate the probabilities of depression and MSI (Rait et al. 2009; Wijlaars et al. 2012). The above, however, do not factor in the physical activity level of an individual and its effect on health conditions risks. In PALMS we estimate the relative risk for developing a condition based on the physical activity level of individuals and the associated costs and quality of life. These risk estimates are based on well-known studies. For CVD and Diabetes, for example, we use a Bayesian meta-analysis study conducted by Kyu et al. (2016) to adjust these risks.

Table 1: Input population selected characteristics.

Selected baseline characteristics of the initial population (n = 9,594)	
Female, n (%)	5,268 (54.9)
Age, (years) median (25th, 75th percentile)	45 (27, 63)
Ethnicity, n (%)	
– **White**	8,470 (88.3)
– **Black**	247 (2.5)
– **Asian**	642 (6.6)
– **Other**	235 (2.4)
Physical activity status, (min per week) median (25th, 75th percentile)	225 (30, 570)
BMI, median (25th, 75th percentile)	25.95 (22.21, 29.72)
CVD history, n (%)	1,033 (10.8)
Type 2 Diabetes, n (%)	527 (5.5)

Typically, the focus of longitudinal studies on physical activity is on specific population groups. To

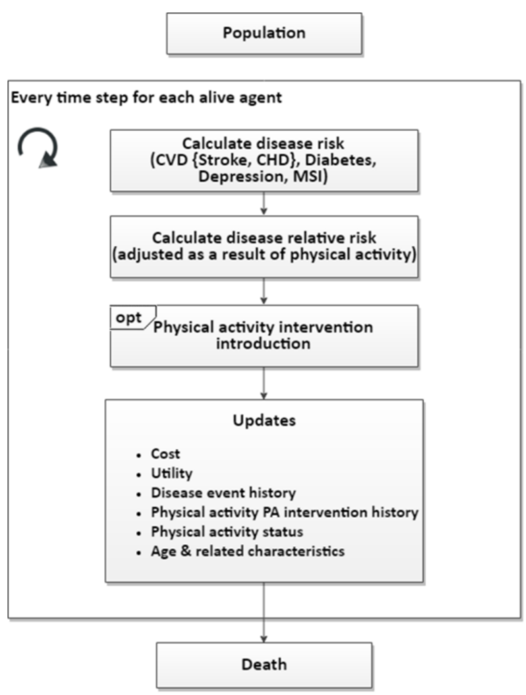

Figure 1: Overview of the PALMS process.

the best of our knowledge there is no longitudinal data for the whole lifetime of the general population. We therefore performed cross-sectional gender specific regression of physical activity on age and medical history, variables that potentially correlate to physical activity changes, on the initial population dataset. The formulae for calculating physical activity changes over time are derived from this exercise where the trajectory of lifelong physical activity is a function of previous activity levels. These are classified into three categories using minutes of vigorous physical activity (MVPA) per week as a measure: *Inactive* (0 ≤ physical activity status ≤ 85 min per week); *Moderately active* (86 ≤ physical activity status ≤ 425 min per week); and *Very active* (physical activity status ≥ 426). An individual's propensity to increase (or decrease) activity is determined by their previous activity level. Individuals can switch activity categories during their lifetime. The formulae used in the model are shown in Table 2.

Table 2: Physical activity changes formulae.

Gender	Activity level	Formula for physical activity status change
Male	**Inactive**	$-4.988 * CVD_{event} - 5.845 * Depression_{event} - 6.598 * MSI_{event} - 0.062 * Diabetes_{event} + 0.002 * Age - 0.0001 * Age_{square}$
	Moderately active	$7.566 * CVD_{event} - 39.901 * Depression_{event} - 10.582 * MSI_{event} - 1.606 * Diabetes_{event} - 0.122 * Age + 0.0001 * Age_{square}$
	Very active	$5.412 * CVD_{event} - 134.971 * Depression_{event} - 10.902 * MSI_{event} - 123.474 * Diabetes_{event} + 3.865 * Age - 0.010 * Age_{square}$
Female	**Inactive**	$-2.267 * CVD_{event} - 4.339 * Depression_{event} - 8.160 * MSI_{event} - 5.959 * Diabetes_{event} + 0.040 * Age - 0.0002 * Age_{square}$
	Moderately active	$-12.414 * CVD_{event} - 1.849 * Depression_{event} - 8.147 * MSI_{event} - 4.229 * Diabetes_{event} - 0.116 * Age + 0.0001 * Age_{square}$
	Very active	$-5.277 * CVD_{event} + 67.087 * Depression_{event} - 93.473 * MSI_{event} - 160.752 * Diabetes_{event} + 4.328 * Age - 0.012 * Age_{square}$

A complementary study was also performed to capture evidence of the success of stratified physical activity interventions. This information was merged with PALMS to allow policy decision makers to investigate the cost-effectiveness of new physical activity interventions on targeted groups in a population. One or more interventions can be introduced to the simulated population. Intervention-related data includes among others estimated impact on physical activity (usually increase), related cost (incurred to one or more public sectors e.g. health, transport, education, etc.) and uptake probability. PALMS simulation output data is recorded in CSV files and includes among others individual variant characteristics and their changes over time, incidents and related costs, physical activity status changes, interventions uptake and related costs, QALYs, etc. Cost-benefit analysis can then be conducted comparing the baseline simulation outputs with the introduced intervention(s) outputs.

Figure 2 illustrates the PALMS simulation and agent logic. At the start of a simulation run, input data sets are read from CSV files and various variables and objects are initialized. At this point, initial population is generated.

Agents represent individuals and each simulation time step accounts for three months of a person's life. Three months was selected as a convenient simulation time unit where the simulation can run efficiently without compromising accuracy of the simulation outputs. The user can choose whether to run the simulation for the lifetime of a cohort or for a shorter period of time.

For each alive agent in the cohort, baseline risks for having medical condition events in each time step are estimated. These risks are adjusted considering the physical activity status of the individuals. We then update several time and/or condition dependent variants. These include aging, physical activity changes, costs, quality of life and medical history. If interventions for promoting physical activity are

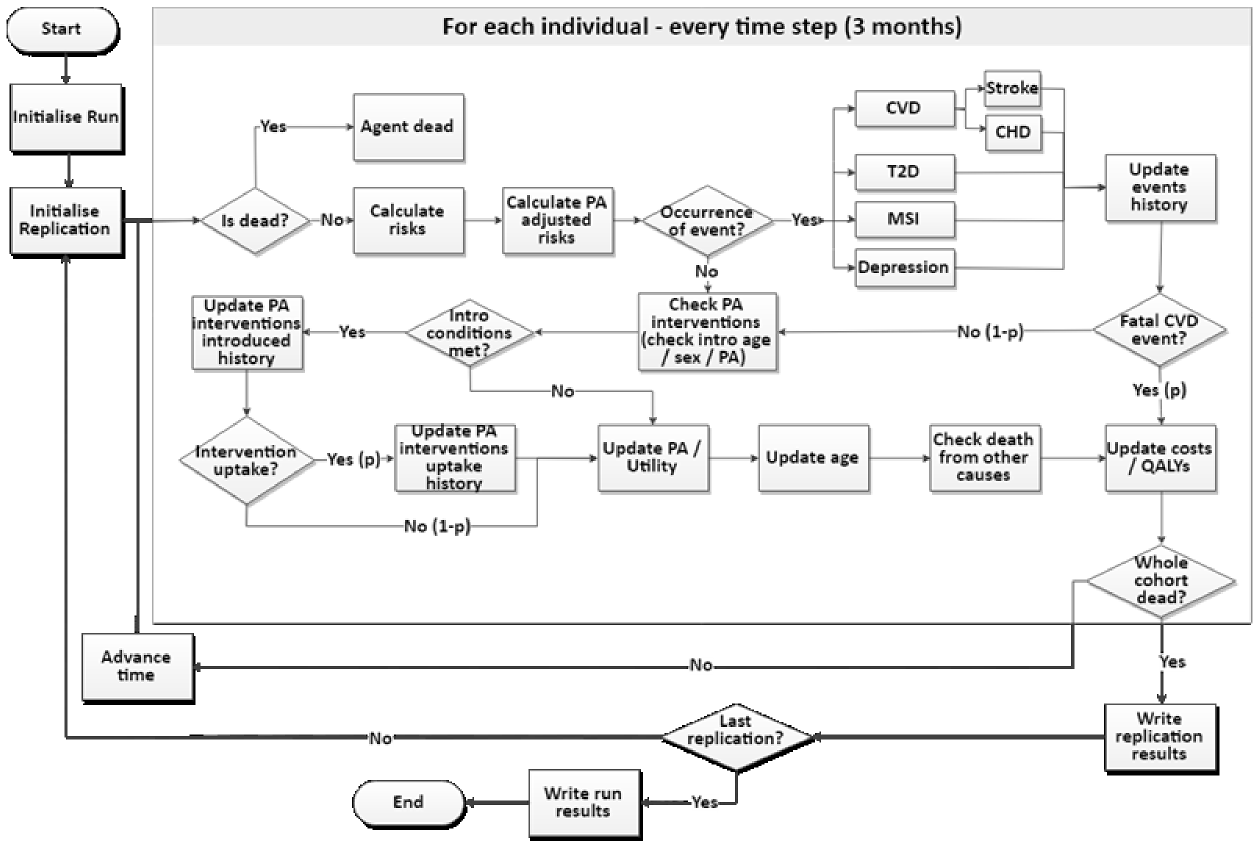

Figure 2: PALMS simulation and agent logic.

included in the simulation run, the eligibility of the individual is checked. Interventions usually are targeted to a population group, for example school children, in which case the age of the agents is an eligibility criterion. Uptake of an intervention changes – usually increases – the physical activity levels and quality of life.

3.2 PALMS Implementation

The PALMS micro-simulation is implemented in REPAST Simphony, a free and Open Source agent-based simulation toolkit (https://repast.github.io/repast_simphony.html). It is a java-based implementation with the respective advantages (e.g. portability, flexibility, etc.) and disadvantages (JVM memory requirements). Alongside the simulation engine, REPAST Simphony provides Graphical User Interface (GUI) tools that facilitate simulation parametrization and visualization. The PALMS GUI can be seen in Figure 3. In this snapshot we can see from left to right:

- Experiment parameters that users can change, for example users can select to run an experiment for a population in a specified age range,
- A grid visualization of the whole population and how the physical activity of the individuals change over time using a color code, for example green cells represent highly active persons, amber and red cells represent moderate and low physical activity, while black cells indicate that the individual is dead, and
- A time series graph plotting the aggregated physical activity trajectory for the whole population over time.

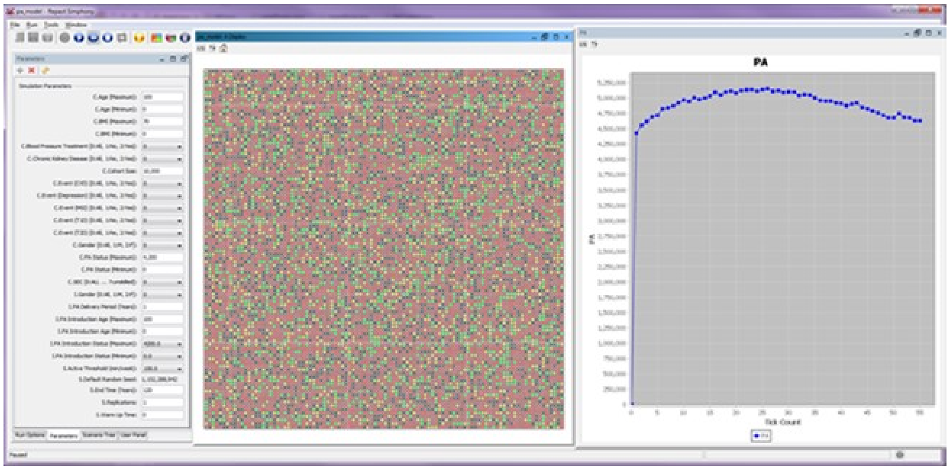

Figure 3: PALMS Graphical User Interface.

PALMS is developed in an extensible manner as it has a well-defined data model. This therefore allows new health conditions, other lifestyle choices affecting NCD and new interventions (e.g. healthy diet promotion) to be added to PALMS in a similar manner as described in the previous section.

3.3 Illustrative example

To show the rationale behind the generated formulae for physical activity progression over time, we ran PALMS for six population subgroups. These subgroups are gender-specific for inactive, moderately active and very active individuals. Table 3 shows the population characteristics of these runs (left column). The right column shows the simulation output for physical activity and CVD events per person-year. The rate is comparable with published literature (Wong et al. 2016; Timmis et al. 2018).

Table 3: Selected experiments' parameters & results.

Seed population characteristics					Simulation output		
Population group	**Cohort size (n)**	**Age (median, 25^{th}, 75^{th})**	**Average physical activity (min / w)**	**CVD history (n)**	**Average physical activity (min / w)**	**Total CVD events (n)**	**CVD events per person years**
M-Inactive	1,271	59 (42, 72)	16.68	290	70.52	1,181	0.04
F-Inactive	1,935	56 (35, 73)	18.20	336	54.28	1,411	0.025
M-Moderately	1,427	42 (25, 58)	237.48	140	294.88	1,358	0.026
F-Moderately	1,765	41 (25, 55)	226.98	111	231.27	1,260	0.018
M-Active	1,628	38 (18, 55)	1,106.34	96	647.96	1,539	0.024
F-Active	1,568	39 (22, 54)	991.29	60	475.89	1,113	0.017

The trajectories of physical activity level for the population subgroups for the lifetime of the cohort stratified by gender and physical activity are shown in Figure 4. In the figure we see the average physical activity level for the cohort over time. On average the trend of the active subgroup is to decrease their physical activity level later in life, while the moderately active and inactive subgroups tend to retain their level of activity.

We now discuss how PALMS was ported to a Science Gateway and how we deposited the experiments' artefacts in an open access repository following Open Science good practices.

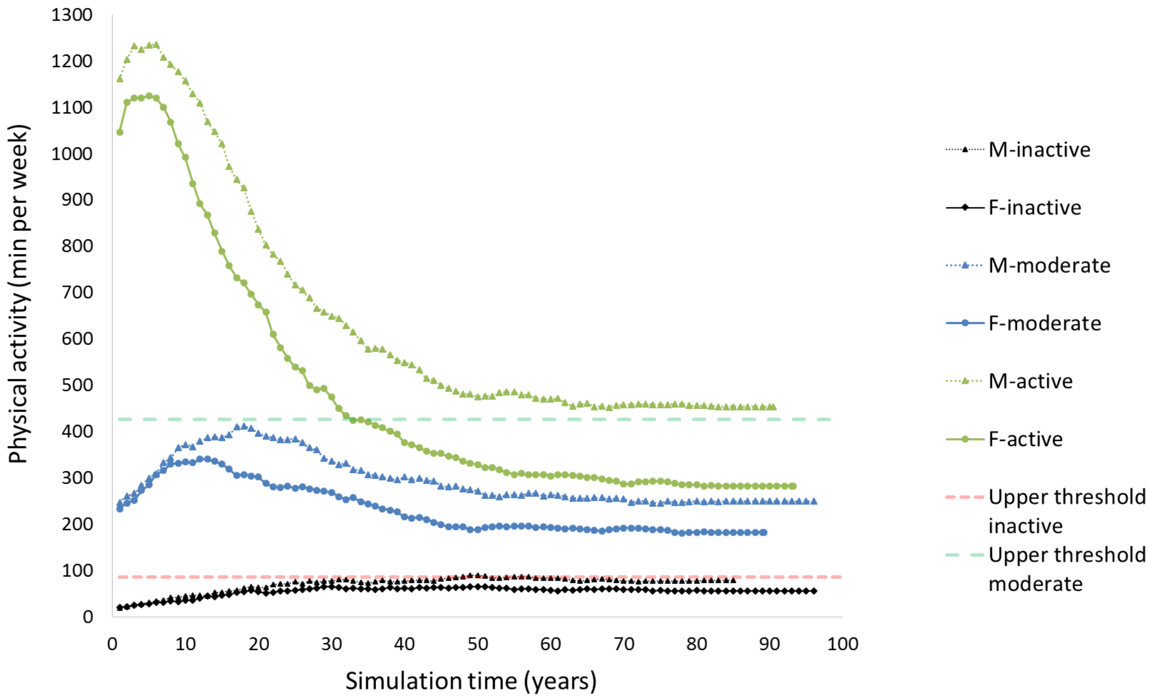

Figure 4: Physical activity trajectory.

4 THE PALMS SCIENCE GATEWAY

To make the PALMS agent-based micro-simulation openly accessible we used a combination of a Science Gateway and an Open Access Repository supported by an e-Infrastructure, as is the practice in numerous scientific communities. Many scientists worldwide use research data infrastructures (e-Infrastructures) to support their work and reinforce knowledge transfer within the community. Our H2020 Sci-GaIA project created the Sci-GaIA Open Science Platform as a way to provide standardized support for Open Science. The Platform consists of a range of technologies that facilitate the development of a range of scientific application services. Aspects of the Platform have been adopted in the development of National Research and Education Services in West and Central Africa.

Figure 5 shows the generic FutureGateway Framework architecture used to create the PALMS Science Gateway. In a high level overview, the FutureGateway Framework consists of three main components. The API Server front-end that accepts API calls, populate a queue table, manages authentication, and manages applications, infrastructures and tasks. The FutureGateway Database that keeps and maintains tasks, applications, infrastructures, task-queue, users, groups and roles. The API Server that polls over the queue table, extracts tasks to submit, checks status and consistency of submitted tasks, and retrieves available outputs. The FutureGateway Framework communicates with Executor Interfaces such as grid and cloud executors to launch jobs on Distributed Computing Infrastructures (DCIs) and with a GUI using a REST API (the DCIs are typically supported by an e-Infrastructure).

PALMS uses a containerized version of REPAST Simphony created by the COLA project (https://project-cola.eu/). When running headless, the parameters discussed in the previous Section are described in an XML file. This parameter XML file together with the compressed model files are taken as input for the containerized software. In the current implementation both input files are stored in a file repository.

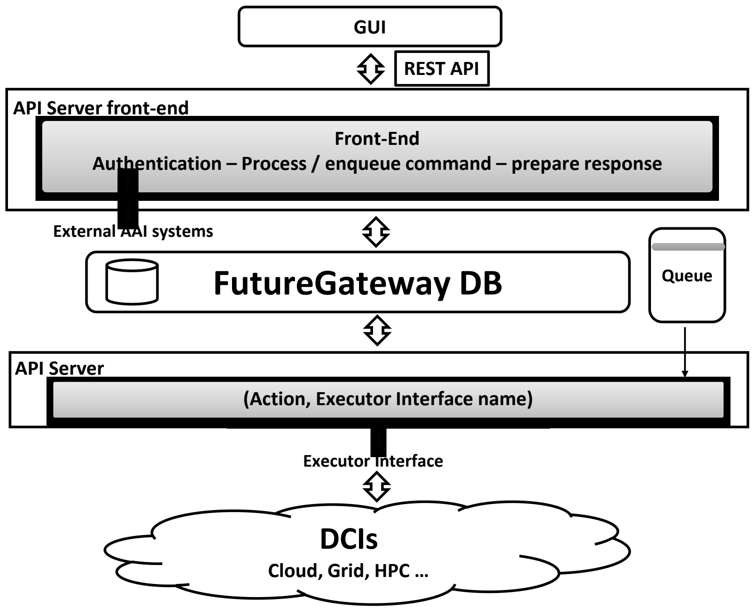

Figure 5: FutureGateway framework architecture.

The PALMS Science Gateway application front-end is shown in Figure 6. It was created to be as simple as possible to enable healthcare scientists to easily use it. The front-end has two fields where the URLs for the model and parameter file inputs are added. Authenticated users can access the portal and submit jobs. Once the simulation is submitted, the job progress can be monitored from the front-end. When the job is finished the output files are stored in a file repository.

The PALMS Science Gateway application is hosted in the EGI's Science Software on Demand site (https://fgsg.ct.infn.it/egissod/web/ssod).

Following the Open Science approach outlined earlier in the paper, we deposited all the artefacts of this work in the INFN Open Access Repository and assigned DOIs to each artefact and followed the STRESS guidelines (Taylor et al. 2018) for agent-based simulation (https://www.openaccessrepository.it/communities/palms-sim). The material and research outputs can be accessed as follows:

- Model https://doi.org/10.15161/oar.it/23467
- Input parameters: M-inactive https://doi.org/10.15161/oar.it/23469
- Input parameters: F-inactive https://doi.org/10.15161/oar.it/23471
- Input parameters: M-moderately https://doi.org/10.15161/oar.it/23473
- Input parameters: F-moderately https://doi.org/10.15161/oar.it/23475
- Input parameters: M-active https://doi.org/10.15161/oar.it/23477
- Input parameters: F-active https://doi.org/10.15161/oar.it/23479
- Results https://doi.org/10.15161/oar.it/23481
- STRESS guidelines https://doi.org/10.15161/oar.it/23485

5 CONCLUSIONS

In this paper we discussed our experiences of using Open Science technologies as enabler for research capacity building in public health. We presented Open Science and how an element of our Sci-GaIA Open Science Platform has been used to support joint research between the UK and Ghana in physical activity agent-based micro-simulation. We developed a Science Gateway for our PALMS model to enable

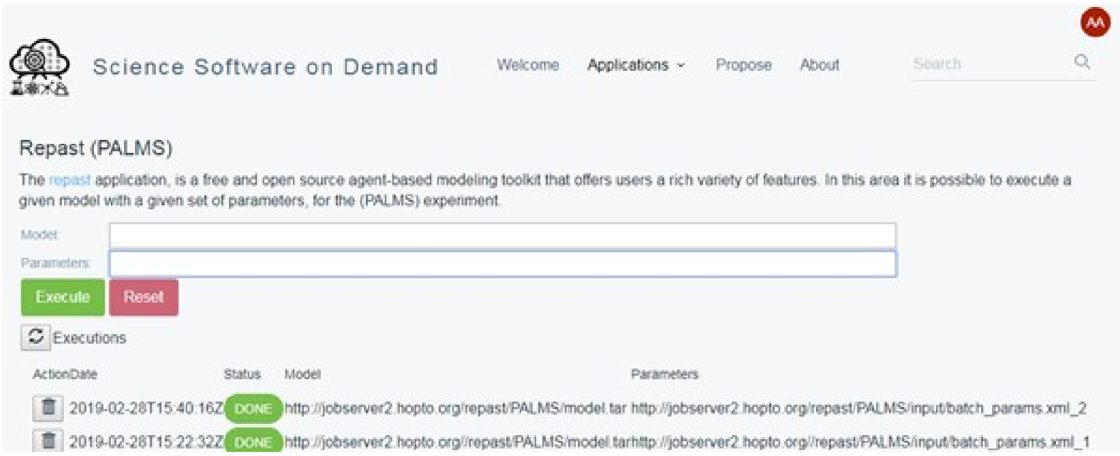

Figure 6: PALMS Science Gateway Application Front-End.

easy and wider access of our model. We then held two workshops in Ghana aiming to disseminate our approach to policy-makers and scientists. The first workshop was held at the Ghana Health Service (GHS) headquarters in Accra where the participants' background ranged from medical professionals to healthcare administrators. The second workshop was held at the Kwame Nkrumah University of Science and Technology in Kumasi where computer scientists and health economists participated in the deliberations. Both workshops generated vibrant discussions. It is very interesting to note that the presented technology served as a springboard to start thinking of local issues that can be studied using simulation. It is worth mentioning also that soon the conversation started to include a wide area of potential topics that need improvements in the local settings and how sharing of information can contribute to potential solutions. To continue the discussions, we will hold follow up workshops in order to address country-specific requirements and forge long-term collaborations.

The main lesson learnt from our experiences is that a very important factor for strengthening research capacity is to enable international research networks and sharing of information, which agrees with the wider views on the area. Often the offered tools cannot be applied directly to local contexts. In the case of policies for encouraging physical activity for example, social and cultural beliefs cannot be ignored. As Onywera (2017) highlights, in many developing countries roundness (or fatness) is often a symbol of wealth, good life and therefore deserves prestige. This view hinders the efforts from public authorities to fight against the negative health effects of physical inactivity. PALMS for example may not be appropriate to use without modifications to include local population data and variables however it can be a starting point for initiating discussions and potential adoption of similar technologies by the local experts. It is very important that local researchers start to take ownership of the research initiatives and develop research culture.

We hope that this example and discussion will facilitate the development of an Open Access Repository to support research in Ghana (in a similar way to that created in Ethiopia see https://nadreweb.ethernet.edu.et/) as well as further Science Gateways in the region. The PALMS Science Gateway follows the WACREN NREN Services Roadmap and is being deployed on the WACREN infrastructure. This will enable Ghanaian academics and policy makers to use the micro-simulation in their research and adapt it to their needs. This will form potentially the basis of guidelines development not only in physical activity but other lifestyle behaviors (e.g. nutrition) as well as facilitating capacity building in developing areas in Africa. Our approach also supports the general objectives of the One Health movement to "designing and implementing programmes, policies, legislation and research in which multiple sectors communicate and work together to achieve better public health outcomes" (WHO 2017) by enabling sharing of resources across sectors and continents.

Additionally, we hope that by creating this demonstration of how PALMS and the Sci-GaIA Open Science Platform can be used to support research and policy making in Ghana, we will show how our research can be applied in an African context and subsequently can impact policies in other African countries such as Nigeria, Kenya, and Ethiopia. To this end, we are currently creating other Science Gateway simulation applications and planning workshops in other African countries. For example, the next workshop is going to be held in Addis Ababa, Ethiopia to demonstrate the use of a Science Gateway of an agent-based simulation built to study forced migration (Suleimenova et al. 2017). Forced migration also affects public health and is of great interest to policy makers globally. This will be part of the Open Science Week workshop aiming to discuss Open Science issues, such as legal and ethical consideration, and train local researchers in applications and services such as Science Gateways and Open Access Repositories.

ACKNOWLEDGMENTS

The authors would like to thank Prof. Tamas Kiss, Mr Gregoire Gesmier, Mr James DesLauriers and Dr Osama, Abu Oun for their support through the EU H2020 COLA (Cloud Orchestration at the Level of Application) project (Project no. 731574).

REFERENCES

British Heart Foundation. 2012. "Coronary Heart Disease Statistics: A Compendium of Health Statistics" 2012 edition https://www.bhf.org.uk/informationsupport/publications/statistics/coronary-heart-disease-statistics-2012, accessed 14th September 2019

Chan, L., B. Kirsop, and S. Arunachalam. 2005. "Open Access Archiving: The Fast Track to Building Research Capacity in Developing Countries". *SciDevNet* https://www.scidev.net/global/publishing/feature/open-access-archiving-the-fast-track-to-building-r.html, accessed 20th April 2019

Creative Commons. 2019. http://creativecommons.org/, accessed 16th April 2019

DFID. 2010. Capacity Building in Research. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/187568/HTN_Capacity_Building_Final_21_06_10.pdf, accessed 18th April 2019

FOSTER. 2019. The Future of Open Science. http://www.fosteropenscience.eu/, accessed 16th April 2019

Hippisley-Cox, J. and C. Coupland. 2017. "Development and Validation of QDiabetes - 2018 Risk Prediction Algorithm to Estimate Future Risk of Type 2 Diabetes: Cohort Study". *BMJ* 2017;359:j5019.

Hippisley-Cox , J., C. Coupland, Y. Vinogradova, J. Robson, R. Minhas, A. Sheikh, and P. Brindle. 2008. "Predicting Cardiovascular Risk in England and Wales: Prospective Derivation and Validation of QRISK2". *BMJ* 2008;336:1475–1482.

HSE. 2012. Health Survey for England. https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/health-survey-for-england-2012, accessed 16th April 2019

OECD. 2015. "Making Open Science a Reality". OECD Science, Technology and Industry Policy Papers, No. 25, OECD Publishing, Paris. dx.doi.org/10.1787/23074957.

Onywera, V.O. 2017. "Physical Inactivity is Hurting the Health of People in Countries like Kenya. There's a Solution". http://theconversation.com/physical-inactivity-is-hurting-the-health-of-people-in-countries-like-kenya-theres-a-solution-73000, accessed 17th April 2019

Kyu, H.H., V.F. Bachman, L.T Alexander, J.E. Mumford, A. Afshin, K. Estep, J. Lennert-Veerman, K. Delwiche, M. L. Iannarone, M.L. Moyer, K. Cercy, T. Vos, C.J.L. Murray, and M.H. Forouzanfar. 2016. "Physical Activity and Risk of Breast Cancer, Colon Cancer, Diabetes, Ischemic Heart Disease, and Ischemic Stroke Events: Systematic Review and Dose-Response Meta-Analysis for the Global Burden of Disease Study 2013". *BMJ* 2016;354:i3857

NCD Alliance. 2019. The Global Epidemic. https://ncdalliance.org/the-global-epidemic, accessed 17th April 2019

Rait, G., K. Walters, M. Griffin, M. Buszewicz, I. Petersen, and I. Nazareth. 2009. "Recent Trends in the Incidence of Recorded Depression in Primary Care". *The British Journal of Psychiatry* 195(6):520–4.

Smolina, K., F.L. Wright, M. Rayner, and M.J. Goldacre. 2012. "Determinants of the Decline in Mortality from Acute Myocardial Infarction in England between 2002 and 2010: Linked National Database Study". *BMJ* 2012;344:d8059-d8059

Suleimenova, D., Bell, D., and Groen, D. 2017. "A Generalized Simulation Development Approach for Predicting Refugee Destinations". Scientific Reports, Vol. 7, Article number: 13377

Taylor, S.J.E., A. Anagnostou, C. Currie, T. Monks, S. Onggo, M. Kunc and S. Robinson. 2018. "Applying the STRESS Guidelines for Reproducibility in Modeling & Simulation: Application to a Disease Modeling Case Study". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A.A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson., 739-748. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Taylor, S.J.E., A. Anagnostou, A. Fabiyi, C. Currie, T. Monks, R. Barbera and B. Becker. 2017. "Open Science: Approaches and Benefits for Modeling & Simulation". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W.K.V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, 535-549. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Timmis, A., N. Townsend, C. Gale, R. Grobbee, N. Maniadakis, M. Flather, E. Wilkins, L. Wright, R. Vos, J. Bax, M. Blum, F. Pinto, and P. Vardas. 2018. "European Society of Cardiology: Cardiovascular Disease Statistics 2017". *European Heart Journal* 39(7):508–579.

UK Office of National Statistics. 2018. National Life Tables: UK. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables, accessed 16th April 2019

UN. 2019. Sustainable Development Goals. https://www.un.org/sustainabledevelopment/health/, accessed 18th April 2019

WHO. 2019. Global Strategy on Diet, Physical Activity and Health. https://www.who.int/dietphysicalactivity/factsheet_inactivity/en/, accessed 17th April 2019

WHO. 2018a. Noncommunicable Diseases. https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases, accessed 16th April 2019

WHO. 2018b. The Top 10 Causes of Death. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death, accessed 17th April 2019

WHO. 2017. One Health. https://www.who.int/features/qa/one-health/en/, accessed 29th April 2019

WHO. 2014. "Seven Principles for Strengthening Research Capacity in Low- and Middle-Income Countries: Simple Ideas in a Complex World". *ESSENCE on Health Research* https://www.who.int/tdr/publications/seven-principles/en/, accessed 18th April 2019

Wijlaars, L.P., I. Nazareth, and I. Petersen. 2012. "Trends in Depression and Antidepressant Prescribing in Children and Adolescents: a Cohort Study in The Health Improvement Network (THIN)". *PLOS One* 7(3):e33181.

Wong, N.D., Y. Zhao, R. Patel, C. Patao, S. Malik, A.G. Bertoni, A. Correa, A.R. Folsom, S. Kachroo, J. Mukherjee, H. Taylor, and E. Selvin. 2016. "Cardiovascular Risk Factor Targets and Cardiovascular Disease Event Risk in Diabetes: a Pooling Project of the Atherosclerosis Risk in Communities Study, Multi-ethnic Study of Atherosclerosis, and Jackson Heart Study". *Diabetes Care* 39:668-676.

AUTHOR BIOGRAPHIES

ANASTASIA ANAGNOSTOU is a Lecturer and the Deputy Director of the Modelling & Simulation group at the Department of Computer Science, Brunel University London. Her email address is anastasia.anagnostou@brunel.ac.uk.

NANA ANOKYE is a Senior Lecturer in Health Economics and the Director of the Division of Health Sciences in the Department of Clinical Sciences, Brunel University London. His email address is nana.anokye@brunel.ac.uk.

SIMON J. E. TAYLOR is Professor of Computing and the Director of the Modelling and Simulation Research Group in the Department of Computer Science, Brunel University London. His email address is simon.taylor@brunel.ac.uk.

DEREK GROEN is is a Lecturer in Simulation and Modelling at Brunel University London, and a visiting Lecturer at the Centre for Computational Science, University College London. His email address is derek.groen@brunel.ac.uk.

DIANA SULEIMENOVA is a Ph.D. researcher in the Department of Computer Science at Brunel University London. Her current research focuses on forced migration and refugee movements models. Her email address is diana.suleimenova@brunel.ac.uk.

RICCARDO BRUNO is a Technologist at the Institute of Nuclear Physics, dept. of Catania, Italy. His main research area is on Distributed computing. His email address is riccardo.bruno@ct.infn.it.

ROBERTO BARBERA is Professor of Experimental Physics of Fundamental Interactions at the Department of Physics and Astronomy of the Catania University, Italy. His email address is roberto.barbera@ct.infn.it.