Proposal to the North West Science Review

The Global Virtual Population Laboratory

Lead Institute: **The University of Liverpool**

North-West UK institutes

Department of Physics, University of Liverpool
Dr T.J.V. Bowcock*,Prof. P.S.L.Booth (Director of Research)
Department of Geography, University of Liverpool
Dr. P. Williamson
Liverpool School of Tropical Medicine
Dr. P. Garner, Dr. M.C.Thomson
North West Public Health Observatory, John Moores University
Prof. M. Bellis, Dr. P Cosgrove
North of England Institute for Health Informatics
L. Isaksen, Prof. C. Taylor, Prof. A. Rector
School of Computing and Mathematical Sciences, John Moores University
Prof. P.J.G. Lisboa
Department of Mathematical Sciences, University of Liverpool
Dr R.G.Bowers, Dr. D.Clancy, Dr. P.R.Williamson

in collaboration with

World Health Organization
Dr. D. Heymann (Executive Director, Communicable Diseases Programme),
Dr. T. Mertens (Director, WHO Tunis Centre), A. Burton (Vaccine Assessment and Monitoring)

International Collaborators

World Bank
Dr. Mead Over
Centers For Disease Control and Prevention
Dr. D. Stroup (Associate Director for Science, Epidemiology Program Office)
National Institutes of Health
Dr. M.Miller (Assistant Director for Research, Fogarty International Center)
Johns Hopkins University
Prof. D. Burke
Brigham Young University
Prof. L. Wolfson

* Contact person for Global Virtual Population Laboratory proposal:
tjvb@hep.ph.liv.ac.uk

1. Overview

1.1 We are proposing the creation of a new facility, based in the North West of England, as a prototype World Wide Centre for the modelling and prediction of health events (such as epidemics). This will be closely analogous to a Weather Prediction Centre but focused on the spread of disease. We are calling this the Virtual Population Laboratory (VPL). This project, which originated from lengthy discussions between the World Health Organization (WHO) and the University of Liverpool, now has the active support of many institutes from the NW and worldwide. We propose to link the VPL with WHO HQ at Geneva and the WHO Mediterranean Centre Tunis (WMC). The project will require interdisciplinary collaboration bringing together information technology developed for the physical sciences with epidemiology, demography, economics, geography, and statistics to address issues in public health policy.

VPL Project

NW & International

inter-disciplinary collaboration

1.2 The bid is for 16 million Pounds Stirling (12 million requested from the North West Science Review and 4 million from Merseyside European Initiatives) to bring this centre to the North West. It will draw together existing expertise including: High Performance Computing from University of Liverpool (the world's largest computer simulation facility is already operating in the Department of Physics), UK population modelling from the University of Liverpool Department of Geography, expertise on climatic and global effects from the Liverpool School of Tropical Medicine, and collaboration with the North West Public Health Observatory (John Moores University) and with the North of England Institute for Health Informatics (NoEIHI) Manchester. The project will exploit the computing and Internet (GRID[*]) technology being developed in the UK. Other collaborating institutes include:

added value

Information Technology

- World Health Organization
- World Bank
- National Institutes of Health (U.S)
- Centers for Disease Control and Prevention (U.S)

1.3 The VPL will provide a permanent resource for the consolidation, analysis, and dissemination of data and for modelling disease patterns and impacts to inform national and international public health policy. This is the first time such a facility will have been available for health policy analysis at the global level. The major activities of the VPL will be:

core activities innovation

- Collection and assembly of global data on health and its determinants.
- Integration and application of this data for building scenarios and evaluating interventions.
- Development of Information Technology in support of this activity.

1.4 The VPL is designed to become an integral part of the Global Health Grid being sponsored by WHO and currently under development. The availability of funding from the NW Science Review would allow the North West to achieve a flying start in this field of global activity, and we foresee the North West becoming the leading centre to emerge in this field in the next 20 years and possibly further – attracting co-operation and investment from international institutes and funding agencies.

long term potential

unique opportunity

[*] The Department of Physics is participating in the EU-GRID development project.

2. Introduction

2.1 Recent advances in IT technology, developed specifically for the next generation of particle physics experiments, have provided new opportunities for creating a global health laboratory for predicting patterns of disease and evaluating prevention and control efforts. Building on established (but smaller scale) modelling techniques used in many countries e.g. Sweden, Canada, and US[1], a global system can now be constructed to inform regional and global public health policies.

precedents

2.2 For years, epidemiologists, statisticians and health organizations have been monitoring, analysing[2] and modelling epidemics such as influenza, measles and HIV infection on a worldwide scale[3]. Predicting the progression of these pandemics has drawn on a variety of techniques. Of particular interest to us has been the pioneering work using micro-simulation[4]. This method assumes that, while the exact mix of characteristics of every individual is unknowable, their aggregate characteristics may be estimated from survey and census data; as may be the range of individual variation in these characteristics, and the consequent relative risks of disease. A large number of virtual populations can then be generated, each one a probable approximation to the unknown true population, and disease events and interventions can be simulated within each virtual population in accordance with known or projected risks. The consolidated results of multiple 'runs' of the model may then approximate to a likely range of future possible scenarios. In addition, this 'Monte Carlo' method is able to model the rapid movement by air and rail of individuals. In the 21st century, where transport networks are key conduits for the spread of disease the ability to accommodate this feature is vital. So far the scope and success of this technique has been restricted by a lack of data necessary to create realistic virtual populations, and by limitations in the processing power to model realistic levels of interaction.

modelling epidemics

virtual population studies

Monte-Carlo method

current limitations: data and processing

2.3 These restrictions are fast disappearing. Work at the University of Liverpool has developed methods of creating realistic virtual populations[5] using UK data. Increased data collection by WHO and national governments is beginning to provide the data to inform fully dynamic models. Finally, physicists working at Liverpool University have developed world-leading expertise in the construction of low-cost computer farms that enable the creation and analysis of large simulated samples of data[6], as part of an ongoing collaboration with the European Centre for Nuclear Research (CERN). Access to the data and analytical power is then made available (subject to authentication and security), world wide, through the Internet.

enabling technology transfer from physical sciences

2.4 The application of this technology to global health modelling has several very attractive features. First, the new low-cost computing power may be used to create virtual populations describing the possible characteristics of the world's 6 billion people. Births, ageing, changing relationships, travel; sickness and death can be modelled. Second, we can now build databases that are large enough to store and to make the simulated populations realistic. Third, modern networking allows these databases to be geographically dispersed, making it easier for each region (whether it be a city, state, country or continent) to maintain and manage its own information resources. Finally the network connections between the user, the processing power and the databases form a "Grid[7]" which will allow worldwide access to the data and its analysis from anyone's desk or portable device.

low cost processing

networking

information "Grid"

2.5 Let us imagine a hypothetical example. It is 2010. The last case of polio has occurred, the virus has been certified as eradicated, and countries have stopped immunizing. A case of polio - wild virus appears. How far will it travel and how quickly? How sensitive is the spread to the location of the initial case? Should countries vaccinate? Which countries should vaccinate? What contingency plans, and what level of surveillance should be implemented to improve the chances of detecting a case? Should we stockpile vaccine? If so, where and how many doses? Will the virus spread far enough before detection to make vaccination an expensive and dangerous irrelevancy? The answers to these questions will depend on the location of the outbreak and the structure and dynamics of demographic, social, economic, ecological, and political circumstances. Similar questions can be asked for other diseases.

example: Post-Polio Eradication

2.6 The process of Monte Carlo modelling yields three classes of benefits. The first is results - quantitative estimates of possible outcomes based on the current state of knowledge. The second benefit, and one that does not depend on the completeness of the existing data, is a quantified evaluation of the very surveillance techniques that produced the data. We will analyze the critical data that are an input to the models, and investigate the degree to which the model's predictive power is degraded by gaps in the knowledge. This

benefits of method

will help modify our techniques for collection of information (perhaps making them more cost effective). Third, the model lends itself readily to a process of successive refinement.

successive refinement

3. Virtual Population Laboratory

3.1 The development and evaluation of public health policy and especially international public health policies is constrained by the lack of a laboratory in which quantitative scenarios on future health events can be developed, impacts of disease patterns studied, and interventions evaluated. The VPL will provide a permanent resource for the consolidation of data and the modelling of disease for informing public health policy and action. The major activities of the VPL will be:

VPL as a centre for public health

- Collection and assembly of global data on health and its determinants.
- Integration and application of this data for building scenarios and evaluating interventions.
- Development of Information Technology in support of this activity.

3.2 The VPL will be part of a programme that will see the creation of regional centres whose aim is:

- The identification and acquisition of regional data that form part of the Global Health Grid
- The consolidation and integration of surveillance data to strengthen local health systems development
- Development of IT based community surveillance systems

VPL as part of world-wide integrated system

The proposal includes resources for co-operation with the WHO Tunis Centre (WMC) that will provide data for the VPL and with WHO-HQ on modelling development.

WHO, WMC

4. VPL Programme

The programme has three main components which are represented graphically in Figure 1 below.

modelling

4.1 *Modelling diseases dynamics and evaluating alternative scenarios*. The VPL will use Monte Carlo methods to generate individual “data records” for each of the approximately six billion individuals, based on known or assumed statistical distributions of human characteristics. In these synthetic or virtual “worlds”, individuals will be created (“born”), will have demographic, genetic and behavioural characteristics, engage in activities, move, and interact with others and die. By changing disease transmission parameters, and modifying behaviour patterns estimates of the likely effects of alternative intervention can be assessed.

4.2 *Data consolidation.* The VPL will provide in-house expertise on the harmonization of a variety of aggregate data sources necessary to build consistent data sets. The service offered will encourage researchers and visitors from around the world to “visit” the VPL, remotely or in person, to contribute in the development and maintenance of relevant data in their area of expertise. Initial data consolidation will focus on UK data, and data supplied by WMC. VPL will negotiate access to existing data and, where appropriate, commission studies.

role in consolid-ating world-wide data

4.3 *Computing infrastructure*. Clearly the volume of data and the calculation of the interaction of 6 billion individuals require significant computing “horsepower”. Fortunately, recent developments in information technology provide a cost-effective solution to developing systems for high performance data storage, manipulation and calculation specifically for Monte Carlo applications.

analytic power

5. Modelling

5.1 The VPL will create the facilities for discrete time, dynamic micro-simulation of populations up to and including the world’s population of 6 billion individuals. This methodology is well established on smaller scales[1, 8] and several Governments currently use similar techniques for their Public Health policy (e.g. Canada, Sweden, Australia.)

existing micro-simulations

structure

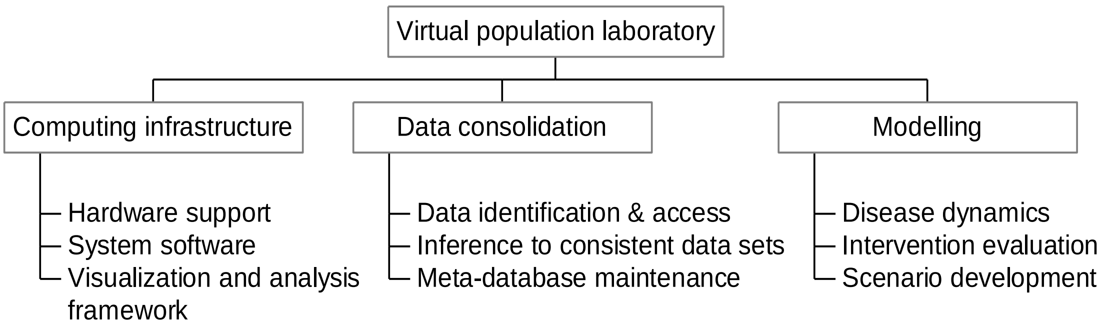

Figure 1: Organization Chart showing the three main contributing areas of expertise to the VPL.

existing expertise

5.2 Substantial research (within our collaboration) has taken place on the creation of locally realistic populations based on anonymized (2%) samples of census data within the UK [5]. In essence, the VPL will apply these methods to the global population. Where the data are sparse or unreliable this will introduce a quantifiable range of uncertainties into the VPL models. There also exists, within our collaboration, statistical expertise on the mathematical modelling of disease and population biology[9].

dynamic modelling

5.3 The VPL approach addresses three problematic aspects of dynamic modelling: the assumption of a closed population, assumption of random mixing patterns and the generation of likelihood distributions of results[10]. A global model is naturally closed – there are no population flows in or out. The microsimulation methodology[11] allows the specification of ranges of mixing patterns within a single population. Furthermore, the Monte Carlo method generates results as a range of outcomes – as against a single prediction.

benchmark with static simulation

5.4 The static properties of the estimated synthetic populations will be investigated, as a precursor (and benchmark) to full dynamic simulation. For example, on the basis of the proposed data consolidation activities VPL will be able to make improved estimates of actual (as opposed to expressed) demand for health care. In this way areas of poor health and areas of severe health care under-provision can be identified. Simple "static" simulation techniques could then be used to estimate the cost effectiveness of various health care provision strategies or the likely impact on the population of alternate health care interventions. This is a process directly analogous to current UK government tax-benefit modelling[12], but on a larger scale and with an emphasis on health.

6. Data Consolidation , Verification and Integration

6.1 In order to create realistic virtual/synthetic populations and their environments, a substantial amount of diverse information will have to be made available to the models within the VPL.

example: climatic data

6.2 Local expertise in accessing climatic data will be incorporated into VPL. For example, a model of malaria transmission is currently under development in the MALSAT research group, Liverpool School of Tropical Medicine. This is essentially a stochastic malaria transmission model - with health outcomes as the output, which is driven by the effects of climate on the vector and parasite population. The model[13] is being developed to support a collaborative study with the European Centre for Medium Range Weather Forecasting to: a) assess risk indicators for the use in malaria early warning systems and b) assess the cost-effectiveness of the timing of interventions in malaria epidemics.

economic factors

6.3 Not all the essential data is demographic or climatic. Important information will be consolidated on the economic microstructure of regions. This is of particular interest to the World Bank who need to compute and model the supply and demand of Health. We will draw on experience and data from World Bank[14] and WHO.

6.4 We will identify and acquire data on the features of the physical world (both natural and human-made) such as terrain, land cover, species distribution, rainfall, and transport networks. The MALSAT research group has extensive experience[15] on the use of environmental data in predicting disease distribution in space and time and currently uses extensive data archives on the global environment. In addition information on social and behavioral characteristics such as family size, educational levels and distributions, population density, health history, nutritional status, levels of poverty, access to heath care, and other determinants of health will be collected.

repository for data

6.5 Two forms of data integrity checks will be carried out as the data are collected; namely imputation and verification. Data imputation to fill lacunae and reconcile disparate categorizations is critical to handle heterogeneous data sets. The JMU group has extensive experience of SOLAS software as part of a collaborative project with industry at Unilever Research. Data verification requires modelling to detect outliers, and segmentation to identify interpretable sub-groups with unique defining characteristics. The mapping of phenomenology where there are significant non-linear interactions between covariates requires the use of optimisation techniques - both of traditional statistical models, as well as neural network models. The JMU group has expertise in both methodologies applied to the segmentation of demographic data, with reference to modelling of complex data for robust predictive modelling[16] and non-linear modelling of censored data and medical data in general[17].

quality and data integrity control

6.6 Knowledge integration and simulation of clinical services will play a key role in the project. The information on which to base these simulations must come from a wide range of sources internationally, and will be collected at differing granularities, and according to differing coding systems. It is widely recognised that the outcome of care is frequently dependent on detailed individual and care factors not covered in standard statistics, which will have to be estimated and imputed from a variety of alternative sources. While some resources use common coding systems, many different countries use different classifications and reference frames, which must be harmonised and integrated in order to be meaningful. The information in these databases is often not directly comparable, but can be integrated successfully if the differences between the different sources can be represented formally. NoEIHI has substantial expertise in this area.[18] Furthermore, the information available will evolve rapidly over time, so that the schemas and data integration must likewise evolve rapidly during the course of the project.

integrating heterogen-eous data

6.7 The North West Public Health Observatory and the NoEIHI will play a major role in reviewing health data available, integrating the data, developing the criteria for evaluating data quality and identifying key missing data sets.

NWPHO NoEIHI

6.8 WMC proposes to serve as a key node within such environment for data collection and consolidation to feed the "Virtual Population Laboratory". WMC will focus on two main aspects. First, enhancing information services allowing secure and reliable communication between primary health care workers, public health centres. The information flow will allow online information to local health authorities and facilitate both the response to problems as well as the collection of statistics. It will form the nucleus of the next generation of health reporting and community- based surveillance. Second, it will contribute key information to the data collection centres created to analyze health, demographic and related information. Information stored will be available globally, subject to user authentication. This information will in turn contribute to the work of the VPL.

WMC role as prototype VPL node

collecting data

6.9 Within a few years such systems are expected to lead to a dramatically improved response to health problems managed by remote communities together with the higher levels of the health system. Preliminary contacts are being made with several mobile telecommunication companies (e.g. NOKIA) to explore conditions of field-testing such systems. Grid tools (e.g. GLOBUS from Argonne National Laboratories) will be used for the sharing and authentication of databases with VPL

6.10 WHO-HQ will be a prime "user-site" of VPL. Access to the VPL from WHO-HQ will be through the "middle-ware" - visualization and access software - that provides the interface between public health experts, the data, and modelling. The development of user-friendly remote-access software is critical to this project and VPL will require staff dedicated to work on this. WHO-HQ will also collaborate in developing the modelling techniques and, where necessary, assist in negotiating access to relevant national data.

WHO-HQ

7. Computing infrastructure

7.1 The advent of low-cost commodity processing, networking and storage has redefined the nature and availability of high performance computing. A revolution in information technology (IT) is well underway with groups across all the Research Councils requiring resources that would only have available (at great expense) to traditionally computer intensive subjects such as Particle Physics. Relatively new disciplines, such as Bio-Informatics, may soon outstrip the Physical Sciences in their need for high-speed, networked access to substantial computing power. First-rate computing facilities represent part of the modern laboratory and in many cases are vital to the success, and leadership, of UK research.

Figure 2: View of the MAP facility. Each rack contains 30 MAP nodes – a total of 300 in all. MAP has been fully operational since December 1999.

enabled by IT revolution

IT vital for research

expertise in low cost super-computing

7.2 In January 1999, the Department of Physics at the University of Liverpool was awarded a JREI grant to build a Monte Carlo Array Processor[4] (MAP). The system, together with the flow control software also developed at Liverpool, has met and surpassed its design goals. To this date MAP has enabled Liverpool University to analyze and study much larger simulated data sets than is possible (or available) at the biggest facilities in the US, Europe or the UK.

7.3 A very important factor for the comparison of MAP with other systems has been its power, low original initial cost, low overheads and great reliability. Unlike other large systems, it does not require expensive recurring costs for batch control software, operating systems or maintenance. MAP is ideal for providing the backbone of the computer infrastructure necessary for the VPL

scale of VPL similar to Met office

7.4 Based on the global population and the required number of interactions per dynamic interval modelled, the VPL has been assessed as needing 5000 1GHz CPU's by the year 2005. The implementation of VPL assumes a level expenditure allowing for large scale testing of the system early on whilst leaving sufficient funding for the purchase of higher power CPU's as VPL becomes more predictive.

Year 1

8. Milestones and Project Phase

- **2001: Establishment of VPL**

Q1†: Establishment of VPL Executive Board from Stakeholders

- Q2: Installation and commissioning of VPL compute hardware
- Q2: Installation of prototype remote data centre
- Q3: Completion of GRID software for authenticated communication between VPL and remote and centres.
- Q3: Completion of Software specification (based on existing models) for: generating populations dynamics, disease processes, interaction physical world
- Q3: Complete staff recruitment for VPL

† **(Q1=1st Quarter, Q2=2nd Quarter, Q3=3rd Quarter, Q4=4th Quarter)**

- Q4: Completion of VPL control software for input of data, visualization, and model parameters.
- Q4: Outline of requirements for data integration and verification.

2002**: Start of modelling and data consolidation**

- Q1: Preliminary tests of large scale synthetic population creation based on UK census data for calibration

Q2: Static coarse level, demographic, economic, climate and social model constructed

Q3: First dynamic processes modelled: birth, death, ageing, migration, travel

Q3: Testing of transfer of data from VPL remote and local centres.

Q4: Benchmark of static modelling

Q4: Complete model framework for general disease processes. Malaria, tuberculosis, influenza, meningitis and HIV/AIDS.

2003: Refinement of prototype

Q1: Incorporation of Climate and Physical Terrain

Q2: Incorporation of Economic Indicators

Q2: Benchmarks of Static Modelling

Q3: Access to VPL by remote centres, WHO HQ and WMC.

Q4: Incorporation of “disaster” scenarios, manmade and natural.

2004: Testing of VPL

Q1: Preliminary results from testbed system including remote access of data from remote sites.

Q3: Results of climate and economic factors on data presented

Q4: Benchmarks of Dynamic models against historical data

2005: VPL readiness for exploitation phase

Q1: Report of Technical Board on VPL to national and international agencies

Q2: Proposal for VPL as central part of Global Health Grid based on the VPL prototype made to national and international funding agencies

9.

Year 2

Year 3

Year 4

Year 5 sustainable phase(new funding)

executive and technical leadership

lead institute

ability to attract industrial collaborators

building partnerships

Organization

9.1 The VPL will be organized around an Executive Board that represents the collaborating institutes and organizations. The Executive Board will be responsible for the strategic planning of the VPL, its reporting to funding agencies, matters of policy, organizing an annual assessment and dissemination of results. The Executive Board will select a Technical Board composed of an Executive Director, and Coordinators (Directors) for each one of the subsystems: Computing/Networks, Statistics, Data Integrity, Epidemiology/Modelling, Surveillance. The Technical Board will be drawn primarily from local institutes and staff recruited for VPL. Within 2 years of operation of VPL, we expect that other areas of expertise will be added to the Technical Board e.g. Climate, Economics, and Health Policy.

9.2 The University of Liverpool will take responsibility for contractual and finance matters. Space for VPL is included in the overhead in Section 12. Discussion is underway with Merseyside European Initiatives for the provision of a new building.

10. Partners

10.1 We are actively pursuing commercial, high technology collaboration with Telecommunications companies (for community-based surveillance) and IT companies (e.g. IBM – see Appendix). These will add to the value of the bid through their existing expertise and strategic development partnerships. The success of the project is not contingent on receiving their active participation.

10.2 The NIH, World Bank, and CDC support this project. These international centres will act as a source of staff and consultants for VPL. We are working to establish formal partnerships with these organizations and funding from, through, and in collaboration with these organizations will be sought. These will be for further posts at the VPL and for the establishment, and enhancement of, remote data-sites such as the Tunis Centre

11. Summary of Bid

world class centre for public health modelling and data consolidat-Ion

11.1 The University of Liverpool Department of Physics and the World Health Organization have initiated the Virtual Population Laboratory project. The collaboration now includes the North of England Institute of Health Informatics, the NW Public Health Observatory, and the Dept of Mathematics of Liverpool John Moores University, the Liverpool School of Tropical Medicine, and the University of Liverpool Departments of Geography and Mathematics. The proposal represents a synergy of expertise in techniques for multiple large-scale simulations of complex multifactorial events (originally developed within the Department of Physics to assist CERN in experiments in particle physics) with expertise in constructing microsimulated human populations (in respect of which the Department of Geography is recognised as the leading UK academic centre). This is complemented by centres of excellence in: mathematics and statistics (Liverpool Mathematics), data extraction and consolidation from health informatics (NWPHO, NoEIHI). The groups combine expertise in, health informatics and computing (JMU Mathematics, NoEIHI and Physics). The LSTM provides expertise on the transmission of disease and particularly its relation to climatic data and information on the effects of interventions designed to prevent or mitigate the effects of disease outbreaks. Our international collaborators provide expertise on epidemiology and public health (WHO, CDC and NIH). The World Bank (and NWPHO) provide expertise on economic factors.

opportunity to attract centre to North West

11.2 Funding from the NW Science Review will establish one part of a prototype three-cornered network - the major analytical centre being linked to the communicable disease programme at WHO (Geneva); and forge a close collaboration with the WHO Mediterranean Centre at Tunis (which would be - in effect - a prototype end-user). This network will develop the potential for exploiting internet technology and the expertise of high-powered analytical centres together with those of epidemiologists, to tackle multifactorial and multidimensional health crises in a world of high mobility and rapid social interactions. In the medium term the population sets generated may be enhanced with descriptors of health systems organisation - e.g. by modelling primary care registration lists - so as to be able to simulate the dynamic processes of health systems pressures and failures, such as 'winter bed crises'. In this context the North West region may also be constituted (through NWPHO) as an additional prototype end-user, subject to additional funding. The project will be directed by a Technical Team drawn from the collaborating institutes who will report to the Executive Board drawn from National and International Health Organizations. The team will develop links with other funding agencies for the long-term finance of the VPL. The team will ensure that the VPL informs and communicates with the public about its activities as a Global Centre for Health Modelling.

outreach

spin-off

11.3 The VPL will, like other international centres, have a positive impact on the future of the North West economic and science base. Its requirements and health-oriented goals will attract satellite hi-technology firms in the IT (networking, software and hardware) sector together with IT-oriented Pharmaceutical Companies to collaborate on the project. There is a large potential for spin-off technology in the region. As a prototype node the NW will, in the long term, benefit from information extracted from VPL for the improvement of its health. The VPL will also act as a centre for the training of a new generation of specialists

training

12 Resource/Budget

The budget is an estimate based on a 5-year programme as described above. Staffing costs are subject to 46% overhead.

	2001	2002	2003	2004	2005	2001/2005
Hardware						
MAP-nodes	£500,000	£500,000	£500,000	£500,000	£500,000	£2,500,000
COMPASS-nodes	£200,000	£100,000	£100,000	£100,000	£100,000	£600,000
Cooling	£10,000	£10,000	£10,000	£10,000	£10,000	£50,000
MAP-infrastructure	£20,000	£5,000	£5,000	£5,000	£5,000	£40,000
Offices	£100,000	£103,000	£106,090	£109,273	£112,551	£530,914
Sub-Total Hardware	£830,000	£718,000	£721,090	£724,273	£727,551	£3,720,914
	2001	2002	2003	2004	2005	Total
Staff Technical						
MAP coordinator	£45,000	£46,350	£47,741	£49,173	£50,648	£238,911
2-MAP systems	£60,000	£61,800	£63,654	£65,564	£67,531	£318,548
2-Globus	£60,000	£61,800	£63,654	£65,564	£67,531	£318,548
2-Data-base	£60,000	£61,800	£63,654	£65,564	£67,531	£318,548
Data-entry	£15,000	£15,450	£15,914	£16,391	£16,883	£79,637
Sub-Total Technical Sta	£240,000	£247,200	£254,616	£262,254	£270,122	£1,274,193
Staff-Public Health/Professional						
Director	£100,000	£103,000	£106,090	£109,273	£112,551	£530,914
2 Epidemiologist	£120,000	£123,600	£127,308	£131,127	£135,061	£637,096
2 Statisticians	£100,000	£103,000	£106,090	£109,273	£112,551	£530,914
4 Monte Carlo Programn	£140,000	£144,200	£148,526	£152,982	£157,571	£743,279
5 Data Integrators	£200,000	£206,000	£212,180	£218,545	£225,102	£1,061,827
Consultants	£500,000	£300,000	£309,000	£318,270	£327,818	£1,755,088
Sub-total	£1,160,000	£979,800	£1,009,194	£1,039,470	£1,070,654	£5,259,118
Administrative						
Manager	£70,000	£72,100	£74,263	£76,491	£78,786	£81,149
3 Support	£60,000	£61,800	£63,654	£65,564	£67,531	£69,556
Travel/Conferences	£100,000	£103,000	£106,090	£109,273	£112,551	£530,914
M&O	£50,000	£51,500	£53,045	£54,636	£56,275	£265,457
Total	£3,600,000	£3,141,100	£3,216,883	£3,294,939	£3,375,338	£16,460,418

References

[1] See, for example, CORSIM programme, http://www.strategicforecasting.com/

[2] Anderson, R and May, R: *Infectious Diseases of Humans: Dynamics and Control.* Oxford U. Press, 1991.

[3] Rvachev LA, Longini IM: *A mathematical model for the global spread of influenza*. Mathematical Biosciences 75:3-22; 1985. Longini IR, Fine PEM, Thacker SB: *Predicting the global spread of new infectious agents*. American Journal of Epidemiology. 123:383-391, 1986. Cliff, A , Haggett P, Smallman-Raynor, M: *Deciphering Global Epidemics, Analytic Approaches to the Disease Records of World Cities*, 1988-1912, Cambridge U. Press, 1998.

[4] Ackerman E, Elveback LR, Fox JP: *Simulation of Infectious Disease Epidemics*. Charles C Thomas. Springfield Illinois. 1984. Bailey NTJ: *The simulation of stochastic epidemics in two dimensions.* Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 4:237-357. Berkeley and Los Angles: University of California Press., 1976. Elveback LR, Fox JP, Ackerman E, et. al: *An influenza simulation model for immunization studies.* American Journal of Epidemiology 103:152-165; 1976. Elveback LR, Ackerman E, Gatewood, et al: *Stochastic two-agent epidemic simulation models for a community of families*. American Journal of Epidemiology. 93:267-280; 1971. Watts Duncan*: Small worlds: The dynamics of networks between order and Randomness.* Princeton University Press, Princeton, New Jersey, 1999. C.F. Citro and E A Hanushek (eds) 1991, *Improving information for social policy decisions: the uses of microsimulation in modelling*, National Academy Press, Washington DC

[5] Voas D. and Williamson P. (forthcoming) *'Evaluating goodness-of-fit measures for synthetic microdata'*, Geographical and Environmental Modelling. Williamson, P. (Forthcoming) 'Synthetic microdata', Ch 18 in Rees P., Martin D. and Williamson P. (eds) *The Census Data System*, TSO: London.. Williamson, P., Birkin M. and Rees P. (1998) *'The estimation of population microdata by using data from small area statistics and samples of anonymised records'*, *Environment and Planning A*, 30, 785-816. Voas D. and Williamson P. (submitted) *'An evaluation of the combinatorial optimisation approach to the creation of synthetic microdata'.*

[6] T. Bowcock, *MAP Project*, Conference on Computing in High Energy Physics, Padova, 2000.

[7] Foster I & Kesselman C: *The Grid: Blueprint for a new computing infrasture*. Morgan Kaufmann Publishers, Inc. San Franscisco, California, 1999

[8] See also, "National Microsimulation Resource", Minnesota,US. http://www.labmed.umn.edu/nmsr/

[9] Reade, B., Bowers, R. G., Begon, M. E. & Gaskell, R. M., 1998*, A model of disease and vaccination for infections with acute and chronic phases.* Journal of Theoretical Biology 190, 355-367

[10] D Clancy(1996) "*Strong approximations for mobile population epidemic models*",. The Annals of Applied Probability 6, 883-895. FG Ball and D Clancy, "*The final size and severity of a generalized stochastic multitype epidemic model*", Advances in Applied Probability 25, 721-736.

[11] G. P. Clarke (ed) Microsimulation of urban and regional policy analysis, Pion London, A Harding (ed) (1996) *Microsimulation and public policy*, North Holland, Amsterdam.

[12] R Eason(1996) *Microsimulation of direct taxes and fiscal policy in the United Kingdom*. Ch 2 A Harding (1996).

[13] Worrall, E., Connor, S.J., & Thomson, M.C. (submitted). *Modelling the cost-effectiveness of epidemic malaria control interventions over time.* Abstract Oxford 2000. Thomson, M.C., Palmer, T., Morse, A.P. Cresswell, M. & Connor, S.J. (2000). *Forecasting disease risk using seasonal climate predictions*. The Lancet. (Letter) 355, 1559-1560

[14] Murray, J.L., Lopez, A.D. *Global Health Statistics*. Global Burden of Disease and Injury Series.Harvard: WHO/World Bank/Harvard University, 1996.

[15] Thomson, M.C., Obsomer, V. Dunne, M. Connor, S.J. & Molyneux, D.H. (in press). *Predicting trouble: satellite mapping of Loa loa prevalence in relation to ivermectin use in West/Central Africa*. The Lancet Thomson, M.C. & Connor, S.J. (2000) Environmental information systems for the control of arthropod vectors of disease Medical and Veterinary Entomology 14, 3,000-000. S.J. Connor, M.C Thomson D.H. Molyneux. (1999) *Forecasting and prevention of epidemic malaria: new perspectives on an old problem.*Parassitologia 41, 439-448

[16] Lee, Y.Y.B., Huang, Y., El-Deredy, W., Lisboa, P.J.G. and Harris, P. *'Robust methodology for the discrimination of brain tumours from in vivo Magnetic Resonance Spectra*' to appear in MEDSIP 2000, IEE Publishing, 2000. Lisboa, P.J.G., Kirby, S.J.P., Vellido, A., Lee, Y.Y.B. and El-Deredy, W. *'Assessment of Statistical and Neural Network Methods in NMR Spectral Classification and Metabolite Selectio*n' Nuclear Magnetic Resonance in Biomedicine, 11, 225-234, 1998.

[17] Lisboa. P.J.G., Vellido, A. and Wong, H. *'Bias reduction in skewed binay classification with Bayesian neural networks*' to appear in Neural Network Letters, 2000. Lisboa, P.J.G., Ifeachor, E.C. and Szczepaniak, P.S. (eds.) *'Artificial Neural Networks in Biomedicine*' Springer, London, 282pp, 2000.

[18] Baker, P., Brass, A., Bechhover, S., Goble, C., Paton, N. and Stevens, R. (1998). TAMBIS: *Transparent Access to Multiple Bioinformatics Information Sources. An Overview*. Sixth International Conference on Intelligent Systems for Molecular Biology, ISMB 98, Montreal:25-34.