AUSTRALIAN BIOINFORMATICS NETWORK
FINAL REPORT

Bioinformatics Australia Management Committee, term 2007-2008 at the face-to-face committee meeting in August 2007. L-R: Dr Anna Lavelle (AusBiotech CEO), Prof Shoba Ranganathan, Dr Bruno Gaëta, Dr Tim Littlejohn (president), Dr Rohan Teasdale, Dr Annette McGrath, Prof Mark Ragan (vice-president), A/Prof Phoebe Chen, Dr Lucia Santoso (ABN EO), Dr Michael Poidinger. Not present: Dr Catherine Abbott.

Bioinformatics Australia inaugural Management Committee, at the face-to-face meeting in August 2006. L-R: Dr Catherine Abbott (vice-president), Dr Dominique Gorse, Dr Anna Lavelle (AusBiotech CEO), Dr Mark Crowe, A/Prof Phoebe Chen, Dr Rohan Teasdale, Dr Jonathan Arthur (president), Dr Bruno Gaëta.

Author of this report: Dr Lucia Santoso,
                           Executive Officer,
                           Australian Bioinformatics Network
                           AusBiotech, Ltd
                           Level 1, 322 Glenferrie Rd, Malvern, VIC 3144

TABLE OF CONTENTS

ACKNOWLEDGEMENT

EXECUTIVE SUMMARY

Bioinformatics has been named as an area of emerging sciences where Australia has strength and that are important to Australia in the Australia's Science and Technology Priorities for Global Engagement, by the PMSEIC working group. It has been considered an important enabling technology to discoveries in life- and medical-sciences, such as the various genome projects, identification of genes pertaining to diseases and prediction of protein structures and functions.

Bioinformatics activities started in Australia in early to mid nineties. Relative to the counterparts in places such as Europe and the USA, the community is relatively young, small and fragmented. The National Bioinformatics Strategy recognised this situation and potential of Australia's role in international bioinformatics arena, and recommended that efforts of coordination be made, in the form of the establishment of the Australian Bioinformatics Network, encompassing six key areas: Infrastructure, Research and Development, Education and Training, Commercialisation, Data Management and Coordination.

This document reports the conclusion of the Australian Bioinformatics Network project, which is the beginning of the development of the Australian bioinformatics community. In the duration of this project, the community has been consulted. The following are the findings resulting from the discussions amongst the community.

Key Findings
1. The greatest need in infrastructure is in human infrastructure, namingly skilled people who are able to understand the life- or medical-science questions, take the available data and wield or develop tools to analyse them.
2. This need in human infrastructure is expected to rise with the drop of the cost of generating data, hence the increase of data size.
3. The capacity of hard-infrastructure facilities are not exhausted. In some cases, they are in fact under-utilised by the bioinformatics community, possibly because many members of the community are not sufficiently skilled to take advantage of high-performance computing capabilities.
4. High-performance computing facilities used by the bioinformatics community are typically staffed by skilled support personnel.

Recommendation 1
Each hard infrastructure provider dedicated to support the bioinformatics community must be equipped with expert/trained support staff and develop strategies of optimal utilisation.
Recommendation 2
Each institution whose activities include management and analysis of large sizes of biological data should have in-house informatics support staff.

---

5. Key-stakeholders of the research community expressed that the success rate of funding for bioinformatics research is poor, particularly for the groups embedded

within medical research institutes, rendering them largely dependent on NH & MRC as a funding source.

6. Key-stakeholders of the research community expressed that the allowance for informatics support in life-science/medical-science research is not sufficient. In many cases, only a fraction of the proposed amount is granted and bioinformatics support to that piece of research is often the component which suffers, which opens possibilties of inaccurate or inappropriate data-handling or – analysis.

Recommendation 3

A review should be conducted of the level of activities in bioinformatics research, as well as use or need of bioinformatics, in medical research, in comparison to the level of funding.

---

7. The bioinformatics community is by-and-large academic and research.
8. Career structure in bioinformatics in unclear and recruitment level has historically been relatively low, evidenced by the small numbers of job advertisement and exit of bioinformaticians to other work sectors. A sharp rise in vacancies in late 2007 suggests a beginning of bioinformatics uptake in related areas.

Recommendation 4

Efforts should be made in provision of intermediary services, to survey the health and biotechnology industries of their possible needs of bioinformatics and explore how bioinformatics can improve their performance, and to leverage the observed uptake of bioinformatics in bioscience research.

Recommendation 5.

Internship programmes should be explored, for placements of bioinformatics students in industries enabled by bioinformatics. This is expected to create the demand in employment as well as enriching the learning experience and

---

9. There is a need in professional development courses in bioinformatics skills for life- or medical-scientists.
10. ANGIS (now Sydney Bioinformatics) seems to be the only successful organisation in the provision of bioinformatics training courses. Its capacity has not been able to meet the demand in professional development in bioinformatics skills.

Recommendation 6

The role of ANGIS as a provider of professional development in bioinformatics skills should be further explored, including the possibility of expanding its operation to meet the demand.

---

11. Dedicated undergraduate degree-course in bioinformatics is not enouraged, given the low employment uptake in bioinformatics.
12. Master degree qualitification is much preferred, with a bachelor degree in one of the core-disciplines. A good master degree course should be tailored to the student's background.

Recommendation 7
Bioinformatics should be offered at the undergraduate level as a stream of one of the core disciplines. Such a course should offer all the abovenamed components (IT, statistics and life-sciences) with a flexibility to vary its proportion within the course, with a component which involves teamwork (such as a group-project).

At the master degree level, bioinformatics courses should be further customised to the background of the students.

---

13. There is currently only a low level of engagement between the mostly academic bioinformatics community and the commercial sector.
14. Industry related to bioinformatics in Australia (biotechnology and ICT) are typically small-to-medium enterprises (SME). The available incentive for interaction (such as ARC linkage grants) are more feasible to very large companies and are not attractive to industry of this size.
15. There is a strong open-source culture in the bioinformatics community, that poses as a hurdle to commercialisation.
16. Freely available bioinfomatics tools are not always easy to use. Commercial packages with embedded work-flows and pipelines are prohibitively expensive. There may be a market in small commercial packages for small-scale needs.

Recommendation 8
More potential and opportunities for the research community to engage with the private sector should be explored. For example:
   (a) funding should become available in the form of seed-funding to develop a product required or developed by the bioinformatics community,
   (b) collaboration should be encouraged between the tool-developing research community and commercial organisations to develop more custom software, with smaller number of components and which would cost less than the currently available commercial products.
Business models should be developed with consideration of the open-source culture of the research community, e.g., in packaging free software to render them easier to use.

---

17. Data management issues are of concern to health-care research, where bioinformatics may play a part therein, for example, in genetic epidemiology.

Recommendation 9

Efforts should be coordinated in education of medical (and life-science) researchers on the existing data standards and the importance of adhering to these standards. The education efforts should be done in conjuction with organisations currently pursuing these endeavours, such as the Western Australian Institute of Medical Research.

---

18. The Australian bioinformatics community has enjoyed and profited from opportunities to meet and conduct discussions, of scientific and community-building issues. Two Bioinformatics Australia conferences held in conjunction with AusBiotech 2006 and 2007, have been very well received. The community strongly indicated the desire for an ongoing annual conference.
19. Australian bioinformatics profile has been noticed and acknowledged in the region (Asia and New Zealand). There is clear potential to play an active part in the Asia-Pacific region.

Recommendation 10

The potential for Australia's leading role in the Asia-Pacific bioinformatics arena should be further explored. This would include:
   (a) continuation of the annual bioinformatics conferences, encouraging the participation of other countries.
   (b) taking the opportunities to host international conferences, such as the proposed Genome Informatics Workshop 2008.

Proposals

The efforts in the Australian Bioinformatics Network project constitutes the beginnings of the establishment of the bioinformatics community. There clearly are ongoing and new activities to be followed, the need of which was identified in during of the project.

The young Australian bioinformatics community is starting to show its potential in the world arena. The recommendations following the findings should be considered, to further foster and nurture this community to its maturity.

CHAPTER 1: INTRODUCTION

Bioinformatics is defined by the US National Institute of Health to be the research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data. Major research areas in bioinformatics include genome and protein sequence analysis, genome annotation, computational evolutionary biology, analysis of gene expression data, analysis of gene regulation, prediction of protein structure, modelling of biological systems, high-throughput image analysis and protein interaction networks.

Bioinformatics activities started in Australia in the early to mid nineties. Being a young and cross-disciplinary, the bioinformatics community in Australia is fragmented. Findings in the National Bioinformatics Strategy recommended, amongst others, that efforts for coordination be made that a bioinformatics network be established.

To carry out the recommendations of the NBS, a call was made coordination efforts to establish the Australian Bioinformatics Network. The Network comprise the community, and means to bring them together, including facilitation of discussions (face-to-face, online and by telephone), representation, coordination of activities and awareness raising (both to the community itself and to the related industries).

The Australian Bioinformatics Community

Bioinformatics activities in Australia occur in three different tiers: bioinformatics research, bioinformatics services and embedded bioinformatics in related research (such as life-science and health-care). Bioinformatics research constitutes development of new bioinformatics algorithms, and this sub-community is the smallest in number. The largest sub-community comprises life- and medical-scientists in whose research bioinformatics is an important tool. The services sub-community are those who provide bioinformatics support, consultancies and services. These layers in the community can be presented in the triangle below:

Bioinformatics researchers

Bioinformatics service providers

Bioinformatics users
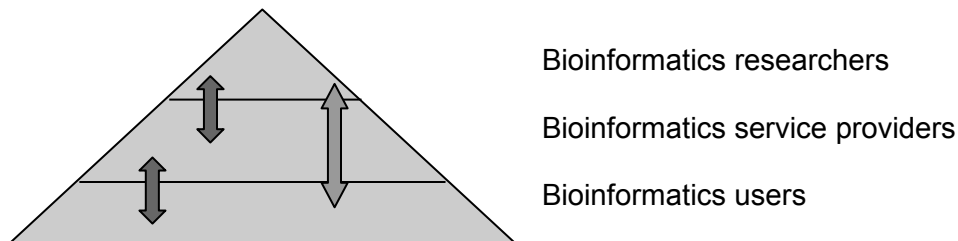
In some cases, the services group acts as an interface between the new algorithms developed by bioinformatics researchers and the research community who uses them. In other cases, developers and users of bioinformatics algorithm interacts directly by means of research collaborations

All three groups compose important aspects of the community. Without the users, there is no purpose to bioinformatics. Without the tool developers, there will be no bioinformatics. Without the service providers, the community cannot function as a whole.

All three tiers of the community have been consulted in this project. In the context of this report, the above terminologies are used. The term "researchers who use bioinformatics" is also used interchangably with "bioinformatics users."

Scope, methods and structure of this report

This report addresses the recommendations on the six issues explored in the National Bioinformatics Strategy (NBS), and stated in the Australian Bioinformatics Network project description. In the course of this project, there are a number of completed and ongoing activities, where findings lead to recommendations (stated in the Executive Summary). This aim of this document is to report on the state of the Australian bioinformatics community, and to make recommendations on future activities for its further development. This main body of this document therefore starts with the key-acitivities which yield in further findings and recommendations. Reporting on completed activities follows.

Information is gathered by consultation with the bioinformatics community, whose scope is described in the previous section. This consultation process is conducted by face-to-face gatherings (such as the Bioinformatics Stakeholder Consultative Workshops and during the two Bioinformatics Australia conferences), site-visitations, via the bioinformatics internet portal (such as the online discussion forum and e-mail discussions) and telephone conversations. The majority of these activities are carried out by the Executive Officer of the Australian Bioinformatics Network, guided by the CEO of AusBiotech and the Bioinformatics Australian management committee (Appendix A). The list of people consulted who are not in the committee is enclosed in Appendix B.

The main body is preceded by a summary of achievements, in the format of the Australian Bioinformatics Network grant deed.

Key activities which lead to ongoing efforts comprises the Bioinformatics Stakeholder Consultative Workshops, the annual Conferences. This chapter also outlines the functions of the Australian Bioinformatics Network website as a portal of information and facilitator of collaborations, and the Networking and Coordination roles that the ABN plays.

These are followed by reporting of completed activities, namingly the Review of ANGIS and the Establishment of a family of Research Fields, Courses and Disciplines classifications.

A summary of recommendations of future directions conclude this document.

ISSUES IN BRIEF

Infrastructure

Three aspects of infrastructure needs have been explored, namingly the two more traditional classification of hard infrastructure (including computer machinery and computer network) and soft infrastructure (including software and networking protocols), and human infrastructure (expert staff).

Hard infrastructure for bioinformatics are provided by the various partnership for advanced computing, notably APAC and SAPAC, as well as the newly established Queensland Facilities for Advanced Bioinformatics (QFAB). High-speed computer network is provided by AARNet and other state-based consortia. Available high-performance computing capability is expected to be able to fulfill computation needs of the bioinformatics community.

The main issues are in the distribution/allocation of which facilities are capable of supporting bioinformatics in terms of available trained support staff.

In terms of soft infrastructure, the bioinformatics community mostly uses free, open-source software. The nature of infrastructure need in this area is in the availability of expert users who are able to wield these tools to analyse life-science data, with the understanding of both the limitation of each tool and the limitation in data-generation accuracy, namingly human infrastructure.

A few commercial software products are considered to be useful and desirable, however the research community and infrastructure providers find the licence cost prohibitive. It is further claimed that these software suites include applications which are not directly relevant, leading many groups to resource to provision of staff with IT-expertise to develop their custom-tools in-house.

Human infrastructure is an area that is clearly unsatisfactorily provided for, and it is expected that this need will increase with the decrease of cost in data generation and hence the rise in data size. It is the most important of infrastructure needs, as human infrastructure provides the interface between the tools (both hardware and software) and their end-users. The required skills are further discussed under "Education and Training."

Research and Development

The research community in bioinformatics loosely can be divided to two groups: those who perform research in bioinformatics and those who perform research in life-sciences or medical-sciences which employs bioinformatics in their methods, as illustrated in the triangle in the previous section. The activities of the former involve the development of new or improved algorithm to enable better analysis of large life-sciences data, sometimes with a focus of addressing particular questions. The latter uses these techniques to answer research questions in life-sciences, which may involve the composition of a selection of tools into a pipeline or minor modifications of existing tools. This community is much greater in number than the former.

Due to its cross-disciplinary nature, the community feels that contributions proposed bioinformatics grant applications are difficult to appreciate by both the informatics (mathematical and information sciences) and life sciences communities. The culture of the former tends to appreciate novel developments, and the latter holds to well-established and -tested methodologies. The perception follows therefore that bioinformatics grant proposals may not be appriopriately reviewed.

Discussions have been held with funding bodies, on behalf of the community, where these agencies disagree that bioinformatics is under-funded in comparison to the average. Statistics to reflect the success rates have been difficult to obtain, due to the absence of RFCD codes for bioinformatics. These codes are being established, and will take effect in 2008. As another result of the consultation process with the Australian Research Council, a member of the bioinformatics research community has been appointed in the College of Experts. This development should help bioinformatics researchers and reviewers with the granting process.

Education and Training

The cry for more human infrastructure is followed by the need for education and training, to add to the pool of skilled work-force. Three tiers of training courses were discussed, short courses to up-skill existing labour force (e.g., informatics skills for bioinformatics users and life science background knowledge for IT professionals), undergraduate and post-graduate by coursework training.

The above-mentioned need in skilled labour and hence training do not seem to be followed by the economic demand in either. Currently there is a low uptake of bioinformatics expertise in the workforce. Career structure in bioinformatics (in particular bioinformatics research) is unclear, leading to exit of mid-career bioinformaticians. Training courses at all levels are mostly under-subscribed, leading to the closure of a number of courses and a course provider organisation.

These blatant inconsistencies calls for further investigations. Possible causes of low employment level and unclear career structure include poor engagement between the research community and private sector and poor awareness of need of expertise in management and analysis of large datasets. The commercial sector should be engaged to explore the need for bioinformatics the the Australian biotechnology industry, to raise awareness of the potential benefits of bioinformatics to that industry and to create networking opportunities and possibly employment opportunities. This issue should be firstly address before any further review of courses be conducted, as the courses should address the skill-set required by employers.

The few employers consulted indicate the demand for the following skills and expertise/knowledge:
- IT, namingly knowledge in databases and the ability to programme in at least one scripting/programming language (such as Perl, Python),
- Statistical knowledge, particularly in statistical analysis of large datasets and at least one statistical programming language (such as R),

- Background in life-sciences, sufficient to understand the hypotheses to be addressed and data-generation processes,
- Communication skills and team-work; as bioinformaticians are inevitably a part of multi-disciplinary teams, they are required to be able to communicate to both technical and biological communities.

The few employers consulted favoured graduates of a master degree in bioinformatics, following a bachelor degree in one of the core disciplines. Second-in-line is an undergraduate degree in the informatics disciplines with life-science components. Due to the low level of employment, dedicated degrees in bioinformatics is less favoured than a degree in one of the core disciplines with a bioinformatics stream. The dedicated courses are not deemed to add value from other courses with bioinfomatics stream.

The proportion of significance of each component in the set of required skills varies, depending on the acitivities of the employer.This implies that the ideal course in bioinformatics should offer all components but with a flexibility in proportion. For these reasons a unified curricullum is not recommended.

Commercialisation

There is currently a low level of engagement between the primarily research-based bioinformatics community and the private sector, which may explain the low economic activity and employment in bioinformatics, despite being named as one of Australia's strengths and potentials, and its increasing relevance to related industries such as biotechnology, IT and medical research. It was reported by a key-player in the IT industry that existing ARC linkage grants is not very attractive to SME (which are likely to engage in bioinformatics activities) due to its degree of bureaucracy and legalities. There were reported interests to from industry to collaborate with the informatics, which could be motivated by incentives in the form of start-up grants.

As previously mentioned, there was evidence that there is low awareness of the needs of informatics in the commercial sector, such as the health-care and biotechnology industries. This is an extremely crucial issue to address, as being an enabling technology, the absence of its uptake robs bioinformatics of its *raison d'être*. There clearly is a need for intermediary services activities to the industries downstream to bioinformatics.

The investigation of possible linkages to the commercial sector needs to be conducted taking into consideration the strong open-source culture in the bioinformatics community. This does not mean a close-door to commercialisation potentials, as there exist open-source business models. However, this difference in culture is likely to pose as an impediment in the partnership between the academic and commercial communities. These opportunities should be further explored, with close consultation with the academic and commercial communities.

Data Management

The pertinence of the issue of integrating inconsistent data formats is more noticable in the space of medical informatics/medical research, where there is more importance in the ability to share data. Some of these inconsistent formats are attributed to the proprietary formats of the equipment manufacturer, and some are due to lack of coordination in data collection, which in turn leads to a laborious effort in data-integration. There are already industry standards, however it is not readily adopted. It is proposed that the latter is mainly caused by the lack of knowledge by medical researchers in the available standards and lack of awareness/appreciation of the significance of consistency in data formats for the purpose of its large-scale management.

Coordination and Networking

Coordination and networking activities underpins the establishment of the Australian Bioinformatics Networks. These are the activities during which the community brings itself together, activities which facilitates the establishment of the network, the relationships, the togetherness which constitute a community. Fragmentation of a community often results from unfounded assumptions, due to lack of communication. These barriers have started to be broken down through facilitated, inclusive discussions.

Even in this era of modern technology, nothing replaces personal contact and face-to-face communication. This was evident in the atmosphere and the dynamics of the conferences, in which the community can gather and discuss matters that are important to them: science and the above-outlined issues, pertaining to community building. Such interactions should be conducted more than once yearly, which is where state-based gatherings such as the Consultative Workshops play an important role. It is crucial for the growth of this nascent community that these activities continue to be supported, in order to nurture it to its maturity.

Online means of discussion play a role in the growth and strengthening of a community, to facilitate continuing communication amongst the community members across the geographical distance. It is a means where ideas can be posed, discussed and debated. Its role has been tested with the discussions around the RFCD code, where the community had to present a united voice to the ABS. This forum, however, needs to be facilitated, mainly due to its non-intrusive nature. Its perpetuity needs to be equipped with the presence of a facilitator, whose role is to initiate and maintain discussions.

One of the challenges in a young community separated by distance is awareness of each other. This can be overcome by an introductory service, some kind of a "bioinformatics online-dating" facility, where one can find prospective collaborators, with particular expertise, in particular geographical locations. An online directory powered by a searchable database has been establish for this purpose. New members to the directory continue to flow, encompassing our keen trans-Tasman colleagues. Possibilities for joint-efforts in facilitation of communication and collaboration are being explored with BioPlatform Australia. As the case with the online discussion forum, this database also needs a manager, whose duties include actively seeking out new members and regular updates and maintainance of the directory.

CHAPTER 2: SUMMARY OF ACHIEMENTS

1. INFRASTRUCTURE	
1.1 Review the Australian National Genomic Information Service (ANGIS)	The review was conducted by a study of documents, statistics from ANGIS staff and feedback from ANGIS users. Its success in professional development activities suggest a real potential in this role and recommended that further investigation and possible investment be made to expand this function.
1.2 Promote Australian bioinformatics facilities, activities, professional services and consultancies, integration services, educational and training programs	The facilities, activities, services and programmes have been promoted in the website (www.ausbiotech.org/bioinformatics), conferences, consultative workshops, online discussion forum and Bioinformatics Australia (BA) mailing list.
1.3 Make recommendations to bioinformatics infrastructure suppliers in Australia	The main need is in human infrastructure, which relates to training and professional development. The hardware infrastructure facilities are often under-utilised as many users are not skilled in wielding high-performance computing, which again points to need of accompanying human infrastructure in the form of skilled support staff.
1.4 Work collaboratively with key National Collaborative Research Infrastructure Strategy (NCRIS) stakeholders	The community has been briefed at both Bioinformatics Australia conferences by Prof Matthew Bellgard, during which discussions occurred.
1.5 Link Australian bioinformatics stakeholders with the broader ICT field	NICTA is now an active participant. Discussions have occurred with IBM and Cray computers, both of which indicated that bioinformatics has not been a major consumer, and therefore is not a focus market.
2. RESEARCH AND DEVELOPMENT	
2.1 Establish a family of bioinformatics Research Fields, Courses, and Disciplines (RFCD) classification codes for Bioinformatics	A set bioinformatics-relevant RFCD classifications, comprising codes for bioinformatics, bioinformatics software, genomics, proteomics and system biology, will be established, to take effect in 2008.
2.2 Monitor grant funding opportunities	Listing of funding oportunities are posted on the ABN website and BA mailing list.
2.3 Help bioinformatics stakeholders prepare grant applications	Conversations have occurred with funding bodies, resulting in the appointment of a bioinformatics academic into the ARC College of Experts, which is expected to improve the review process of grant proposals.
2.4 Assist networking amongst the Australian bioinformatics community	Networking occurs through the means mentioned in Item 1.2.
2.5 Build ties with regional and international	Closer link is established with the New Zealand and Asian community, with participation in conferences and hosting of

bioinformatics organisations	a regional conference. Affiliation has been sought to the International Society of Computational Biology.
3. EDUCATION AND TRAINING	
3.1 Review current bioinformatics training in Australian and international institutions.	A master degree qualification is considered ideal, with a bachelor degree in a core discipline, with a stream in bioinformatics. Degree courses in bioinformatics are not favoured. Professional development programmes are sought in bioinformatics skills for established life-scientists.
3.2 Provide information on jobs in the Australian bioinformatics industry	Jobs are advertised on the ABN website, BA mailing list and online discussion forum.
3.3 Create a virtual faculty in bioinformatics	A common curriculum is not favoured by the community, hence this activity has been abandoned.
3.4 Sponsor small grants, fellowships and travel scholarships to assist students in developing their bioinformatics education	Travel scholarships were awarded to a number of students attending BA 2007 conference.
4. COMMERCIALISATION	
4.1 Link bioinformatics companies and bioinformatics research groups to the larger Australian biotechnology community	Current engagement with the commercial sector is poor. The role of bioinformatics in biotechnology SME (majority of Australian biotech industry) needs to be further investigated. Incentives for engagement and intermediary services are recommended for future projects.
5. DATA MANAGEMENT	
5.1 Coordinate data standards for bioinformatics	Inconsistent data standards are greater concerns of medical research, such as genetic epidemiology. Conversations have occurred with a major stakeholder, who is undertaking standardisation efforts.
5.2 Seek to provide mirrors of major international databases are provided for bioinformatics researchers	A list of existing mirrors can be found in the ABN website, under "Infrastructure."
5.3 Seek to provide the Australian Bioinformatics community with access to relevant research databases	See above. During consultative workshops, no requests for other databases were voiced.
5.4 Encourage Australian bioinformatics service providers to deliver appropriate products to end-users	Interaction between users and providers occurred during the consultative workshops and conferences, where bioinformatics service providers were present and contributed in the discussions.
6. COORDINATION	
6.1 Develop a useful and relevant ABN website	A comprehensive website which functions as both a portal of information and tool for collaboration has been developed.

6.2 Organise and run an annual Australian Bioinformatics Conference	Bioinformatics Australia 2006 and 2007 were successfully run alongside AusBiotech 2006 and 2007. It is proposed that BA 2008 is merged with a Asia/Pacific regional conference, efforts for the organisation of which are in planning.
6.3 Travel support for Conference Attendees	Funding provisions were made to bring international experts to give keynote lectures at both conferences. Travel scholarships were awarded to a number of students.
6.4 Coordinate, publicize and sponsor national and state networking events, and business meetings to foster the collaboration and development of bioinformatics business partnerships	Events are publicised in the AusBiotech website, BA mailing list, online discussion forum. A series of discussion fora and annual conferences were organised, where opportunities for interaction were offered.
6.5 Broaden current Bioinformatics Australia network	The networking events have brought new players to the community. The New Zealand community has also expressed keen interest in participating.
6.6 Publish information relevant to the Australian Bioinformatics community in relevant journals, magazines, and other media including opinions, editorials, letters, and broad-based research articles	Information is published in the Australian BioTechnology quarterly journals (circulation size 3000), articles on Australian bioinformatics was published in the Asia-Pacific Biotech News, members of the bioinformatics community have been interviewed by a journalist from The Age, and a career profile in bioinformatics along a video clip will be published at the ABC website: http://www.abc.net.au/acedayjobs/

CHAPTER 3: DETAILS OF MAIN ACHIEVEMENTS

3.1 Bioinformatics Stakeholders Consultative Workshops

A series of workshops were held in the following capital cities: Adelaide, Brisbane, Canberra, Melbourne, Perth and Sydney, to engage with the community, promote the networking facilities and website as an information portal, and to incite discussions on various issues concerning bioinformatics in Australia. Invitations were sent by email to more than 300 names in the Australian Bioinformatics Network database, comprising individuals from student/early career professionals to key-players in the community. No workshop were held in Tasmania or Northern Territory due to the lack of critical mass, however members of the community in Tasmania have been consulted prior to the events.

A total of fifty delegates partipated in the workshops. In addition, a number of people who could not attend submitted their input via email prior to the workshops and were included in the discussions,totalling to up to seventy participants in-person and online. Furthermore, the outcomes of each session were posted in the Australian Bioinformatics Network online forum, allowing any member of the community to participate in a continuing discussion.

Discussions were held to cover the following topics: current and future infrastructure needs, research grants, bioinformatics job-market and how bioinformatics training meets the needs of the market, and how bioinformatics as an enabling technology fits into the bigger picture of life- and medical-sciences, and biotechnology industry. Detailed agenda and notes can be found in Appendix C.

The main findings are as follows:
- that overwhelmingly the main infrastructure need is in human infrastructure, namingly expertise in analysis of large sets of biological data, who are able to understand the biological questions and wield appropriate bioinformatics tools to analyse the data. This leads to a need in professional development for the current and prospective users of bioinformatics.
- that the career path in bioinformatics is unclear, especially in the mid-career stage.
- that an appropriate training constitutes a master degree course, after an undergraduate qualifications in one of the core disciplines.
- that a unified curriculum in bioinformatics is contrary to the demands of employers.

3.1.1. Infrastructure

Infrastructure needs were explored in three different categories: the two more traditional classification of hard infrastructure (computing machinery and network cabling), soft infrastructure (software, databases, networking protocols) and human infrastructure (expert staff). The main issue with infrastructure needs is the availability of human infrastructure.

The community felt that the main driving phenomenon for infrastructure needs is the explosion of biological data, which gives rise to the needs for storage and interpretation. They expressed that the needs for storage is lesser of an issue than that of skilled people trained to analyse and interpret the fast-growing dataset, namingly human infrastructure.

The community believes that the issue surrounding the lack of human infrastructure outweighs that of computer hardware, software or databases. The availability of storage and high-performance computing at the various Partnerships in Advanced Computing consortia is well known. However, it is only accessible to few, that is members of the community with high-level expertise in computing - a distinct minority in the bioinformatics community. For example, high performance computing typically

has different hardware architecture and different operating system and set of software packages to the everyday desktop. This is also acknowledged by APAC, expressing that the National Facilities are under-utilised by the bioinformatics community. Therefore the availability of skilled support staff are crucial to ensure available computing infrastructure is well utilised.

The need for skilled people, with expertise in computing and good appreciation of the biological context is necessary to take the data and questions from the biological laboratories, and either use an array of existing tools or develop a new algorithm to analyse the data. There is also a need for the development of middle-ware, to connect the strong available computing power to more users, for example to enable storage and easy (e.g. web-based) remote access.

The need for human interface brings about the issue of education and training. This is discussed further below.

High performance computing are provided by the various partnerships for advanced computing, notably APAC and SAPAC for the availability of expert support staff. These support team aid users "translate" their programme to suit the heavy-duty computing machinery. A new facility, the Queensland Facilities for Advanced Bioinformatics (QFAB) also provides high performance computing facilities, as well as software and expertise which is specially catered for bioinformatics.

Issues with hard infrastructure unrelated to the advanced computing consortia above concerns the replacement of aging computing machinery. Representatives of the research community which develop bioinformatics algorithm expressed the needs for non-human infrastructure, such as collaborative tool software, hardware and fast network. These forms of infrastructure are currently available, but aging, hence rate-limiting. Furthermore, they are expected to be obsolete in the next few years. It was remarked as well that in the area of algorithm development research, algorithm speed is a performance-measuring factor, which is affected by the capability of the hardware at-hand. The need to replace hard infrastructure brings about the issues of cost and management, particularly that there is only very limited amount of grant available from research funding bodies for infrastructure.

A great number of bioinformatics software packages are freely downloadable from a number of websites, a list of some of which is given below. A good number of them, however, may need a level of IT expertise to wield. This is an area where facilities such are ANGIS are popular, as the web-based BioManager package is web-based and relatively easy to use, with good documentations and tutorials, and where training courses are available. SAPAC also has specialised support staff for the R statistical programming language commonly used for the analysis of gene expression/microarray data.

Name	Web address	Facility focus
Bioconductor	http://www.bioconductor.org/	mostly microarray data analysis
NCBI	http://www.ncbi.nlm.nih.gov/Ftp/	databases and sequence analysis
EBI	http://www.ebi.ac.uk/	various bioinformatics applications

The availability of human infrastructure is the real area of need, and it is expected that this need is on the rise, due to the continual drop of the cost of generating data and consequently rise of data size.

The data-handling need can be met in a number of ways:
- by provision of professional development courses, to train existing staff appropriate informatics (IT and/or statistics) skills
- by provision of suitably trained work-force

These are education and training issues, which will be discussed further in subsection 3.1.3.

3.1.2 Research and Development

Bioinformatics research discussed here encompass both research in developing new bioinformatics methods and research where significant amount of bioinformatics is used to analyse life-science or health-care data. Both research communities claim that bioinformatics is more poorly funded than average, and that due to its cross-disciplinary nature, it is hard for reviewers in both biological and information sciences to judge the value of the contribution in bioinformatics research. These claims do not correspond the points of view of the funding bodies and as it stands it is hard to ascertain the statistics.

The research community believe that bioinformatics research and bioinformatics in research are not very appropriately funded. In the former case, it is believed that since bioinformatics straddles life and mathematical/information sciences, each of which has a different tradition of valuing research outcomes, the grant applications are not appropriately judged. Since bioinformatics is still a relatively new field, the majority of reviewers are likely to come from one of the disciplines encompassing bioinformatics, therefore it is very possible that the proposed research outcome is undervalued. In 2007, there have been a number of success stories where large grants are given to centres or groups, but not to individual researchers (particularly for mid-career researchers), which would have significant impact on their career progression. This has been evidenced by an exodus of researchers to non-bioinformatics careers.

With regard to bioinformatics components in life science research, it is believed that the expertise in data management and analysis in undervalued, that there is a perception that bioinformatics analysis can be performed by simply using freely-available web-based tools, which can lead to to inappropriate analysis and inaccurate conclusions. There are also issues on the level at which bioinformatics support should be provided, as one research project does not necessarily require one full-time bioinformatician. It is believed that a shared bioinformatics resource at the departmental or university level is ideal, however there are no framework to fund this human infrastructure need.

Conversations with the funding bodies (ARC and NH & MRC) reveals the converse view. The ARC, for example, believes that they understand very well the existence and growth of multi-disciplinary research subjects, and that all due efforts have been made to ensure that grant proposals are appropriately allocated to suitable reviewers. There is currently no means to ascertain the success rate of bioinformatics in comparison to the average grant success rate, as new RFCD codes (see Section 4.2) will not take effect until 2008, and there is no effective means to compile the success-rate statistics.

3.1.3 Education and training

The discussion on education and training was held in light of the perceived need of human infrastructure, i.e., how to train people to meet this need. Two forms of training were considered: university qualifications in bioinformatics, and professional development/short courses for life- or health-scientists. Also discussed was the observation of how this need in skilled work-force has not been met by a demand in the job market.

Professional development courses were raised in particularly by members of the community who either come from the life sciences background, or work in collaboration with life scientists. It was felt that better awareness and understanding of available tools (and their limitations) would serve as valuable skills to and were desired by established (often mid-career) medical- or life-scientists.

However, low economic demand seem to follow this need. Existing bioinformatics courses are mostly under-subscribed and some have been closed. This scenario is common across the academic and industrial training sectors. Bioinformatics undergraduate courses at Flinders and Sydney Universities have been cancelled. Discussions with Box Hill Institute, which approach biotechnology companies and design customised professional development packages, revealed that this situation also applies to them. Half- to one-day courses in bioinformatics skills have been undersubscribed, and bioinformatics components in the courses they designed for the industry have been minimal.

The only success story in bioinformatics professional development seems to be Sydney Bioinformatics, whose courses have been over-subscribed, with long waiting lists, as outlined in the table below. This prompts for further exploration of their role as a provider of professional development.

Course	Length	Date	Location	Enrolment	Capacity	Wait list
Custom	1 day	May	Sydney	20	N/A	N/A
Intro. to Microarray analysis	1 day	June	Sydney	20	20	45
Applied Bioinformatics	3 day	July	Melbourne	17	17	5
Applied Bioinformatics	3 day	July	Sydney	20	20	5
Applied Bioinformatics	3 day	July	Brisbane	14	16	0
Introduction to Bioinformatics	1 day	July	Brisbane	9	16	0
Proteomic Bioinformatics	1 day	Aug	Sydney	20	20	20
Intro. to Microarray analysis	1 day	Oct	Sydney	20	20	35
Custom	2 day	Oct	Townsville	15	N/A	N/A
Phylogenetics	1 day	Nov	Sydney	TBC	20	TBC
Custom	1 day	Dec	Canberra	TBC	N/A	TBC

Bioinformatics is offered as university courses in a number of ways. At the undergraduate level, there are dedicated bioinformatics degree courses and bioinformatics as a stream in other courses (usually biotechnology, science, IT or mathematics). There are also graduate certificates, graduate diplomas and master degrees in bioinformatics.

There were criticisms against bachelor degree courses in bioinformatics, in particular dedicated degree courses. Some consider it to spread the subject matters too thin, producing graduates with no solid background in anything. Another criticism relates to the quiet job market for skilled bioinformaticians.

Contrary to the findings in the National Bioinformatics Strategy, the projected skill demand has not been reflected in the job market. The expected number of vacancies of 370 in the past two years foreshadowed in the NBS did not eventuate. For example, in the past six months, bioinformatics job advertisements number around ten, with employers coming from the public sector (academic, research, infrastructure provider). This number is reflective of the state of the job market in the past two years. The projected uptake of bioinformatics skills in the biotechnology industry has not been manifested itself for most of the time period between 2002 and 2007.

However, this grey sky is projected to lift, with the rise of vacancies in the last two months of 2007. Over forty openings were advertised. They are mostly in the public sector, and mostly positions where bioinformatics is embedded, in support of other sectors (such as bioscience research), with a small handful from private enterprises. This suggest that the bioinformatics is starting to be harnessed as an enabling technology, at least by the research community. Further growth can be projected by awareness-rising activities to the private sector, particularly the biotechnology industry.

In terms of suitable training for the existing job market, a master degree course is considered to be the appropriate level at which to train bioinformatics skills, where students are expected to have completed a bachelor degree in either mathematics/statistics/IT or biological sciences. This framework would allow students to develop a strong background in at least one of the core disciplines. Due to the diverse backgrounds of prospective students, a master degree course needs extensive course advise to tailor a curriculum which suitably fills the gaps in the students' skills and knowledge. An example of such a course is offered at the University of Sydney.

At the undergraduate level, due to the uncertain employability and career path for graduates, a stream in bioinformatics within a related degree course is considered to better serve the students rather than a devoted degree course in bioinformatics, so graduates have other avenues in which to seek employment. Of the array of courses with bioinformatics components or streams, those with greater emphasis in informatics (such as computer science, software engineering, or mathematics courses with subjects in life sciences) are preferred by bioinformatics academics, of informatics and life sciences backgrounds alike.

The community is in general not in support of establishment of a common curriculum in bioinformatics. It was felt that firstly, more effort into establishing a career path for graduates would be of greater value. This can be achieved via a vibrant biotechnology industry driving demand. Furthermore, employers of bioinformatics-skilled work-force believe that the variety of courses, hence variety in graduates and potential applicants to job openings, enriches the work-force. Employers prefer to have the choice rather than uniformity in applicants.

It is believed, however, that the demand for bioinformatics-skilled work-force is in the horizon. The decrease in cost of data generation (such as genome sequencing) is expected to stimulate the surge in skill demand, in areas of applications such as personalised medicine and agriculture, just to name a few. This is a projection for the five to ten year time-frame, which is in fact the correct time-frame for bioinformatics training.

3.1.4. Commercialisation

The Australian bioinformatics community is largely an academic/reseach community, along with infrastructure providers who mostly serve the needs of bioinformatics research. There is currently a low level of engagement with the commercial sector. This causes the community to depend to a great extent to public funding (such as research and infrastructure grants).

There is potential for greater economic activity and therefore high employment rate and more established career paths for graduates, which has not yet been pursued. It was suggested that biotechnology would indeed benefit from a work-force trained in bioinformatics, however the biotechnology industry still has low awareness of what bioinformatics has to offer. This calls for awareness-raising activities, such as intermediary services to biotechnology organisations on how bioinformatics can contribute to their productivity.

There is also potential in partnerships between research groups and companies in joint-projects which could result in a commercial product. The existing ARC linkage-grants are considered unattractive by interested companies, which are mostly SME's, for requiring a significant financial investment towards endeavours with still uncertain outcomes. It is recommended instead that incentives in the form of seed-funding be made available, to motivate more engagement between the research and commercial communities.

3.1.5. Data Management

The issues of consistent data formats are of greater concerns to the medical community, where there are efforts towards integrating large medical databases. Major efforts to consolidate inconsistent data formats have commenced, spearheaded by the Western Australia Institute of Medical Research funded by an NH & MRC infrastructure grant.

Discussions are still underway on initiatives towards data standards within the Australian National Data Service under the umbrella of NCRIS 5.16, Platforms of Collaboration. These workshops aim to address policies and technologies around data access and discovery, storage and management including:
- Policy/Standards/protocols development
- Technology/tools development underpinning services/processes/workflow
- Trusted repositories

Further details can be found in the Platform for Collaboration website:
http://www.pfc.org.au/twiki/bin/view/Main/Data

3.1.6. Regional variations in the responses

The Victorian community most clearly voiced the need for human infrastructure, along with the desired training for a praciticing bioinformatician. The community shows strength in medical research, with delegates from the Walter and Eliza Hall Institute, Peter MacCallum Cancer Research Institute, NeuroSciences Victoria, as well as biosciences, represented by Bio21 Institute, the University of Melbourne and Monash University. A typical arrangement of bioinformatics researcher-user interface is by collaboration. The PeterMac is equipped by a bioinformatics unit, which is still understaffed and overworked, but openings are still yet to be made. This is a gathering which shows a large bioinformatics user:bioinformatics expert ratio, giving the first indication of the need to raise awareness of the role of bioinformatics in downstream disciplines.

The New South Wales community claims success in education, with representation from the University of New South Wales (with their Bachelor of Software Engineering majoring in bioinformatics course), the University of Sydney (with their customised curriculum Master in Bioinformatics course) and Sydney Bioinformatics (with their over-subscribed professional development short courses). There is more representation of bioinformatics researchers, from the above universities as well as CSIRO Mathematical and Information Sciences division, driving the discussions towards the need for fast computing facilities for effective tool development. These hardware issues have been relayed to and responded by APAC by means of the online discussion forum and face-to-face discussions in Canberra.

The community in Canberra constitutes a strong pure-research culture and support infrastructure, with representations from the Australian National University, CSIRO and APAC/ANU Supercomputing Facilities. Grant reviewing process was a main concern in this workshop. This should be addressed with the appointment of a bioinformatician to the ARC College of Experts. APAC/ANUSF is addressing the gap in infrastructure provision to the middle-users, by the development of a web-based middleware.

The Queensland community has a strong research background in the biosciences as well as bioinformatics tool development, with delegates from the ARC Centre of Excellence for Bioinformatics, CSIRO and QUT. Two delegates travelled to Brisbane from James Cook University at Townsville, which showed the strong desire to be involved in the community building activities. Infrastructure

providers were represented by QFAB. Hardware infrastructure is also an issue here, but the emphasis is on the replacement of aging equipment, at which level to manage this, rather than access to centralised, national facilities. Emphron Informatics was also consulted separately, where the concern of economic sustainability due to lack of engagement with the commercial sector was raised.

The Western Australian community was strongly represented by medical research, with the Western Australian Institute of Medical Research and the Telethon Institute of Children's Health. The need for bioinformatics expertise is met with a demand in the job market here, addressing the stretched capacity of statistical/IT support staff in the medical research institutes. Data management was discussed at length, as well as the expressed need for hardware infrastructure with the appropriate software for statistical analysis.

The South Australian community presents strengths in agricultural bioinformatics and statistical analysis of gene expression data, with delegation from the University of Adelaide and Australian Centre for Plant Functional Genomics. Infrastructure is very well provided for by SAPAC, which has a team of support staff with expertise in the statistical software for gene-expression data-analysis. Proposals of bioinformatics courses were discussed here, unhindered by the current low level of unemployment, as there is an expectation of uptake from industries such as personal medicine and agriculture in five to ten years' time, resulting from the continuing drop in sequencing costs.

3.2 Bioinformatics Australia annual conferences

The success of the inaugural Bioinformatics Australia 2006 conference serves as a platform to the second annual Bioinformatics Australia 2007 conference. Both conferences are held in conjunction with the AusBiotech National Conference, in Sydney and Brisbane in the years 2006 and 2007 respectively. This presents opportunities for bioinformaticians to interact with users of the technology in the biotechnology sector.

The first conference was attended by delegates from all over Australia. The second conference had in addition involvement from the New Zealand bioinformatics community; in the programme committee, in financial sponsorship and in the invited speaker list, as well as contributed papers and posters, which resulted from promotion of the Australian Bioinformatics Network to the New Zealand community (further on this under Regional and International Development section, subsection 3.4.3).

The Bioinformatics Australia 2007 (BA 2007) conference grew from its inaugural predecessor in the number of countries represented and in the number of sponsors. As well as the New Zealand sponsorship and delegation, BA 2007 boasts delegation from the USA, South Africa, France and Finland (a delegation from which was also a sponsor). The number of sponsoring institutions grew from eight to twelve. These further indicate a growth in interest and the clear potential of an event which is gaining reknown and reputation. The next step, of a joint event with an established international conference, is only logical in further establishment in the region and wider promotion of bioinformatics in Australia.

The conferences have also served as a forum for the community to meet and discuss issues pertaining to networking and community-building, as well as a scientific meeting. Discussion sessions were held on NCRIS (in which the following key-players have chaired: Prof Matthew Bellgard, convenor of the bioinformatics platform of capability 5.1, Dr Rhys Francis, convenor of capability 5.16 and Andrew Gilbert, Manager of BioPlatform Australia, realisation of capability 5.1) and the Australian Bioinformatics Network. These discussions proved to be very productive and popular, with delegates coming to mainly participate in them.

The scientific programme includes international key-note speakers, national invited speakers, contributed papers and posters. It provides a medium for scientific discussions and networking, and particularly an opportunity for the local community to interact with international experts. These opportunities were particularly appreciated by early-career bioinformaticians and newly established groups, making the event an incubator for a growing community.

3.2.1 Bioinformatics Australia 2006 Conference

A more comprehensive reporting of Bioinformatics Australia 2006 was presented in the Progress Report submitted on 30 November 2006. A copy of this comprehensive coverage can be found in Appendix E.

It suffices to say in this report that the inaugural conference was a great success, attended by 120 delegates which include key-players in the Australian bioinformatics community and sponsored by the NSW government Office of Science and Medical Research, CSIRO, National ICT Australia, Australian Genome Research Facilities, APAF, the Ramaciotti Centre, Supamac and the Garvan Institute.

3.2.2 Bioinformatics Australia 2007 Conference

Following the success of the inaugural conference, Bioinformatics Australia 2007 was held in Brisbane on 23-24 October 2007, also in conjuction with the AusBiotech 2007 National Conference. It was attended by 109 delegates, from Australia, New Zealand, Korea, USA, South Africa, France and Finland. It was sponsored by the ARC Centre of Excellence in Bioinformatics, BioPlatforms Australia, the Allen Wilson Centre for Molecular Ecology (New Zealand), QFAB, Sydney Bioinformatics, Medicel (Finland), the Walter and Eliza Hall Institute for Medical Research, the Australian Mathematical Sciences Institute, the Garvan Institute, and Cambridge University Press. As in the previous year's conference, it also include a scientific programme, discussion sessions, and opportunities to interact with the biotechnology community during the common refreshment breaks held in the Exhibition Hall.

Delegate exit survey feedback indicated that the conference was deemed to be of good quality, both on the scientific programme and organisation, and most respondees indicated that they would be very likely to attend future Bioinformatics Australia conferences. Many delegates enjoyed the networking opportunities that the conference offered, including the discussion sessions in the establishment of a community.

Scientific programme

The programme comprise two plenaries, a poster session and five scientific symposia in: Computational Biology, Methods and Tools in Bioinformatics, Proteomics and Structural Bioinformatics, Biological System and Comparative Genomics and Evolution. Two experts from Australia and New Zealand are invited to present at each symposium and two oral presentations are selected from submitted abstracts. Whenever possible, priority is given to abstracts by students and early-career researchers.

Two plenary lecturers were invited: Dr Michael Hucka from the California Polytechnique and Prof Win Hide of the South African National Bioinformatics Institute (SANBI). Dr Hucka is one of the principal developer of the System Biology Mark-up Language (SBML), which facilitates researchers to share and build on each other's work more effectively and directly. Prof Hide is the founder and Director of SANBI, and visiting fellow at the Harvard School of Public Health, whose main research interest include the bioinformatics in tropical diseases (such as malaria and HIV) and cancer gene discovery.

Poster prizes were awarded to encourage even better submissions in future conferences. One first and two runner-up prizes, of $300 and $100 each respectively, were awarded to Mr John Hawkins (PhD student, University of Queensland), Dr Bruno Gaëta (lecturer, University of New South Wales) and Mr Ashley Waardenberg (PhD student, CSIRO).

Delegates expressed their satisfaction with the scientific programme in the exit survey, rating each component (overall scientific programme, keynote lectures, contributed papers, poster exhibition and abstract book) an average of 2.75-2.94 out of 4.

Discussion sessions

Two discussion sessions are organised; one on NCRIS Capability 5.1 and one on the Australian Bioinformatics Network. Prof Matthew Bellgard gave a presentation to lead discussions on NCRIS 5.1, with Mr Andrew Gilbert (General Manager of BioPlatform Australia), Dr Annette McGrath (Bioinformatics Manager, AGRF/Genomics Australia), Dr Vladimir Likic (Bioinformatics Manager of Bio21, in charge of bioinformatics developments for Metabolomics Australia) and Dr Margaret Kahn (Scientific Consultant, APAC/ANU Supercomputing Facilities, a major player in NCRIS 5.16) present in the audience.

Prof Bellgard presented the structure of BioPlatform Australia, how the three -omics platforms (Genomics Australia, Proteomics Australia and Metabolomics Australia) relate to the Australian Bioinformatics Facilities (ABF), how the facilities are managed to serve the needs of the -omics communities, and the governance of the ABF. There were questions on whether non -omics will be catered for, whether NCRIS 5.1 would support training to address the shortage of human infrastructure and what the mechanism would be to obtain feedback from the community. It was explained that at this early stage it was deemed most strategic to be focused and only address the needs of the -omics platforms included in NCRIS 5.1, but at a later stage, other issues (such as training) and the needs of other areas (non -omics) can be reviewed. For example, tools will be developed to address the needs of the -omics community, but these tools will be made available to the community at large. The feedback mechanism and evaluation of BPA was outlined, that these will be performed by the Bioinformatics Management Group of the ABF and BPA, with representatives from the three -omics platforms. There were suggestions on plans of interactions with bioinformatics researchers to take advantage of the novel technologies and encouraging communication with users to access the available services, which was taken on board by Prof Bellgard.

Further discussions were conducted with Mr Andrew Gilbert on links between BioPlatform Australia and Bioinformatics Australia. BPA expressed interest in seeking affiliates to use its capacity, and to build collaborative tools. They are particularly interested in the ABN community directory, and

proposed possibilities of joint efforts in furthering the directory and establishing an Australian bioinformatics toolbox registry.

The discussion on ABN updates is led by Dr Lucia Santoso, Executive Officer of the Australian Bioinformatics Network and Prof Mark Ragan, vice-president of Bioinformatics Australia, in place of Dr Tim Littlejohn who was absent for health reasons. The findings of the Bioinformatics Stakeholder Consultative Workshop (see Section 3.1) were presented, outlining the needs for human infrastructure and professional development, which are contrasted with the lack of employment opportunities and defined career path.

Delegates from Finland and New Zealand contributed with experiences in their respective countries, including past experience in the UK. It was outlined that the situation in Finland was not much different, but the UK community had been more successful in creating employment opportunities by raising awareness of bioinformatics capabilities to the biotechnology industry. It was then emphasised that the nature of the industry would be different in Australia, where no drug-development activities occur. It was further suggested that universities approach the existing companies to establish internship programmes, aiming to both enriching students' learning experience, building networks between educational institutions and industry, and raising awareness of bioinformatics technologies and capabilities to existing industries.

It was also suggested that collaborations be sought with institutions where bioinformatics are better established and funded, such as Europe and the US, with the European Commission Framework Programme 7 and NIH grants, for example. It was recommended that joint-application with local (relative to funding body) researchers as principal investigators would be more strategic for success.

It was revealed that most graduates of bioinformatics undergraduate courses entered research higher-degree programmes, and it was discussed whether less research and more technical/industrial focus should be adopted in the courses. Representatives of two educational institutions with good success track in undergraduate bioinformatics education (the University of New South Wales and the University of Auckland) revealed the relatively low numbers in enrolment (between 10-20 in each university). It was agreed overall that a master degree course would be preferred as bioinformatics training. A feedback mechanism was suggested, to filter the required skills from job advertisements and provide input to curriculum development.

Conference attendees expressed that they have enjoyed the discussion fora, with a good number naming them to be the best aspect of the conference.

Networking

As in last year's conference, opportunities to network with the greater biotechnology community were offered in the BioIndustry Exhibition, where the refreshments were taken. Delegate responses from this year's exit survey indicated, however, that this opportunity was not fully as in last year's event.

Student support

Travel sponsorships are available for students, sponsored by the Australian Bioinformatics Network grant, the Walter and Eliza Hall Institute and the Australian Mathematical Sciences Institute. AusBiotech student members are eligible to apply for the AusBiotech travel scholarship, the selection of which is based upon the quality of submitted abstracts. Students and early career researchers from AMSI member institutions are eligible to apply for travel sponsorship through their departments. Some of the recipients of the scholarships are the following:

Name	Institution	Scholarship
Mr Bryan Essien	University of Southern Queensland	AusBiotech
Ms Chinfoon Khoo	Australian National University	AMSI
Mr Leo McHugh	University of Sydney	AusBiotech
Ms Armita Zarnegar	University of Ballarat	AusBiotech

Recipients of the AusBiotech travel scholarship.
L-R: Dr Rohan Teasdale (chair of BA 2007 conference), Bryan Essien, Leo McHugh, Armita Zarnegar, Dr Lucia Santoso

3.2.3 Future conference plans

The organisation of the Genome Informatics Workshop 2008 conference has been offered to Australia, and preparations have started towards its organisation. The Genome Informatics Workshop is the longest-running conference in bioinformatics (the first one being in 1990) and has been held in Japan until this year, when it start being a travelling conference.

This is further discussed under "Regional and international development."

3.3 Australian Bioinformatics Network website

http://www.ausbiotech.org/bioinformatics

The ABN website plays a pivotal role in the establishment of a community whose members are geographically scattered. It serves the purposes of both being an information portal, first point-of-call, to the community and general public by promoting Australian bioinformatics facilities, activities and various events and programmes, and facilitating networking and collaboration within the community.

Information on bioinformatics in Australia, particularly on infrastructure, research institutes, educational programmes, jobs and events are available in the website.

- The welcome page gives an outline of Bioinformatics Australia, including its aims and objectives.
- Infrastructure facilities are listed in categories of database mirrors and software facilities, research facilities and high performance computing facilities.
- The research and development page contains a list of research groups in Australia and New Zealand are posted, outlining the areas of research, location and contact person, aiming to facilitate research collaborations. It also would contain notices on available research grants, and briefs of successful grants.

- A list of educational programmes can be found in the Education and Research page, along with a list of skill-set sought by employers.
- Listing of events and jobs can also be found on the website. Eight bioinformatics jobs have been advertised since July. Events webpages advertise all bioinformatics events, both nationally and internationally, where facilities to register online are available for AusBiotech-organised events.

A community directory and online discussion board are developed as collaborative tools to aid communication and discussions across the distance.
- The directory is powered by a database of community members, which is searchable by name, expertise and location. It is aimed to fulfill a "find-a-collaborator" function for the community, or a "find-an-expert" function for the general public.
- The online forum enables discussions to continue after face-to-face discussion opportunities (such as the consultative workshops and conferences). It was launched with discussions on RFCD codes and continued with a series of follow-up reporting and discussions after the consultative workshops.

The online discussion forum has proven to be a popular means to communicate, with a steadily rising number of participants and discussion threads.
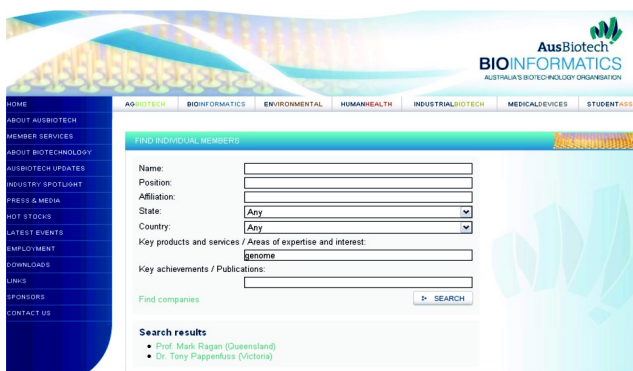

(a)
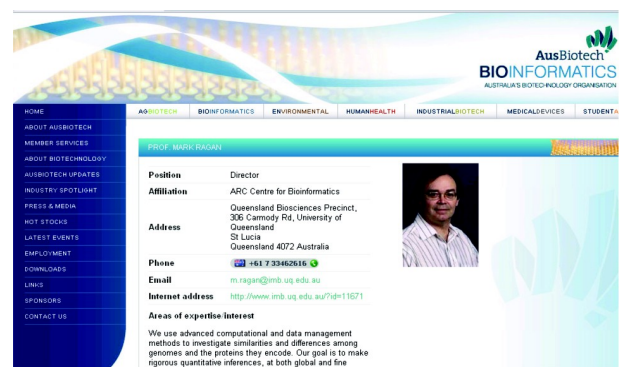
(b)
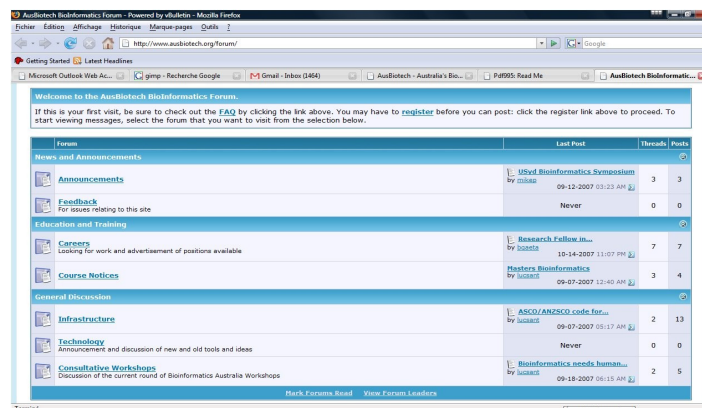
(c)

Figure 1: (a) Australian Bioinformatics Network database search functions – experts can be searched under name, organisation, location and area of expertise. (b) Example of a search result in the database. (c) Front page of the Australian Bioinformatics Network online discussion forum.

3.4 Networking and Coordination

3.4.1 Networking between members of the bioinformatics community

Networking and communication occurred both in person and online. Three opportunities to meet face-to-face and discuss issues pertaining to community building occurred: during Bioinformatics Australia 2006 and 2007 conferences and during the Consultative Workshops.

The online discussion forum served the purpose of conducting discussions and information dissemination. For example, an intense discussion occurred in the process of establishing the RFCD classifications for bioinformatics, where in the start the community was divided as to whether an RFCD code for bioinformatics should sit in the mathematical sciences, information sciences or biological sciences. A resolution and favourable outcome was achieved in the end. More detail on this process is outlined in the section on Establisment of RCFD Classifications for Bioinformatics.

The Australian Bioinformatics Network community directory is another means to facilitate communication and collaboration. This directory is powered by a database, searchable by name,

3.4.2 Networking between the bioinformatics community and related industries

Communication has been established with the two named organisations: National ICT Australia and Health Systems Research Corporation, as well as with two large ICT organisations: IBM and CRAY computers. A relationships has been established also with a consortium in the space of life-science and health data integration, Molecular
Medicine Informatics Model (MMIM), which operate under the Bio21 Consortium based in Melbourne.

National ICT Australia has moved to the Life Sciences space, with the establishment of the ICT for Life-Sciences programme in the NICTA Victorian Research Laboratory, which has a bioinformatics group. They are active members of the Australian Bioinformatics Network, with participation in the form of sponsorship and provision of a speaker at Bioinformatics Australia conferences.

The activities of Health Systems Research Corporation is mostly in the space of Health Informatics; in integration of clinical information and capabilities of its remote access. Collaborative projects are conducted with MMIM and IBM. There is little overlap with bioinformatics activities, however it is a potential to be pursued in the future.

Health System Research Corporation is an informal alliance between IBM, Intel and the e-Health research unit at Monash university, headed by Michael Georgeff. Their main activity lies in integration of health databases, with the purpose of connecting the healthcare systems throughout the continuum of care to bring better healthcare
outcome. Their connection with bioinformatics is through their involvement in the Molecular Medicine Informatics Model (MMIM) programme in the Bio21 consortium.

Communication has been established with MMIM programme director, and discussed their involvement in biological informatics. Their role is mainly on project and data coordination, and thus falls in the category of infrastructure provider. Their information can be found in the ABN website.

Conversations with both IBM and CRAY computers revealed that the market for high-performance computing amongst the bioinformatics community is at present immature. Both firms are reducing their marketing efforts in the life-science research domain, devoting more energy to the healthcare space.

Similar opinions also of the software market, low demand of commercial life-science tool software in Australia, culture of free-software, brings about the rise in demand of human infrastructure to wield the free software packages to a pipeline more capable to attend to the informatics needs of wet-laboratory scientists.

3.4.3 Regional and international development

Relationships have been established with the New Zealand and Asian communities, evidenced by a growing New Zealand participation in the Bioinformatics Australia conference and Australian Bioinformatics Network community directory, and the proposed hosting of the Genome Informatics Workshop 2008 conference by Bioinformatics Australia.

There has been an existing relationship with the New Zealand community, namingly between the ARC Centre of Excellence for Bioinformatics and the Bioinformatics Institute based at the University of Auckland. This trans-Tasman network is further broaden by an awareness-raising visit to the New Zealand community. It was revealed that there is strong bioinformatics capability in New Zealand, with a particular strength in evolutionary genomics, and there is a desire for some form of coordination, however the size of the community is too small to warrant a separate society. They are then included in Bioinformatics Australia activities, for example, NZ research groups and courses in bioinformatics are listed on ABN website, members of the NZ bioinformatics community included in the ABN community directory, and the Bioinformatics Australia 2007 conference included an invited speaker, a member of the programme committee, a number of posters and a sponsor from New Zealand. They actively participated in the discussions and further expressed a keen interest in being included in the ABN.

Bioinformatics Australia (past president Dr Jonathan Arthur and present vice-president Prof Mark Ragan) participated in the board meeting of the Austrasian Association for Societies in Bioinformatics (AASBi) in Japan, during the Genome Informatics Workshop 2006 conference. The Australian presence was well received, resulting in the proposal that Australia holds the secretariat of AASBi in 2008, which will involve the hosting of the flagship conference Genome Informatics Workshop 2008.

Bioinformatics Australia has submitted a proposal for an affiliation with the International Society for Computational Biology (ISCB), which would leverage its profile in the international arena, with the provision of webpage at the ISCB website and the authorisation to use ISCB logo on its website and advertising material, and the advertising of the conference and other events at the ISCB newsletters. A decision from ISCB board is expected to be made within the next twelve months.

3.4.4. Representation of the community to relevant agencies

Bioinformatics Australia under the Australian Bioinformatics Network project has liaised with agencies such as the Australian Research Council (ARC), the National Health and Medical Research Council and the Australian Bureau of Statistics, on behalf of the community.

The community voiced a concern that bioinformatics is under-funded, relative to other disciplines, and that grant proposals are not appropriately reviewed. Statistics on success rate was sought from the NH & MRC and advice was sought from the ARC. Discussions were conducted with Executive Directors of the ARC on the processing of grant proposals for bioinformatics and counsel was sought in effective grant-preparation techniques. It was been advised that the factors that truly underpin the success rates

of grant proposals are the presence of community representatives in the ARC reviewing panels, namingly the College of Experts and the list of ``OzReaders,'' who directly assist the Executive Directors with the reviewing process. As it stands there are no
bioinformaticians in either panels. It followed that two members of the bioinformatics community: Prof Mark Ragan and Prof Geoff McLachlan, both of the University of Queensland, have been nominated for membership in the College of Experts, and Assoc Prof Wise has been recommended to the ARC to be an OzReader. Prof McLachlan has been appointed to the College of Experts, taking effect in 2008.

The bioinformatics community through the ABN was actively involved in the establishment of a family of Research Fields, Courses and Discipline codes by the Australian Bureau of Statistics. Active discussion and consultation process occurred between the bioinformatics community and the ABS, with the ABN acting as a facilitator.

3.4.5 Liaison with key National Collaborative Research Infrastructure Strategy stakeholders

Discussions occurred at a semi-regular basis between Bioinformatics Australia, through the ABN project, and key-stakeholders of NCRIS. The first discussion was held at the Bioinformatics Australia 2006 conference, with presentations by Prof Matthew Bellgard, convenor of the bioinformatics platform in NCRIS 5.1 and Dr Rhys Francis, convenor of NCRIS 5.16. A series of discussion followed with Dr Sue Forrest of the Australian Genome Research Facilities (Genomics Australia), Prof Mark Baker and Mr Brett Cooke of the Australian Proteomics Analysis Facilities (Proteomics Australia), Dr Vladimir Likic of Bio21/University of Melbourne (Metabolomics Australia) and Mr Andrew Gilbert, manager of BioPlatforms Australia. The community was further briefed during the Bioinformatics Australia 2007 conference.

BioPlatform Australia expressed interests in collaborative efforts in the establishment of internet networking tools, such as the ABN directory, and possibly on a repository of bioinformatics tools.

3.4.6 Coordination of education and training

Feedback was gathered from representatives of the community present at the consultative workshops. These participant comprise representatives from academia, research community, research support, and students.

The outcome of a good bioinformatics education is competence in the required computation tools and statistical techniques, with a good understanding of the biological questions and limitations that the experimental procedures or equipments place on the data. This is achieved by having a strong grounding on both information sciences
and biological sciences.

The undergraduate degree programme was deemed not suitable to start a focused training in bioinformatics. There were criticisms that the breadth offered in a bachelor degree in bioinformatics does not provide its graduates with the depth required to be able to thoroughly perform the data analysis with the available tools, and
develop new tools where required.

Undergraduate degree courses leading to a master in bioinformatics should be those which gives a solid background in the core disciplines underpinning bioinformatics practice. There was an expressed opinion, which is up to debate, there should be more emphasis in the mathematical and information sciences in the core disciplines. Undergraduate students pursuing a career in bioinformatics should take biological subjects, to give them an appreciation of the biological context and the processes that

generate the data. This is not by all means the total amount of required biological knowledge; this will come through engagement. The biological component in an
undergraduate degree leading to a career in bioinformatics should equip the graduate to communicate and collaborate with their biological peers, and adequately understand the biological questions.

It was proposed that a master degree course is a more appropriate level to train the vocation of bioinformatics, at which the two disciplines can be married and further hands-on training is given. The ideal situation would be exposure to the biological problems by immersion, where students with solid foundation in informatics can
be put in partnership with biological counterparts and exposed to real biological problems by collaboration.

There exists an online education consortium, the S* Alliance, http://www.s-star.org/, arising from the collaboration between universities from Australia (Macquarie and USyd), Sweden (Karolinska Institute and Uppsala), Singapore (NUS), South Africa (U Western Cape) and USA (Stanford and UCSD).

Further standardised curricula are not recommended, as it would lead to production of standardised graduates. It was established from the employers' perspective during the consultative workshops that variety in the work-force is more desirable, hence standardised curricula are not.

There is an expressed need for professional development, short courses in programming for biologists to be comfortable with the available tools, in biology
for IT professionals interested to enter the field. This need, however, has been followed by an economic demand, evidenced by the demise of companies such as BioLateral due to a market failure, and the low enrolments of bioinformatics short courses offered by Box Hill Institute. The latter also offers customised training packages for industry, bioinformatics components of which has also been quite low.

This situation does not seem to apply to Sydney Bioinformatics, whose courses have been either full or oversubscribed, as shown in the table in subsection 3.1.3. This illustrates that there is a need that ANGIS is fulfilling. Their potential role as a key deliverer of bioinformatics training should be further explored.

3.4.7 Promotion and awareness-raising

Currently awareness-raising of the existence and potential of the bioinformatics community are conducted by publications of articles in journals such as the Australasian BioTechnology and the Asia-Pacific Biotech News, in the AusBiotech website, and in liaison with the ABC in the production of a career information clip and profile for school-leavers, in the "Ace Day Jobs" project, http://www.abc.net.au/acedayjobs/

There could be potential for more active and directed awareness-raising activities, aimed towards the industries that bioinformatics enables, such at the biotechnology industry.

CHAPTER 4: DETAILED REPORTING OF SPECIFIC TASKS

4.1 Review of the Australian National Genomic Information Services (ANGIS)

Summary

The purpose of this review is essentially to compare the services provided by ANGIS against the infrastructure needs of the community, and consider possible roles beyond ANGIS' current ones in the provision of bioinformatics services and support. ANGIS was restructured in mid 2007, and now exists under the umbrella of Sydney Bioinformatics. In this light, we review the past and present services of ANGIS, but future recommendations are made in relation to the services and potential of Sydney Bioinformatics.

ANGIS underwent a series of crises between 2004 and 2005, due to lack of leadership and funding, compounded with the need to replace hardware and software. The University of Sydney carried out a review of its bioinformatics capabilities, particularly of ANGIS and the Sydney University Biological Informatics and Technology (SUBIT) in 2005. The review considered ANGIS and SUBIT to have unrealised potential in building bioinformatics strength in the university and found it strategic for the university to put further investment into bioinformatics. This leads to a merger between ANGIS and SUBIT and the formation of Sydney Bioinformatics.

Sydney Bioinformatics offers bioinformatics services/expertise and professional development/training for life scientists as well as the BioManager/ANGIS services. Anyone can access the infrastructure services provided by Sydney Bioinformatics. Rates vary, depending on whether users are members of the University of Sydney, non-profit or commercial organisations.

Feedback survey of ANGIS users indicates that ANGIS facilities, both in the form of software/databases and training, are still fulfilling a need. Users have been by and large satistfied with the services they receive. In addition, it is worth noting that ANGIS is the only provider of professional development in bioinformatics – raised in the series of consultative workshops as a need – who can claim success, in the form of good enrolment levels, and even waiting lists. Possibilities of expanding its professional development arm to further serve the needs of the community should be considered.

Introduction

ANGIS was established in 1990, as a successor and upgrade to a service then provided by the University of Sydney named Sydney University Sequence Analysis Interface (SUSAI). Its establishment was proposed by the Commonwealth Department of Industry, Technology and Commerce (DITAC)

ANGIS was established in 1990, as an outcome of the proposal of the Commonwealth Department of Industry, Technology and Commerce (DITAC). The University of Sydney won the proposal to develop Sydney University Sequence Analysis Interface (SUSAI) to

become a National Genomics Information Services, which then became ANGIS. On 5 August 1991, Australian Genomic Information Centre was formed by the Senate of the University, with the responsibility for managing ANGIS.

The objectives of ANGIS are to:
1. maintain sequence and other genome-related databases and programs, using an integrated system for retrieval, manipulation and comparison;
2. provide immediate access to this system and the associated databases and programs, via the internet and other means, seven days a week, 24 hours a day;
3. facilitate access to specialised local and overseas databases and programs which are not available within the ANGIS system;
4. develop easy-to-use menu-based interfaces while retaining the command-driven system, which is more responsive for experienced users;
5. provide documentation (e.g. manuals) and tutorials for beginners and experienced users;
6. provide help to users via e-mail and/or telephone;
7. provide assistance in training undergraduate and postgraduate students in genome data retrieval and analysis; and
8. facilitate the formation of user groups in major centres of genome research, and collaborate with these groups.

Upon its establishment, ANGIS was equipped through a donation from Sun Microsystems and a number of small grants.

Since 2000, ANGIS derived income from fee (subscription, courses and consulting) and grants. The summary can be found below:

	2000	2001	2002	2003	2004/2005
Fee	$ 399,491	$ 272,748	$ 344,320	$ 325,261	N/A
ARC (RIEF)	$ 250,000				
NH & MRC	$ 75,000	$ 75,000	$ 255,000	$ 75,000	$ 0

These figures are obtained from the University of Sydney internal reviewed, conducted in 2005, at which time the figures for the financial year 2004/2005 was not yet available.

ANGIS has operated at a significant net annual loss after the decline in grant income. The service is sustained in a limited way by accessing surplus funds which remained following the operational closure of Entigen (the commercial arm of ANGIS) in 2001. About half of the income needed for sustainability is obtained from subscriptions.

As outlined in the National Bioinformatics Strategy (NBS), the community has a divided view on ANGIS. High-end researchers found that ANGIS has completed its useful life, it no longer performs any unique function and ought not be considered as a national level service. The alternate view suggests that ANGIS maintains a unique role in providing software and database access, and education services - particularly for those in the start-up phase of 'doing' bioinformatics.

For this reason, it was recommended in the NBS that ANGIS' services be reviewed, to ascertain how it addresses the infrastructure needs of the bioinformatics community, and whether it warrants further funding in order to continue its existence.

Previous Reviews of ANGIS

A number of reviews and evaluations of ANGIS have been conducted, all of them indicating that ANGIS has unrealised potential. The earlier reviews, a report by the Bioinformatics Review Working Group (2000) and the "Pittard Report" (2000) recommended that ANGIS funding be increased, a view supported by the report of a National Bioinformatics Workshop (2002), stating that "a lack of funding may make ANGIS operation unsustainable. If it ceased to operate, universities and other organisations would have to duplicate these activities. Provision of similar services by the private sector would likely occur at greater expense."

Following the financial and leadership crisis in 2004, the University of Sydney commissioned a review of its bioinformatics potential and capabilities, a significant component of which involves review of ANGIS. The University of Sydney review was conducted in two stages, of the bioinformatics activities across the university and affiliated research institutes, with special regard to AGIC, which runs ANGIS, as a centre hosted by the university. The findings of the first stage is more relevant to the future of ANGIS, and is hence summarised here. The second stage pertains more to internal strategic recommendations for the University of Sydney, and is not deemed so relevant for the purpose of this review.

The first stage of the University of Sydney review aims to indentify and describe existing and potential bioinformatics capacity at the University of Sydney and its affiliated research institutes and organisations, including AGIC and ANGIS, and to assess the university's verall needs for capacity in bioinformatics. It is conducted by literature searching of electronics databases, study of documentation describing the establishment, development and achievements of the relevant centres, site visits of the offices and newly allocated spaces and interviews of several individuals who had an interest in bioinformatics and/or involvement in bioinfomatics research, distribution of e-mail questionnaire/survey and analysis of data describing bioinformatics research and teaching provided by the University of Sydney.

This review outlines the aspects of bioinformatics to be research and development (including those concerned with developing new techniques and those the outcomes of which rely on the use of bioinformatics techniques), support function (concerning training and assistance in data processing and management) and infrastructure (comprising computer hardware, software, networks and databases). The review also reports the recognition by both state and commonwealth levels of governments, that bioinformatics is essential for many fields of research, evidenced by the establishment of various bioinformatics task force and working groups.

Most relevantly to ANGIS, the University of Sydney review recommended that ANGIS and the Sydney University Biological Informatics and Technology (SUBIT) Centre was merged under a new umbrella of Sydney Bioinformatics. As a result this current review under the Australian Bioinformatics Network project is conducted in light of the new structure, and recommendations are made with Sydney Bioinformatics' services in mind.

Purpose and terms of reference of this review

The objectives of this review are:
1. critically review the services provided by ANGIS compared against the needs of the bioinformatics community - including basic, intermediate, and high-end users.
2. consider the mertis of broadening ANGIS' base of knowledge, expertise, and service provision beyond genomics into other fields of bioinformatics.
3. compare the services provided by ANGIS against those provided by other national bioinformatics service providers.
4. examine possible alternative roles for ANGIS as a central provider of bioinformatics services and support.
5. explore opportunities for partnership and networking between ANGIS and other organisations.

Methods

This review is conducted by the Executive Officer of the Australian Bioinformatics Network together with Dr Annette McGrath of the Australian Genome Research Facilities (AGRF), a member of the Bioinformatics Australia management committee selected by the Commonwealth based on the least conflict of interest.

Information is collected from the following sources.
1. Study of documentation
   The following documentations have been received for study:
      a. Bioinformatics in the University of Sydney: a review – stage 1, April 2005.
      b. Bioinformatics in the University of Sydney: a review – stage 2, April 2005.
      c. Bioinformatics Expert Task Force, Interim Bioinformatics Report, May 2003.
      d. Bioinformatics: Issues and Opportunities for Australia, Emerging Industries occasional paper 15, Department of Industry, Tourism and Resources, January 2002.
      e. NHMRC Research Committee, Report of the Bioinformatics Review Working Group (also known as the Pittard Report), January 2000.
2. ANGIS data
   This is provided by the CEO of Sydney Bioinformatics, on the request of the ABN EO.
3. Questionnaire survey
   This is to be circulated to ANGIS users. The CEO of Sydney Bioinformatics has been consulted on the distribution method. The questionnaire is conducted online, and is hosted at the AGRF webserver.

ANGIS' structure and services

ANGIS is managed by Sydney Bioinformatics, which is a Centre of the University of Sydney. Its operation is the responsibility of the CEO of Sydney Bioinformatics, who in turn answers to the Deputy Vice-Chancellor (Research) of the University of Sydney and an Advisory Board.

ANGIS is staffed by four people: a director, educational officer, system administrator and bioinformatics programmer. It is the Australian node of EMBnet, a collaboration of bioinformatic institutes and resources based in Europe, and is also a member of the Asia-Pacific Bioinformatics Network.

ANGIS provides access to analysis tools software and training in bioinformatics. The software is packaged in a platform named ``BioManager,'' which is easy to use, with extensive on-line help system. BioManager provides access to over two hundred programmes for analyses such as sequence database similarity search, pairwise sequence comparison, nucleic acid structure and analysis, gene detection and translation, multiple sequence analysis, statistical analysis, motif and pattern analysis, protein structure and function, molecular modelling, molecular evolution, PCR primer design and mapping. BioManager also links to databases such as GenBank, SWISS-PROT, Blocks, Prosite, Enzyme, Pfam and StackDB. ANGIS also provides educational services in the form of short courses (one to three days) in bioinformatics analysis and databases, at various venues in Australia. Of these packages, gcg, blast and clustalw are used most frequently by clients (full statistics can be found in Appendix G)

ANGIS also provided professional development courses in use of biomanager and other bioinformatics skills (microarray analysis and proteomics), which has been full or oversubscribed (table of courses and enrolment sizes can be found in subsection 3.1.3)

Results of ANGIS user feedback survey

Thirty seven users submitted their feedback forms, out of which five used ANGIS facilities for educational purposes as well as research. Most users use the facilities for research purposes, as summarised in the ANGIS statistics enclosed in Appendix ...

Half of the respondents claimed to have their own bioinformatics facilities, and only a small percentage also use other bioinformatics service providers. A great majority expressed that ANGIS has met their needs well, or very well, although those who use ANGIS for teaching purposes feel that it is slow. This suggests that new computing hardware may be required.

A number of respondents use the following webtools which they would like ANGIS to provide: Swissprot, NCBI, Invitrogen, Biology Workbench 3.2. Approximately a third of the respondents have taken ANGIS courses, most of which do not want more courses. This could mean that the course they took have satisfied their need.

Conclusions and recommendations

ANGIS under the new structure of Sydney Bioinformatics serves the functions as infrastructure (software and database) provider, education provider and bioinformatics technical service provider. This review examines mainly the services provided by the BioManager webtool, both to the research and education communities, and found that users are by and large satisfied with the level of service, with complaints revolving around the lack of speed, suggesting the necessity of equipment upgrade.

ANGIS' role in the provision of professional development has been successful. The courses have been fully subscribed, quite often with a waiting list. This has been thus far the only successful enterprise in bioinformatics professional development, which was found in the Consultative Workshops to be a need. It is recommended, therefore, that ANGIS' role in professional development be further examined, especially in how it could better address the demand in the community.

4.2 Establishment of a family of bioinformatics Research Fields, Courses, and Disciplines Code

Research Fields, Courses and Disciplines (RFCD) codes are developed by the Australian Bureau of Statistics (ABS) to classify research areas. The RFCD codes are organised as a family tree, as illustrated in the following example:
    270000 BIOLOGICAL SCIENCES
        270100 Biochemistry and Molecular Biology
            270101 Analytical Biochemistry
The largest class is called a Division, the next level a Discipline and the lowest a Subject, nick-named two-, four- and six-digit codes respectively.

These codes are used by research funding bodies to group grant proposals, in order to send the proposals to appropriate reviewers. The Australian Research Council, for example, classifies grant applications by the two-digit RFCD in the first instance.

The last RFCD codes were reviewed in 1998, at which time bioinformatics was very much at its infancy, and the community was far too small to warrant a code of its own. The absence of an RFCD code for bioinformatics has caused a level of grief amongst the bioinformatics research community. Since bioinformatics span the disciplines of life sciences, and mathematical and computing sciences, it is not clear to researchers, to which Division an application for research grant should be sent, and further, which Subject best approximates the piece of research proposed.

The absence of a bioinformatics RFCD code further renders it practically impossible to ascertain the grant success rate for bioinformatics. Only anecdotal evidence from the various sub-community exarcerbates the impression that bioinformatics is more poorly funded than average. As it stands, there is no means to (dis)prove this impression.

This frustration of the research community lead to a proposed recommendation in the National Bioinformatics Strategy, that an RFCD code be establish for bioinformatics. This is no trivial task, as an RFCD code is unique, and a decision remains on which Division, firstly, it should sit in. Since the bioinformatics community comprises members of multiple disciplines, it was expected that each would prefer their own discipline to host the new bioinformatics RFCD code.

The tree of codes is developed by the Australian Bureau of Statistics, guided by an Expert Committee from each Division. After initial contacts, the bioinformatics community has been invited to give input to the process. The ABN EO plays a role as a facilitator of discussion and spokesperson to the ABS.

The initial recommendation, by the Expert from the Information Sciences Division, is that bioinformatics should be placed in either Mathematical or Information Sciences, for reasons that the research method used is understood to be predominantly mathematically or computationally based. This is also supported by the OECD Fields of Science classification, which places bioinformatics under ``informatics.''

A representative of the community was consulted and an email discussion was conducted. The ABS' provisional placing was met with a mixed reaction, as expected. Approximately half responded in accordance to their own backgrounds, however, a good number gave the opinion that bioinformatics should be housed within the Biological Sciences division, including those of mathematical background. They argued that the multi-disciplinary nature of the subject matter does influence the research methods, that it indeed a mixture of methods from biological sciences and mathematical/computational sciences, and not only the latter.

A position paper summarising the results of these discussions was drafted (enclosed in Appendix H) and submitted to the ABS. This paper was well-received by the ABS, and was taken into consideration in formulating the first draft of the 2007 revision of the RFCD classifications, then re-named as the Fields of Research (FOR) classifications. This draft was released in September 2007, with a much more favourable result than the community had expected. Bioinformatics is placed in Biological Sciences, and along with it, a number of bioinformatics-related Subject codes were created:

Division Mathematical Sciences	Biostatistics
Division Biological Sciences	Bioinformatics
	Genomics
	Proteomics
	System Biology
Division Information Sciences	Bioinformatics Software

This draft along with transcripts of conversations with the ABS was published in the ABN online discussion forum (www.ausbiotech.org/forum) for further comments and feedback, to finalise the FOR classifications. The community was delighted with these developments, as can be viewed in the forum. This activity is then closed, as a success.

5. CONCLUSION

Coordination efforts have been conducted to establish the Australian Bioinformatics Network. A thorough analysis of the bioinformatics landscape has been performed, by close consultation of the community. A database of community members have been built and all members have been invited to participate in discussions. In addition to the original proposal, a second conference has been organised, which proved to be a success. A new committee comprising a number of key-players in the community is in office, actively planning future efforts to further bring the community together.

As this Australian Bioinformatics Network is at its nascence, to further consolidate the gains achieved in these beginnings, continuing efforts need to be made to nurture the network to its maturity. The consultative process yield a number of findings, giving rise to recommendations which in turn need to be addressed in the coming two year period.

Options for future developments include:
1. Supporting GIW 2008 in Australia:
   - it will be the first time for GIW to be hosted by Australia.
   - it is proposed to be held in Queensland, for its proximity to Asia, with some support of this from local organisations.
   - a staff member is is required to assist with logistics, venue liaison, international speaker programme, marketing, and other activities associated with conference organisation.
2. Further development of electronic communication via the world-wide web, by improving and updating content, management of online discussions, and active promotion of the online collaborative facilities on offer.
3. Identification of the potential role bioinformatics can play within the biotechnology industry, and assistance in development of career paths.

6. Financial Summary of ABN Grant Expenditure

Australian Bioinformatics Network
Financial Report

	1/7/06-30/6/07	1/7/07-31/12/07	Total
Salaries	$52,013.74	$40,986.26	**$93,000.00**
Operating Costs	$24,529.48	$12,121,22	**$36,650.70**
Conference Costs	$15,000.00	$19,295.45	**$34,295.45**
Travel - Committee	$7,385.25	$8,728.39	**$16,113.64**
Travel - EO	$1,265.99	$14,734.01	**$16,000.00**
Website	$18,181.82	$0.00	**$18,181.82**
Affiliations	$0.00	$1,818.18	**$1,818.18**
Student Scholarships	$0.00	$4,500.00	**$4,500.00**
Event Sponsorships	$863.64	$5.849.30	**$6,712.94**
	$119,239.92	$108,032.81	**$227,272.73**

Note 1: As at 31/12/2007 Accrual of $30,766.35 for: 2008/2009 Planning, website upgrade, travel and meetings expenses.

Contacts:

Dr Tim Littlejohn
President, Bioinformatics Australia
Business Development Manager, IBM Life-Sciences and Healthcare,
Acting CEO, The Australian Employers' Network on Disability
E-mail : tglittle@au1.ibm.com

Dr Anna Lavelle
Chief Executive Officer
AusBiotech Ltd
Level 1, 322 Glenferrie Rd
Malvern VIC 3144
Telephone: 03 9828 1400
Fascimile: 03 9828 5188
E-mail : alavelle@ausbiotech.org

Abbreviations

AASBi	Australiasian Association for Societies in Bioinformatics
ABS	Australian Bureau of Statistics
ANGIS	Australian National Genomic Information Services
APAC	Australian Partnership for Advanced Computing
ARC	Australian Research Council
BA	Bioinformatics Australia
GIW	Genome Informatics Workshop
MMIM	Molecular Medicine Informatics Model
NH & MRC	National Health and Medical Research Council
QFAB	Queensland Facilities for Advanced Bioinformatics
RFCD	Research Fields, Courses and Disciplines
SAPAC	South Australian Partnership for Advanced Computing
WAIMR	Western Australian Institute of Medical Research

Appendix A - Bioinformatics Australia Management Committee

Inaugural committee, term August 2005 - June 2007

Dr Catherine Abbott (Vice President)   Flinders University
Dr Jonathan Arthur (President)   University of Sydney
A/Prof Phoebe Chen   Deakin University
Dr Mark Crowe   Catapult Genetics
Dr Bruno Gaëta   University of New South Wales
Dr Dominique Gorse   Bio-Layer
Dr Annette McGrath   Australian Genome Research Facility
Dr Michael Poidinger   Johnson & Johnson Research
Dr Rohan Teasdale   University of Queensland

New committee, term July 2007 - June 2008

Dr Catherine Abbott   Flinders University
A/Prof Phoebe Chen   Deakin University
Dr Bruno Gaëta   University of New South Wales
Dr Tim Littlejohn (President)   IBM Australia
Dr Annette McGrath   Australian Genome Research Facility
Dr Michael Poidinger   Johnson & Johnson Research
Prof Shoba Ranganathan   Macquarie University
Prof Mark Ragan (Vice-president)   University of Queensland
Dr Rohan Teasdale   University of Queensland

The committee comprises seven elected members and up to two AusBiotech-appointed members. The appointments are made to balance geographical or industrial composition of the committee.

Elections are held annually, where a part of the elected positions are open. This part-election is conducted to ensure continuity in the committee. Each elected and appointed committee member holds a two-year and one-year office respectively.

Appendix B – members of the bioinformatics community consulted
Academic and research community

Prof David Adelson	University of Adelaide
Dr Douglas Anderson	Box Hill Institute
Dr Melanie Bahlo	Walter and Eliza Hall Institute of Medical Research
Dr Tim Bailey	University of Queensland
Dr Swati Baindur-Hudson	Victoria University
Dr Jim Bashford	University of Tasmania
Prof Philip Batterham	University of Melbourne
Dr Ute Baumann	Australian Centre for Functional Plant Genomics
Dr Alex Beesley	Telethon Institute of Child's Health Research
Prof Matthew Bellgard	Murdoch University
Dr Mikael Boden	University of Queensland
Dr Cindy Bottema	University of Adelaide
Dr Sarah Boyd	Monash University
Dr David Bulger	Macquarie University
Dr Conrad Burden	Australian National University
Prof Kevin Burrage	University of Queensland
Dr Kim Carter	Western Australia Institute of Medical Research
Ms Kao Casle	University of Sydney
Mr Cheong-Xin Chan	University of Queensland
A/Prof Danny Coomans	James Cook University
Prof Ross Coppel	Victorian Bioinformatics Consortium
Dr Brian Dalrymple	CSIRO
Dr Melissa Davis	University of Queensland
Mr Saravanan Dayalan	RMIT University
A/Prof Nick de Klerk	Telethon Institute of Child's Health Research
A/Prof Brian Dean	Mental Health Research Institute
Prof Simon Easteal	Australian National University
Dr Noel Faux	National ICT Australia
Prof Simon Foote	Menzies Research Institute
Dr Sylvain Fôret	Australian National University
Dr Murray Grigor	Uni Auckland Bioinformatics Institute, NZ
Dr Ilka Havukkala	Auckland University of Technology, NZ
Ms Manal Helal	University of New South Wales
Prof Mike Hendy	Allen Wilson Centre for Molecular Ecology, NZ
A/Prof Jim Hogan	Queensland University of Technology
Mr Fabien Huard	Macquarie University
Dr Gavin Huttley	Australian National University
Ms Izmira Ismail	University of Sydney
Dr Lars Jermiin	University of Sydney
Dr Warren Kaplan	Garvan Institute for Medical Research
Dr Karin Kassahn	University of Queensland
Dr Gavin Kennedy	CSIRO Plant Industry
Dr Adam Kowalczyk	National ICT Australia
Prof Don Kulasiri	Lincoln University, NZ
Mr Braddon Lance	Macquarie University
Ms Melanie Lehman	Queensland University of Technology

Appendix B – members of community consulted (continued)
Academic and research community

Dr Vladimir Likic	Bio21 Institute and Metabolomics Australia
Dr Guangbin Liu	University of Southern Queensland
Dr Jesus Lopez	University of Southern Queensland
Dr David Lovell	CSIRO Statistical Bioinformatics
Mr Geoffrey Macintyre	National ICT Australia
Prof Geoff McLachlan	University of Queensland
Dr David Mitchell	CSIRO Mathematical and Information Sciences (MIS)
A/Prof Pablo Moscato	University of Newcastle
Prof Ajit Narayanan	Auckland University of Technology, NZ
Prof Collen Nelson	Queensland University of Technology
Prof Frank Nicholas	University of Sydney
Ms Maree O'Sullivan	CSIRO MIS
Dr Alicia Oshlak	Walter and Eliza Hall Institute
Dr Tony Papenfuss	Walter and Eliza Hall Institute
Dr Victoria Perreau	Neurosciences Victoria
A/Prof Tuan Pham	James Cook University
Prof Peter Reeves	University of Sydney
Dr Peter Ritchie	Victoria University Wellington, NZ
Dr Shamith Samarajiwa	Monash University
Dr Andreas Schreiber	Australian Centre for Plant Functional Genomics
Dr Moh Shoaib Sehgal	University of Queensland
Mr Gowri Shankar	RMIT University
Dr Alex Smola	National ICT Australia
Dr Gordon Smyth	Walter and Eliza Hall Institute
Prof Terry Speed	Walter and Eliza Hall Institute
Prof Mike Steel	University of Canterbury, NZ
Dr Natalie Thompson	Peter MacCallum Cancer Institute
Prof Peter Timms	Queensland University of Technology
Dr Guy Tsafnat	University of New South Wales
Mr Jonathan Tuke	University of Adelaide
Dr Anna Tsikin	University of Adelaide
Mr Philip Uren	University of Tasmania
Dr Wynand Verwoerd	Lincoln University, NZ
Dr Matthew Wakefield	Walter and Eliza Hall Institute
Prof James Whisstock	Victorian Bioinformatics Consortium
Prof Marc Wilkins	University of New South Wales
Dr Rohan Williams	Australian National University
Dr Bill Wilson	CSIRO MIS
Prof Sue Wilson	Australian National University
A/Prof Michael Wise	University of Western Australia
Prof Graham Wood	Macquarie University
Dr Ian Wood	University of Queensland
Dr Meridee Wouters	Victor Chang Cardiac Research Institute
Dr Jean Yang	University of Sydney
Dr Lauren Young	Central Queensland University
Prof Justin Zobel	National ICT Australia

Appendix B – members of community consulted (continued)

Infrastructure provider

Dr Jonathan Arthur	Director, Sydney Bioinformatics
Prof Mark Baker	CEO, Australian Proteomics Analysis Facilities
Mr Jeremy Barker	CEO, Queensland Facilities for Advanced Bioinformatics
Prof Matthew Bellgard	Australian Bioinformatics Facilities (NCRIS 5.1)
Dr Markus Buchhorn	AARNet/Australian National University
Dr Sonia Cattley	Sydney Bioinformatics
Dr Paul Coddington	South Australian Partnership for Advanced Computing
Mr Brett Cooke	Bioinformatics, Australian Proteomics Analysis Facilities
Dr Sue Forrest	CEO, Australian Genomics Research Facilities
Mr Andrew Gilbert	Manager, Bioplatforms Australia (NCRIS 5.1)
Dr Dominique Gorse	Queensland Facilities for Advanced Bioinformatics
Dr Marienne Hibbert	Manager, Molecular Medicine Informatics Model (Bio21)
Dr Craig Hill	South Australian Partnership for Advanced Computing
Dr Margaret Kahn	ANU Supercomputing Facility/APAC
Ms Cath Lawrence	ANU Supercomputing Facility/APAC
Dr Annette McGrath	Bioinformatics manager, Australian Genomic Research Facilities
Dr Jamie Sherman	Bioinformatics, Australian Proteomics Analysis Facilities
Mr Paul White	Western Australia Institute of Medical Research

Commercial/private sector

Dr Peter Amer	Abacus Bio, NZ
Mr Keith Antonisz	ISA Technologies
Dr Leonard Bloksberg	Cartesian Gridspeed
Dr Nick Conomo	SGI Computers
Dr Mark Crowe	Catapult Genetics
Dr Peter Fenessy	Abacus Bio, NZ
Mr John Henderson	CRAY Computers
Dr Grant Jacobs	BioInfoTools
Dr Bonni Reichhelt	Gene Search
Dr Bruce Ross	IBM Life Sciences and Healthcare
Dr Ken Simpson	Watermarks Patent Attorney
Dr Mervyn Thomas	Emphron Informatics
Ms Candace Toner	CEO, Biomatters Ltd (NZ)
Dr Christopher Wilkinson	Madderns Patent Attorneys

Government agencies

Dr David Brett	Australian Bureau of Statistics
Prof Margaret Clayton	Australian Research Council
Dr Kate Le Strange	Department of Industry, Tourism and Resources
Prof Jonathan Manton	Australian Research Council
Ms Therese McLennan	Department of Industry, Tourism and Resources
Mr David Wilson	Department of Education Science and Training
Mr Roland Wise	National Health and Medical Research Council

Appendix C - Discussion points from the consultative workshops

Current infrastructure needs:
- What are your infrastructure needs (e.g. database mirrors, software, high-performance machine)?
- Do you have current unfulfilled infrastructure needs?
- What stands on the way on these needs being unfulled? (e.g. software/database not available or accessible, or they are available but too expensive, or others . . .)

Future infrastructure needs:
- What do you foresee your research would need in the future, that are currently unavailable?

On research grants:
- Have you applied for research funding to the ARC, the NH & MRC, or similar bodies? If yes, were you successful? If no, were there any factors discouraging you? In either case, do you have any general comments about Australian funding for research on bioinformatics?

On bioinformatics skills and employability:

Does bioinformatics training meet the needs of the job-market?
1. If you have been in the position of leadership, do you find there is a shortage or surplus of bioinformatics candidates? Have you been satisfied with the range and levels of their skills? The following is a list of skills collected from a number of employers we have consulted:
    - algorithmics: programming skills and familiarity with a few programming languages,
    - experience in genomics databases, knowledge of statistical analysis
    - knowledge of life sciences (genetics, molecular biology, biochemistry, cell biology)
    - ability to work in multidisciplinary teams

    Do you have anything to add?
2. Looking at the number of courses mentioned, do you think there are enough training courses to meet the demands of the job-market? If not, what is the nature of this need of more bioinformaticians? (e.g., should there be more short-courses or continual education to train the existing workforce, or should there be more degree courses?)
3. If you are a student or recent graduate, what is/has been your experiences of job-seeking? Did you know the employers to target? Did you know what they expected? Did you feel that your training has given you enough skills to practice as a bioinformatician? If not, what factors make the gap between what you have learned and what you find you need at work or what you would have liked to learn?

On bioinformatics curriculum

Bioinformatics undergraduate courses come in different shapes and forms, as a Bachelor of Science (Genetics, Biochemistry or Molecular Biology majors), Applied Science, Biotechnology, Information Technology, Computer Science, Biostatistics, and a couple of universities (Latrobe and USQ) offered Bachelor of Bioinformatics. A number of universities offer the stream of multiple courses (e.g., BCS and BSc majoring in biochemistry from UniMelb). Postgraduate degrees by coursework follow similar course titles.

There is also an online education consortium, the S¤ Alliance, http://www.s-star.org/, arising from the collaboration between universities from Australia (Macquarie and USyd), Sweden (Karolinska Institute and Uppsala), Singapore (NUS), South Africa (U Western Cape) and USA (Stanford and UCSD).

1. The titles and foci of these courses are typically historical, where they first arose in a given institution. In a sense, this also reflects the nature and origins of the community. Do you think this is a good thing, or do you think it is better to have a unified curriculum in bioinformatics that all universities follow?
2. If you think a unified/common/shared curriculum for all institutions is a good idea, what are the essential subject matters that it should cover (e.g., which life sciences subject matters and which IT/mathematical/statistical techniques), and what are the essential skills that should be honed during the courses (e.g., programming, databases, bench- work)? Should it be implemented at undergraduate or postgraduate levels? Do you think an online course such as provided by the S¤ Alliance is a good framework for bioinformatics education?
3. If you think it is not a good idea, what are your reasons?

On related industries

Which industry is most closely associated to your research?
For example:
1. which industry would you approach to commercialise a discovery?
2. which industry would you approach should you decide to leave the academic career path?
3. which communities use the results of your work?

Appendix D Summary of Bioinformatics Australia 2007 Exit Survey

Demography

The data quoted here concerns registrants to the Bioinformatics Australia conference only. Registrants to the AusBiotech 2007 conference could also attend the Bioinformatics Australia conference without having to register separately. There were no means to record these attendants.

There were 109 registrants at the BA 2006 conference, 26 of which were students. The majority of registrants, numbering 79 delegates, belong to universities; 11 to research institutes and facilities, seven to infrastructure providers, three to the private sector and six from government departments.
Based on their geographical origins, a delegate each came from Brazil, France, Finland, USA and Korea, seven from New Zealand, and the rest from Australia.

Exit survey summary

Delegates were asked to complete an exit survey to give the committee feedback on the quality of the BA 2006 conference, and to help in preparation of a better BA 2007 conference. Of the 109 delegates, 48 responses were collected, of which 16 were students. The responses of this
sample are summarised as follows.
1. On the matter of scientific content: 8 thought it was not enough, 1 too much, and an over-whelming 38 out of 48 thought it was just right.
2. On the ratio of invited version contributed papers, 11 respondents thought there should be more invited talks, 14 thought there should be more contributed papers and 21 thought the proportion was about right.
3. Thirty one respondents said they had enough time to view the posters, 11 disagreed and 6 did not respond.
4. Out of the 48 responses, 1 thought it was too short, 2 too long, and an overwhelming 43 thought it was the right length
5. On the question whether they would be likely attend future BA conferences, 9 responded likely, 17 very likely and 16 definitely, with 4 less committed responses of `probably.'
6. Quality of the various aspects of the conference is summarised in the table, which is a reproduction of the tick box the delegates were asked to fill. In the processing, the responses: poor, acceptable, good and outstanding are given the scores 1,2,3,4 respectively. The average scores sit between 2.75 and 3.0, indicating that on average respondents think that these aspects areby and large good. It needs to be noted, however, that since there are very few responses on the conference dinner question, the results are not representative of either the sample or the population of delegates.

	Poor	Acceptable	Good	Outstanding	Average
Overall scientific programme	0	13	33	1	2.75
Keynote presenters	0	13	28	5	2.94
Contributed papers	0	6	40	0	2.87

Poster session	1	7	33	1	2.81
Printed programme book	0	12	32	3	2.83
Exhibits	0	13	21	5	2.82
Refreshments	0	14	25	7	2.87
Conference dinner	1	0	3	4	2.75
Overall conference organisation	0	5	38	5	3.00

7. The table below summarises the suggested topics for future conferences. Where a topic is suggested multiple times, it is indicated in the number next to that topic.

Topic	
System biology	8
Transcriptomics	2
Proteomics	2
Commercialisation	2
More statistical component	2
Bioinformatics application	
Bioinformatics software	
Data management	
Emerging technologies	

8. The table below summarises the responses on how they benefit from the BA conference association with AusBiotech:

Topics	
Networking/interactions/collaborative potentials	10
Exhibits	5
Expressed great benefit but not specify	6
Informative/exposure	5
Little interaction/no benefit at all	11
Useful background	
freebies	

9. The table below summarises what the respondents think are the best aspects of the conference:

Topic	
Networking	12
Discussions	8
Scientific content	5
Small, friendly and informal	4
AusBiotech association	

10. The aspects upon which we can improve are listed below. There are less multiple-responses on this question.
   - Longer time allocation for contributed talks
   - More, shorter contributed talks
   - dissociate with AusBiotech
   - closer interaction with AusBiotech conference
   - software demonstration
   - Bioinformatics Australia Annual General Meeting to be held at the conference
   - time-keeping
   - copies of presentations be made available after the conference
   - results of feedback be shared after the conference
   - first plenary was not very interesting but the second one was great
   - incorporate social/touristic activities
   - student symposia
   - tutorials for experimentalists

Based on these responses, we concluded that the majority of delegates were happy with the conference, on both scientific content and organisational matters, and that we can expect a great majority of them to attend future conferences.

Appendix E Bioinformatics Australia 2006 Conference – comprehensive reporting (copy of reporting from Progress Report, 30 November 2006)

The inaugural Bioinformatics Australia conference was held on 21-22 November 2006, in conjunction with the AusBiotech 2006 national conference, at the Sydney Convention and Exhibition Centre. It was attended by 120 delegates, from all over Australia and overseas, comprising academics, industry and government representatives and students. It was sponsored by the NSW government Office of Science and Medical Research, CSIRO, National ICT Australia, Australian Genome Research Facilities, APAF, the Ramaciotti Centre, Supamac and the Garvan Institute. It comprised discussion sessions and scientific sessions, both of which were deemed to be more than satisfactory by delegates in their feedback survey. They particularly enjoyed the opportunity to meet, network and have discussions both on scientific and community-building issues.

The conference was a success, with the majority of delegates indicating that theyare happy with the content and organisation of the conference, and are supportive of future conferences with their attendance.

An exit survey was conducted to give the committee feedback on the current conference and to help organise a better conference in the following year. A summary of the results can be found in Appendix F of this report.

Conference - Australian Bioinformatics Network issue sessions

There were two discussion sessions devoted to issues surrounding the establishment of the Australian Bioinformatics Network. These sessions were conducted in the form of panel discussions, lead by key-players in the Australian bioinformatics community.

The first session focused on issues of infrastructure, especially on the National Collaborative Research Infrastructure Strategy (NCRIS), with the following expected key outcomes:
1. that the bioinformatics community be informed on the infrastructure which already exist and which is being developed,
2. that the members of the community with most interest in infrastructure issues be identified, both the contributers and users.

Of the recommendations in the National Bioinformatics Strategy, the items addressed in this discussion include making recommendations to bioinformatics infrastructure suppliers in Australia (under this heading including gathering feedback from the community on the use of infrastructure facilities, determine effectiveness of existing infrastructure and making recommendations on improvements) and working collaboratively with NCRIS stakeholders.

The panel consists of Professor Matthew Bellgard from Murdoch University, Dr Rhys Francis, previously from CSIRO High Performance Scientific Computing now seconded to the Department of Education, Science and Technology, and Professor Ross Coppel from the Victorian Bioinformatics Consortium. Professor Bellgard is responsible for the implementation of the Capability 5.1 investment in bioinformatics infrastructure and Dr Francis is the facilitator of NCRIS Capability 5.16 respectively. They therefore

represent the providers of bioinformatics-relevant infrastructure. Professor Coppel gave the point of view of users of infrastructure, to ensure that the needs of infrastructure to support research, to ensure that the infrastructure suppliers meet the needs of research in bioinformatics/biosciences.

In the second session, themed ``What kind of bioinformatics does Australia need?", the discussions surrounded the the issues of having an identity as a community, such as `` who we are" and ``where we are going." The panel comprised Professor Terry Speed of the Walter Eliza Hall Institute, Dr Sue Forrest of the Australian Genome Research Facility and Professor Marc Wilkins from the University of New South Wales.

This session aims to address recommendations in the National Bioinformatics Strategy in the areas of Research and Development, and Education and Training, by stimulating
a discussion amongst the bioinformatics community about who we are, where we sit in the bigger scheme of things, and where we are going as a community.
To start this discussion, the panel to gave their thoughts in the issues of
1. connecting bioinformatics to the broader Australian community and emphasising its benefit to this broader community,
2. realising the potential of bioinformatics to greater benefit,
3. making bioinformatics outcomes relevant to the wider community and how the Australian Bioinformatics Network can help,
4. breaking open the mindset of "bioinformatics" and achieving bigger outcomes.

Both sessions were very successful, with significant participation from the floor. The discussions were recorded, to be transcribed and posted on the website, with great anticipation from the participants of continuing web-based discussion fora.

Conference - scientific programme

The scientific programme comprised oral presentation sessions and a poster session. Each of the four oral sessions started with a plenary, delivered by national and international invited speakers from Oxford University, the US National Institute of Health,
the Korean Advanced Institute of Science and Technology and CSIRO, and followed by peer-reviewed contributed papers.

There were fifty posters from the four scientific themes displayed in the poster hall, where delegates could browse at any time during the conference as well as during the allocated poster reception.

The scientific sessions are deemed to be a success, with a majority of the respondents stating that they were good or outstanding.

Conference – networking

The ABN EO made contact and held discussions with key-players of the bioinformatics community, such as the above-mentioned panelists, representatives of NICTA, QFAB, APAC.

During one of the discussions APAC informed the ABN EO that an Australian mirror of
the major international databases included in www.biomirror.net is hosted at the Australian Academic and Research Network (AARNet) server. This information and a link to it has been included in the ABN webpage. QFAB briefed the ABN EO on its establishment, progress in its nascence and future directions.

Two state networking events occured in conjunction with the conference. Discussion sessions on the establishment of the NSW bioinformatics network were included in the conference programme, and a BA-sponsored SA networking event was held after the conference, on 23 November 2006, where the invited speaker from US NIH also spoke and networked with the community in Adelaide. The SA event was attended by fifty people, twenty of whom were students.

The conference also gave the bioinformatics community an opportunity to network, amongst each other and with the greater biotechnology community, which in their exit survey they indicated as a highlight of having attended the meeting. Members of the community who wish to play an active role in discussions and management matters are
identified amongst both established key-players and new players.

Appendix F Summary of Bioinformatics Australia 2006 Exit Survey

Demography

The data quoted here concerns registrants to the Bioinformatics Australia conference only. Registrants to the AusBiotech 2006 conference could also attend the Bioinformatics Australia conference without having to register separately. There were no means to record these attendants.

There were 119 registrants at the BA 2006 conference, 44 of which were students. The majority of registrants, numbering 81 delegates, belong to universities; 27 to research institutes and facilities, nine to the private sector and two from government departments. Based on their geographical origins, 78 came from NSW, 13 from ACT, ten from Victoria, nine from Queensland, three from SA, one from WA and one from Tasmania. There were four registrants from overseas, three of which were invited speakers.

Exit survey summary

Delegates were asked to complete an exit survey to give the committee feedback on the quality of the BA 2006 conference, and to help in preparation of a better BA 2007 conference. Of the 119 delegates, 56 responses were collected, of which 22 were students. The responses of this
sample are summarised as follows.

2. On the matter of scientific content: 9 thought it was not enough, 1 too much, and an over-whelming 45 out of 55 thought it was just right.
2. On the ratio of invited version contributed papers, 10 respondents thought there should be more invited talks, 9 thought there should be more contributed papers and 33 thought the proportion was about right.
3. Forty two respondents said they had enough time to view the posters, 9 disagreed and 4 did not respond.
4. Out of the 55 responses, 3 thought it was too short, 2 too long, and an overwhelming 50 thought it was the right length
5. On the question whether they would be likely attend future BA conferences, 8 responded likely, 27 very likely and 12 definitely, with 4 less committed responses of `probably.'
6. Quality of the various aspects of the conference is summarised in the table, which is a reproduction of the tick box the delegates were asked to fill. In the processing, the responses: poor, acceptable, good and outstanding are given the scores 1,2,3,4 respectively. The average scores sit between 2.75 to 3.00, indicating that on average respondents think that these aspects areby and large good. It needs to be noted, however, that since there are very few responses on the conference dinner question, the results are not representative of either the sample or the population of delegates.

	Poor	Acceptable	Good	Outstanding	Average
Overall scientific programme	0	21	31	3	2.75
Keynote presenters	0	6	30	18	2.94
Contributed papers	1	15	39	0	2.87

Poster session	1	15	32	5	2.81
Printed programme book	0	12	35	8	2.83
Exhibits	3	19	22	2	2.82
Refreshments	1	23	25	5	2.87
Conference dinner	2	1	5	1	2.75
Overall conference organisation	0	10	33	9	3.00

7. The table below summarises the suggested future topics for Bioinformatics Australia 2007. Where a topic is suggested multiple times, it is indicated in the number next to that topic.

Topic	
Evolution/phylogeny/comparative genomics	7
System biology	6
Computation/algorithms (incl data mining & machine learning)	5
Gene expression	3
maths (modelling) and stats	3
medical bioinformatics (incl drug discovery)	3
review of bioinformatics tools and software	2
structural bioinformatics	2
Protein structure prediction	2
Proteome	2
Networks: interactomes	
rational gene design (?)	
review on experimental design	
marine biotech/bioinformatics	
Population genetics	
Microbial biotechnology	
metabolomics	
Collaboration with regional consortia (NZ & Asia)	
Commercial bioinformatics	
Biological data capture and management	
RNA biology/mRNA detection	

8. The table below summarises the responses on how they benefit from the BA conference association with AusBiotech:

Topics	
exhibits/trade displays	
networking	
Meet people	
little/not much	
give good/new ideas/info/vision	
Not at all	
broad overview of biotech	
opportunity to meet industry	
learned	
Talking to peers	
great/yes	
Useful background	
freebies	

9. The table below summarises what the respondents think are the best aspects of the conference:

Topic	
keynote	
Issue session	
networking	
Scientific content	
topic diversity	
Meet people	
Good speakers	
Poster session	
Research discussion	
Size (allow discussion)	
venue	
organisation	
Single track	
freebies	
Good atmosphere	
Meeting of the community	
Not travel	
Red wine prize	
Mix of science and discussion	

10. The aspects upon which we can improve are listed below. There are less multiple-responses on this question. Where these exist, the number of responses are indicated in brackets.
    - more time for questions more time for discussion

- more info week prior more info in website
- more scientific paper affordable dinner
- cheaper venue extended abstract
- longer conference, say 3 days more contributed talks (5)
- more keynote (3) condition - environment
- broader presentation more technical talks
- free/easy parking poster prize
- longer contributed talks clustering of major interest
- timing shorter keynote
- parallel session talk on expt design
- more info for interstaters poster session
- invite comp sci speaker closer catering venue
- too bio focused
- separate programme to "pure" and "applied"
- catering - better lunch box, more substantial vege food
- more choice in operating system for AV
- programme - w index, in order of presentation, more info on room
-

Based on these responses, we concluded that the majority of delegates were happy with the conference, on both scientific content and organisational matters, and that we can expect a great majority of them to attend future conferences.

Appendix G. ANGIS statistics

Package	Research jobs	Training jobs	Total	% Training jobs
gcg	3030	1714	4794	36%
blast	1271	1527	2798	55%
clustalw	1508	359	1867	19%
phylip	929	249	1178	21%
mbat	651	78	729	11%
emboss	222	232	454	51%
fasta	149	112	261	43%
primer3	216	40	256	16%
boxshade	113	89	202	44%
flip	126	76	202	38%
gscan	39	20	59	34%
tacg	24	34	58	59%
blocks	16	5	21	24%
wise	3	10	13	77%
phred	13	0	13	0%
protml	9	0	9	0%
hmmer	1	3	4	75%
fastdaml	0	3	3	100%
special	0	1	1	100%
Total	8370	4553	12923	35%

Summary of ANGIS users:

No of logins	Academic	Institute/hospital	Government	Commercial	Total
0-10	26	17	5	2	50
11-60	26	6	6	0	38
61-120	8	2	0	0	10
121-180	2	2	2	0	6
Total	62	27	13	2	104

Total number of logins subscribed to is 2872.

Appendix H Bioinformatics Australia position paper on RFCD code submitted to the Australian Bureau of Statistics

Introduction

This position paper has been prepared by the Executive Officer of the Australian Bioinformatics Network in order to open a discussion with the Australian Bureau of Statistics regarding the establishment of the Research Fields, Courses and Disciplines for bioinformatics. It aims to convey the views of the Australian bioinformatics community on where in the RFCD family tree the bioinformatics code should sit.

The paper presents opinions collated from selected members of the Bioinformatics community, in discussion with the Australian Bureau of Statistics. It is not meant to be proscriptive or final. Its purpose is to state the perspective of the Bioinformatics community on where it sits as a research subject and give recommendation on the position of RFCD for bioinformatics to the ABS and its panel of experts.

Background

The term "bioinformatics" was coined in 1990, and with the research community developing in Australia around the mid-nineties. The community was at its infancy at the last revision of RFCD in 1998, which accounted for its absence in the current classification.

It has since then gained momentum, as a fast-emerging field. National Bioinformatics Strategy (2002) reported Australia's potential for research outcome (such as the high-standard research capabilities, substantial biological and biodiversity knowledge contained in the range of datasets, as well as a unique biodiversity and well defined population for genetic studies) and Australia's growing number of institutions undertaking bioinformatics R&D. It also reported that this growth is not reflected in the number of bioinformatics published articles. Australia was shown to be lagging behind overall global growth in research outcomes, inconsistent with the potential it presents.

Bioinformatics researchers feel that due to its multi-disciplinary nature it is non-trivial to place their proposals in the current RFCD classifications. It is unclear to the researchers that reviewers with expertise in either RFCD divisions which span bioinformatics have enough understanding of the other aspect, to be able to judge the grant-worthiness of bioinformatics projects. In discussions with key-players in the Bioinformatics community, the expert task-force found that the establishment of an RFCD classification code would better identify bioinformatics as a component of research applications.

Bioinformatics Australia is the peak coordinating body for the bioinformatics community, with the aim of providing leadership in bioinformatics research, education and commercialization, and providing representation to authorative/governmental bodies. One of the activities it has been commissioned to do, as a part of a Commonwealth Government project, is to liaise with the Australian Bureau of Statistics and provide input on behalf of the community in the process of establishing an RFCD code for bioinformatics.

Discussion points

Bioinformatics is a research discipline involving 'input' from many diverse fields, including mathematics and statistics, chemistry and physics, biological, medical and agricultural sciences, and information

science. It refers to a range of cross-disciplinary endeavours rather than a single discipline, and members of the bioinformatics community come from the various backgrounds that compose it.

The ABN EO invited discussions amongst the research-members of the Bioinformatics Australia management committee and other representative members of the bioinformatics community, to encompass the diverse background that composes bioinformatics and to include members from five states (QLD, NSW, VIC, SA, WA) in Australia. The question was whether bioinformatics would be most suitably placed in the division of Mathematical Sciences (230000), Biological Sciences (270000) or Information, Computing and Communication Sciences (280000). This initial discussion followed from communications with the ABS, on the establishment of a bioinformatics RFCD code and preliminary proposals on its location, that there were thoughts that it could sit either under Biotechnology, which will be placed under Engineering and Technology (290000), or under ICCS.

The respondents and responses are summarised in the following table, with the details provided below.

	23xxxx Mathematical Sciences	27xxxx Biological Sciences	28xxxx Information, Computing and Communication Sciences	29xxxx Engineering and Technology
Respondent Background	5	6	2	1
Response	1	7	2	0

Considering the diverse backgrounds of the community, the responses are understandably varied. It is expected that each subgroup in the community would support placement of the RFCD under the discipline of their respective Interestingly, however, an overwhelming majority of responses favours the placement of bioinformatics RFCD under Biological Sciences, including from members of the community of mathematical sciences and IT backgrounds. This paper therefore is composed to present the view of the majority, with particular attention to the arguments coming from respondents who do not support their own background.

All respondents believe that purpose of bioinformatics is to advance biological and medical science, not to advance mathematics, information science or engineering, although those might sometimes arise as a bi-product. The respondents who support RFCD placement within biological sciences (of both biological and mathematical/IT backgrounds) also believe that the process involved in bioinformatics research comprise mixtures of the discipline components in bioinformatics, with different proportions from one individual to another. This belief was reinforced by assessor reports of ARC grants that they have viewed. The research intention affects the process; if the intention to solve biological problems, this transforms the entire research process. Methods which originate from the mathematical or information sciences domain are adapted to enable address of biological questions.

There is a case example where a discipline whose process involving a substantial amount of mathematics sit in another Division, namingly 250600 Theoretical and Computational Chemistry, including subjects such as Quantum Chemistry and Statistical Mechanics, which thread the fine line between chemistry and mathematics (with the latter being housed in the mathematics department in a number of universities).

The community is in favour of a similar treatment to bioinformatics, that it be made a discipline named "Theoretical and Computational Biology." Having understood that this may be premature at this stage of the community growth, it is still a strong preference that the subject-level (six-digit) RFCD code should sit within biological sciences, with the location of "Theoretical and Computational Chemistry" within the Chemistry division as a case-point.

Simiarly to bioinformatics, theoretical and computational chemistry is the study of chemistry undertaken using mathematics and information technology. As such, it receives a classification in the 2506xx range under Chemical Sciences. This presents an exact analogy to bioinformatics. Bioinformatics can be considered as another name for theoretical and computational biology. The bioinformatics research community studies biology using mathematics and information technology.

For example, molecular dynamics simulations would fall under statistical mechanics (250602) and may involve simulating on the computer the movement of atoms under mathematical models of the chemical forces between the atoms. These motions then form input data for statistical mechanical equations to calculate chemical properties of the system. [e.g. Arthur and Haymet, J. Chem. Phys. 1998, 109, 7991] This is exactly analogous to work in bioinformatics. For example, the prediction of protein structure by comparative modelling involves using a computer to implement a theoretical model of sequence evolution in order to find homologous sequences whose three-dimensional structure can be used to as a basis to predict the structure of the unknown sequence. The methods employed in theoretical and computational chemistry are very similar to those used in bioinformatics. In fact, Theoretical and Computational Chemistry overlaps with bioinformatics once the chemical entities under consideration become "biochemical" entities such as peptides, proteins, DNA, etc.

The sense that bioinformatics is a biological science rather than mathematical or information science is strengthened by international publishing bodies. For example, the Oxford University Press, who publishes "Bioinformatics," places the journal under its Life Sciences selection (http://bioinformatics.oxfordjournals.org/). Another main bioinformatics journal, "BMC Bioinformatics," is published by BioMed Central (http://www.biomedcentral.com/bmcbioinformatics/), clearly a publishing operation for biological and medical sciences.

Biological topics studied in bioinformatics spanned the disciplines of Biochemistry and Cell Biology (270100) and Genetics (270200) – approximately equal proportions of the community indentify their research to belong in either disciplines.

Conclusion

It is a voice of the majority within the bioinformatics community that the RFCD code should be located within the Biological Sciences (270000) division. This opinion is shared by those of biological and mathematical/information sciences backgrounds alike. It is understood that RFCD is a process classification. The community believes that the process in bioinformatics research is a marriage of both biological and mathematical sciences, i.e. adapted from mathematical origins to specifically answer biological question. A case-example analogous to this is presented in Theoretical and Computational Chemistry, which sits in the Chemical Sciences discipline.

The prospect for the establishment of Biostatistics under the Statistics discipline would provide an alternative for bioinformatics researchers whose work is more mathematically bent. This further strengthens the argument for the placement of Bioinformatics subject code under Biological Sciences, to cater for bioinformatics research which is further removed from rigorous mathematics, where the process is closer to the biological sciences side of the continuum.

AusBiotech Ltd
Level 1, 322 Glenferrie Rd
Malvern, VIC 3144
Australia
Ph: +61 3 9828 1400