Unexpected Paths in Applied Computing: A Journey from eHealth to Sports Science with Other Applied Computing Ventures Along the Way

David Paul

2024-07-24

This presentation traces my research journey, beginning with the development of a secure portal for the Australian Schizophrenia Research Bank (ASRB). This led to my PhD in distributed systems, where I focused on Web services transactions. The experience of managing large datasets for the ASRB was helpful in agricultural applications, including systems like ASKBILL for sheep wellbeing. More recently, my expertise has been applied to sports science, where I have contributed to gait analysis and performance monitoring in rugby league. This experience demonstrates the versatility of applied computing across various domains and shows the value of interdisciplinary research in solving complex, real-world problems. This unplanned research path, following my nose and seizing opportunities as they arose, has led to many unexpected and fulfilling ventures.

Introduction

Research Island

Hi, everyone, and thanks for coming along.

I volunteered to give this discipline research seminar, but I wasn't really sure what I was going to talk about. Last month, Vera spoke about her research in error-correction codes, and it's obvious she's had a very focused and direct research trajectory, so I thought it might be useful to talk about a path that wasn't quite so clear. So, I'll talk more generally about my research journey—where I started, where I am now, and the paths I took to get here.

I won't be able to cover everything, but I'll give some details on how I started and where it's led. Recently, I reread Treasure Island for the first time since primary school (still a great book, by the way), so I've asked ChatGPT and DALL-E to help theme this presentation around it - it seems appropriate for talking about unexpected paths. You can see it really loves compasses and is terrible with text, but please forgive me, that's just how it is.

I was really lucky in that I've always loved mathematics and started loving computing when I was about 14. Around that time, I also found out that my local university had a combined Bachelor of Mathematics and Bachelor of Computer Science double degree. So, I knew from then what I wanted to do when I got to university.

After finishing my undergraduate degrees, I needed to decide whether to complete honours in mathematics or computer science. Mathematics is my first love, but I was more interested in pure than applied maths, so I chose computer science because it has more immediate practical applications than most pure mathematics. I chose an honours project related to one of my favourite computing units: compiler design. My project was related to optimizing compilers, specifically optimizing HTML to reduce the size while keeping visual and semantic information as in tact as possible, mainly by optimising CSS classes to minimize the amount of style information that needed to be sent.

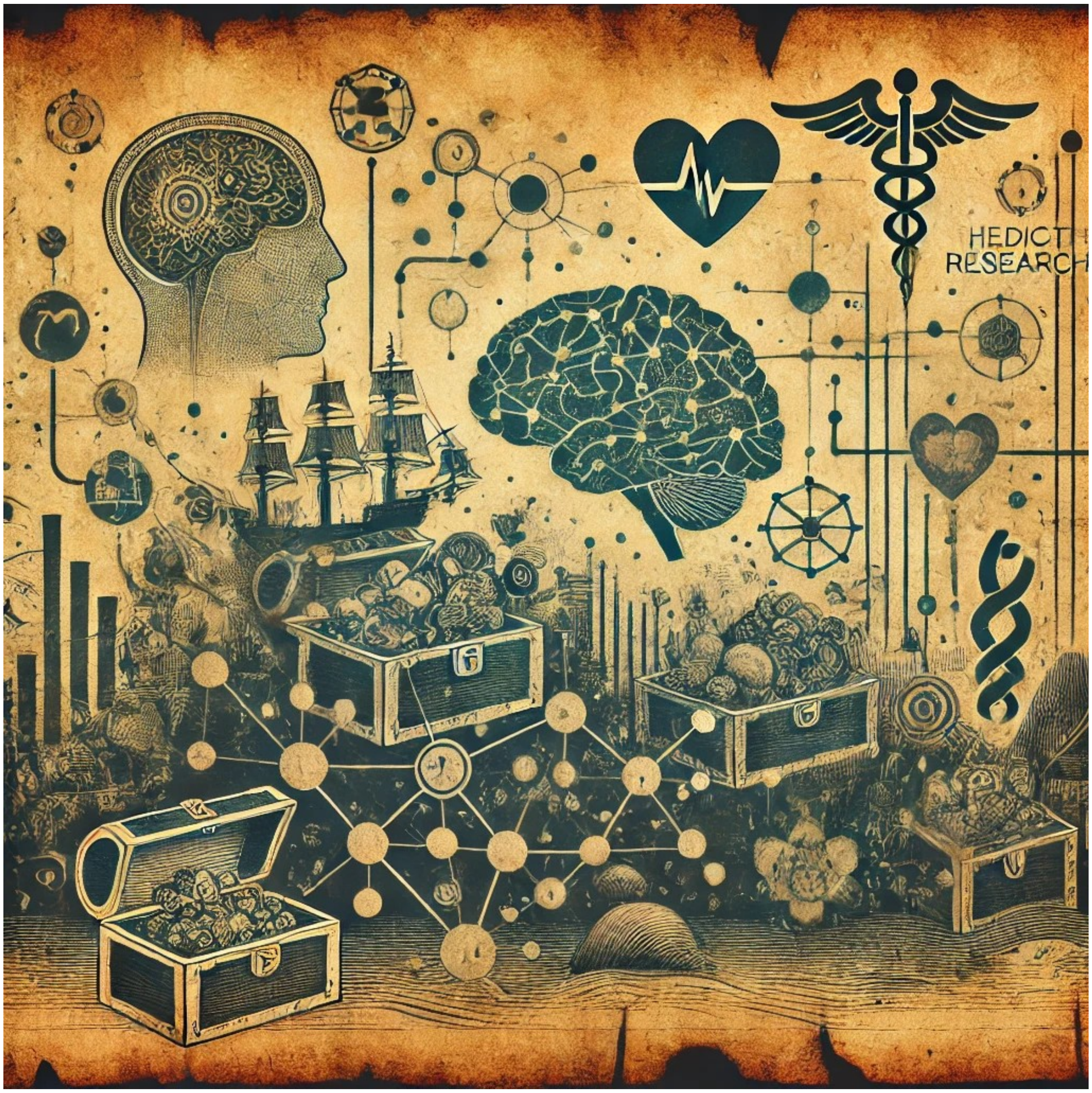
After that, I wasn't really sure what I wanted to do. I applied for a few jobs, but nothing I was really passionate about, so I didn't have any clear direction. Luckily, someone who had the office next to my honours supervisor managed to get a big grant in eHealth and needed a developer. So, I spent a year working with him before starting a PhD under his supervision while continuing with the eHealth work.

Today, I'll tell you about what I did in eHealth and my PhD. I continued working in eHealth until I got to the University of New England. At UNE, I quickly moved into agriculture, although I did a bit of eHealth here as well. Since then, I've done quite a lot, including sports science, which I'll talk about today. There are other paths I've gone on, in applied computing areas like technology acceptance and optimization, that I won't cover today.

In this talk, I'll focus on eHealth, my PhD, agriculture and sports science, sharing some of the things I've done, how they've led to other opportunities, and how my experience has guided me. This gives an example of what a research path can look like when you don't really plan ahead but follow your nose and seize good opportunities when they arise.

---

eHealth



eHealth

On Friday, I submitted my honours thesis, and the following Monday, I started working for the Australian Schizophrenia Research Bank (ASRB). Schizophrenia is a mental illness that we don't fully understand, and while it has genetic components, that's not the whole story. Additionally, the tests needed for research in this area, like MRIs, are expensive. Many studies only get funding for about 20 MRIs, which includes both people with the condition and matched controls. With such a small sample size, the results are not statistically strong.

The idea behind the ASRB was to combine data collections of MRIs from around Australia into a single collection that researchers could access. We aimed to use grid computing for this, which has since been largely superseded by cloud computing. The idea was that you could access computing resources like you access the electricity grid—without knowing the exact source.

At the time, there was also the concept of a data grid. The idea was to request and access data seamlessly. The grant my boss received was to develop a grid-based system for the ASRB. However, when I looked into the tools available, there was plenty of support for compute resources, but almost nothing for data management.

We envisioned a system where you could search for an MRI of a 24-year-old woman with schizophrenia and a matched control, without knowing which database the data came from. Access would be based on permissions granted by different sites. Unfortunately, getting agreement from the Human Research Ethics Committees at all the different sites proved impossible.

So, I spent a year developing a prototype system using PostgreSQL and the Globus Toolkit for grid computing. The key requirement was role-based access control, which PostgreSQL didn't have built-in at the time. I created a workaround that was functional but not production-ready. The bigger issue was the lack of agreements from the various data-holding institutions to combine their data into a single system.

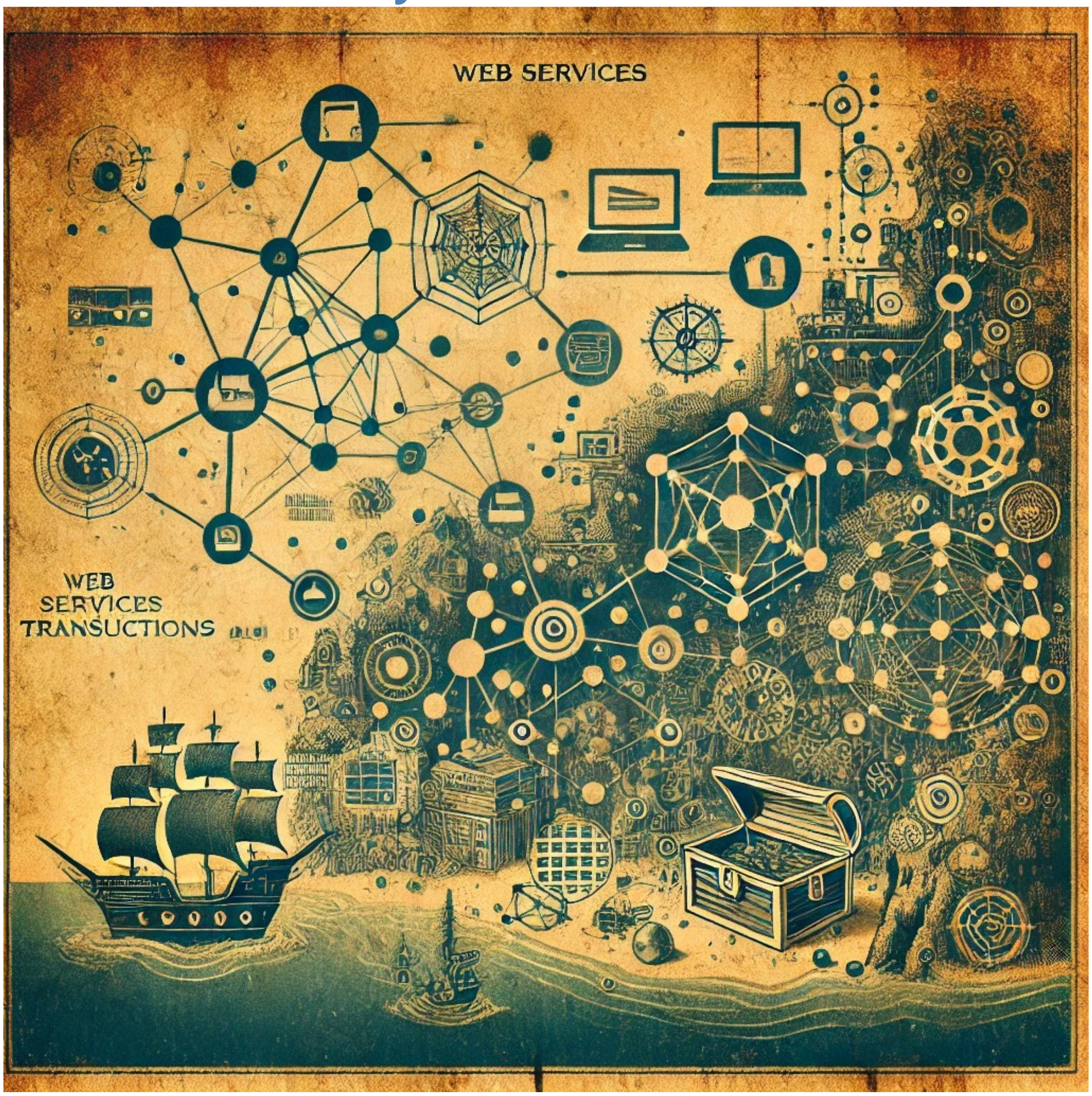
After the initial grant ended, we decided I would start my PhD. However, the ASRB received more grants to collect data into a single, large collection. We collected data from about 4,000 people, including MRI scans, blood samples, and detailed clinical interviews, ensuring privacy and data security.

While starting my PhD, I continued working casually on the ASRB until I finished my PhD and beyond. Additionally, I joined the Health Behavior Research Group at the University of Newcastle, where we developed an online survey system for medical research. This system could handle stratified, randomized controlled trials and adaptive questioning.

One project involved a smoking cessation system that was personalised by user responses, while another addressed issues with appointment attendance. We created a robust survey system that allowed multiple people to input data and adjust questions based on previous answers. This led to several good publications.

In parallel to my eHealth work, I pursued my PhD.

PhD in Distributed Systems



Web Services Transactions

I earned my PhD in distributed systems. The idea was to develop something helpful for the ASRB but not reliant on it. However, it quickly became clear that what was done at the ASRB wasn't enough to build a PhD thesis on, so I shifted focus to web services transactions.

Web services transactions rely on web services, which are requests made over the web that receive responses. Before this, transactions were mostly in databases and typically only ran for a short time. With web services transactions, you don't control the systems you communicate with, and processes can take much longer. For example, requesting the manufacture of a new product involves production and delivery, which can take weeks or months.

Traditional database transactions follow the ACID properties: Atomicity (an operation either happens completely or not at all), Consistency (the system remains in a consistent state), Isolation (transactions don't interfere with each other), and Durability (once a change is made, it remains). However, web services transactions often require relaxing these properties.

For instance, you might book a hotel and cancel within a few days for a full refund, or book something without paying upfront. Existing systems like BPEL supported some of these concepts, but there wasn't enough flexibility.

I identified about 150 different interactions possible with web services and developed a system where clients could negotiate with service providers to determine the level of transaction support each was willing to provide. This allowed clients to assemble complete transactions that met their needs and ensured that, even if everything didn't work perfectly, their losses were limited to an acceptable amount.

After finishing my PhD, I continued working on some eHealth projects before moving to UNE.

Agricultural Applications

Digital Agriculture

When I came to UNE, it was a busy time for computer science as we were redeveloping all of our units for the courses. The courses had been restructured before I got here, but the new units hadn't been developed. I was given three units to develop while teaching out three of the old units, making it a busy time.

Research is also part of the job, and I quickly got involved in agriculture. Given UNE's strengths, that shouldn't be surprising, but it was an area I never would have thought of before coming to UNE.

In particular, the Sheep CRC wanted systems to help farmers make decisions, and we received grants for a system called ASKBILL. A farmer's job is to manage risk, so we needed to provide them with useful information to help them make decisions.

We developed a system that collects on-farm data and includes extensive weather data—both past and predictive. One of my roles was integrating historical weather data from the 1980s to the 2010s into the system in a usable format. The system also includes results from models of pasture growth, parasite infection, and other relevant factors.

For example, instead of simply telling farmers that it will be cold in winter, we provide specific information like, “Next Thursday there’s going to be a cold snap, so you should delay shearing your sheep until the following Monday.” Or, “If you supplement feed for the sheep this week, you’ll have enough pasture to last through the winter.” This kind of targeted advice is incredibly useful.

ASKBILL combines disparate datasets to provide actionable insights, similar to how the ASRB combined blood work, MRI scans, and clinical interviews. ASKBILL automates this process and sends notifications to farmers, alerting them to unusual risks and helping them make informed decisions.

I’ve continued working in agriculture, using machine learning and computer vision. For example, we’ve developed biometric identification of cows by photographing their muzzles, which are unique like human fingerprints. This allows us to identify individual cows.

I’ve also worked with PhD students on weed detection in pastures. This is challenging because green weeds grow among green grass. The goal is to identify weeds so we can target them with herbicides or mechanical devices, leaving the useful crops untouched.

These projects are things I never would have thought of doing, but they are clearly very useful and impactful.

Sports Science

Sports Science

So, I continued working in agriculture for a while. My work with the Sheep CRC finished around early 2020, just as COVID hit. The CRC had completed its term, and my involvement ended as well. I needed new research projects, and while I had plenty of PhD students, I lacked a clear direction for my own work.

One day, the head of our school literally came running up to me and asked if I would be interested in working in sports science. Sports science researchers suddenly have access to a lot more data than they know how to handle. For example, in Rugby League, especially Women's Rugby League, elite players now wear devices during every game and often in training. These devices capture GPS positions, speed, accelerometer information, and even detect tackles.

Sports scientists have all this data but often don't know how to utilize it effectively. That's where I came in. Basic machine learning techniques, like clustering, helped us discover that women Rugby League players play quite differently from their male counterparts. For instance, in men's Rugby League, there are typically two main positions: forwards and backs. Training is tailored accordingly.

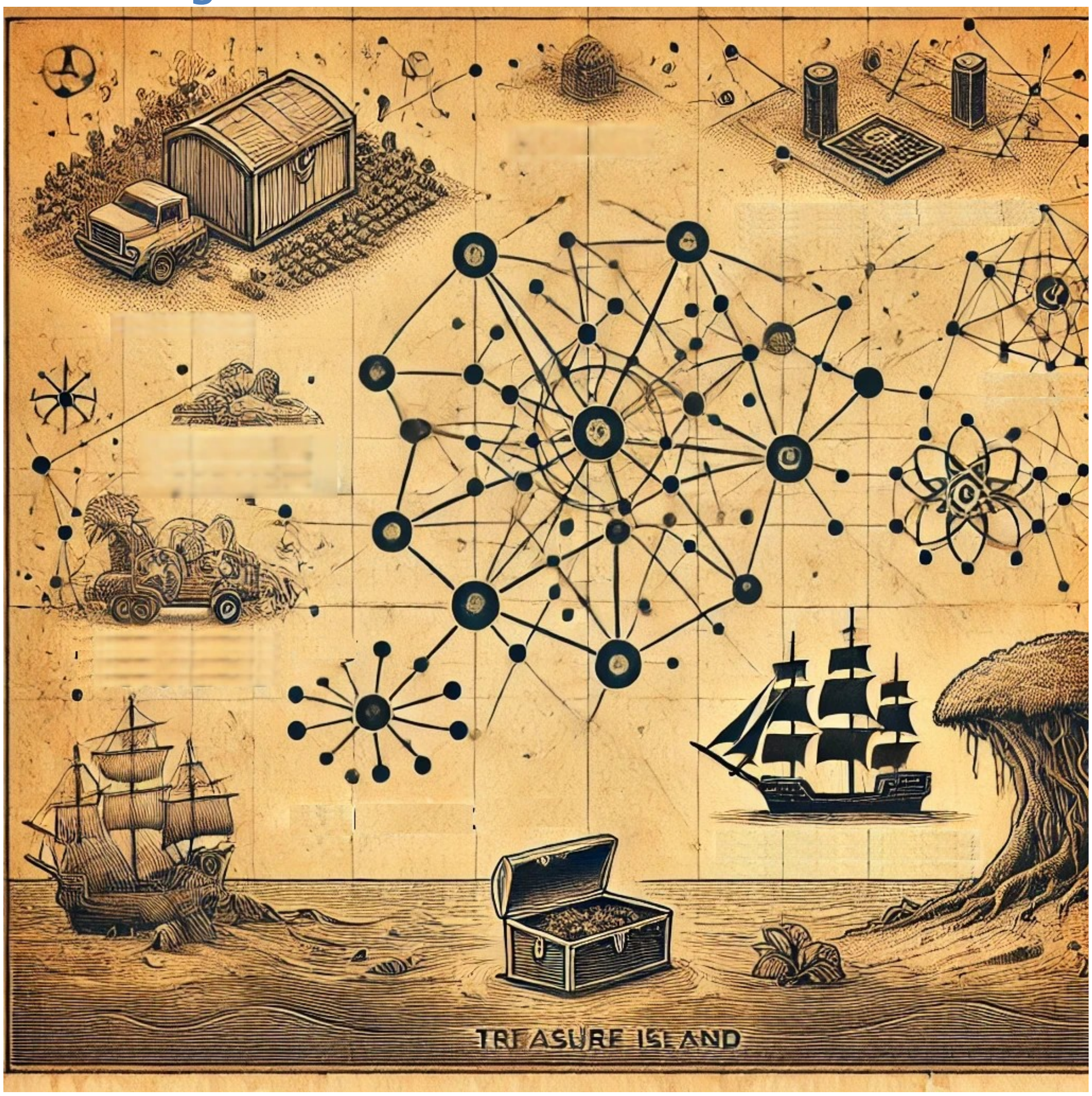
However, our data revealed that in Women's Rugby League, there are not just forwards and backs but also "adjustables"—players who fall between these two categories. Training these players as either forwards or backs alone wouldn't yield optimal performance. This insight led to altering training methods to better suit these unique roles, improving overall performance.

Another issue is that much of the existing research is based on men, even though women's sports have grown immensely. For example, the speed categories used to classify player movements (walking, jogging, running, very fast running) were based on male data. Women are different, so we analyzed data from an entire season of the National Rugby League for Women and established new speed ranges specific to female athletes. These new speed zones are now used by the National Rugby League to report statistics in women's matches.

This work has real-world impact. It helps ensure that we are training female athletes correctly, reducing the risk of injury by tailoring training to the actual demands of their game. There's a lot we can do in sports science, thanks to the availability of data. However, having data is different from having useful information. In computer science and data science, we are accustomed to working with large datasets to extract meaningful insights.

This is yet another area I never would have thought about working in before coming to UNE, but it has turned out to be a natural fit.

Connecting the Dots



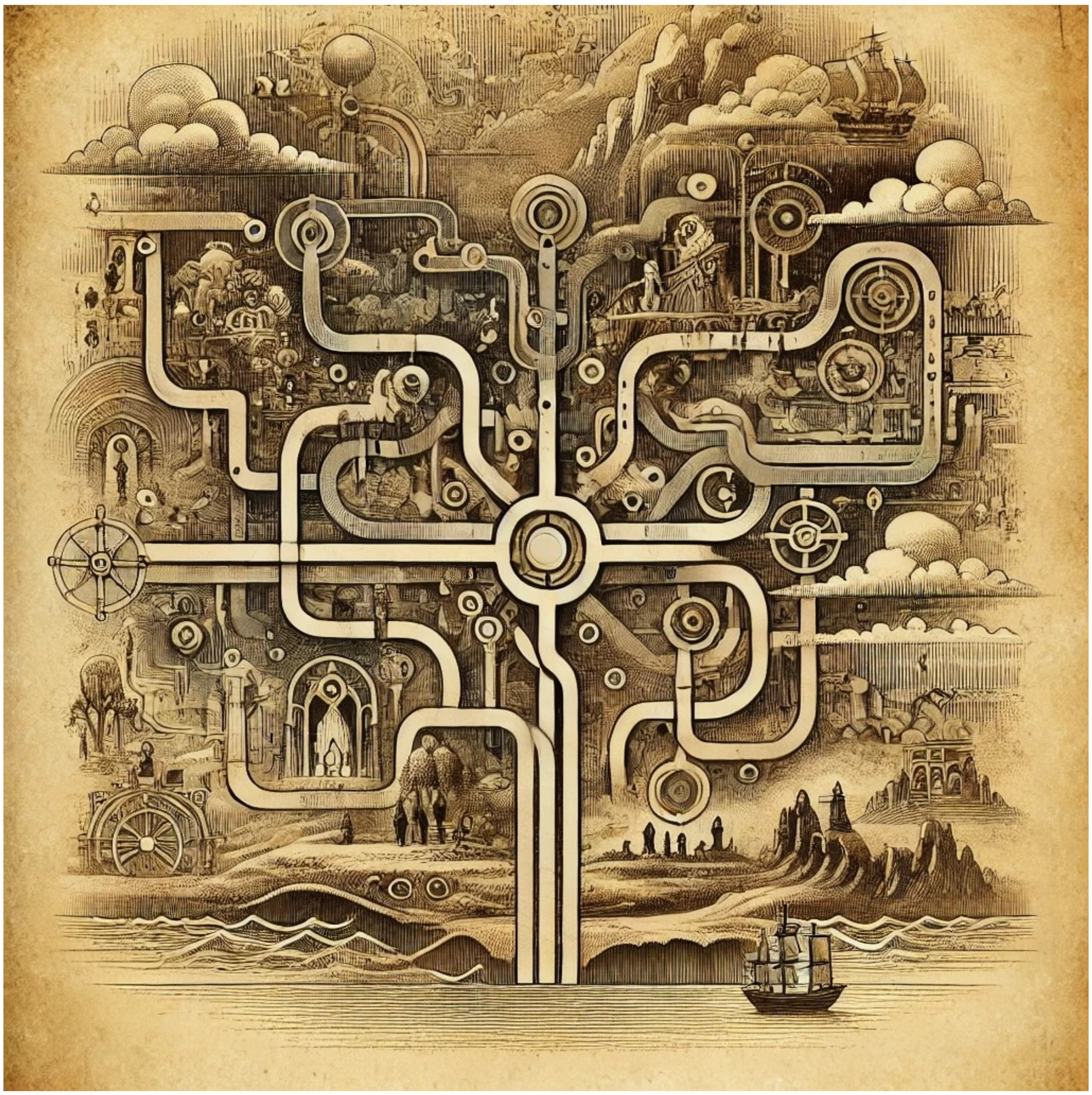
Applied Computing Connects Domains

My real takeaway here is that applied computing can enhance virtually any domain. Good computer science involvement can significantly improve outcomes across various fields. With the vast amounts of data available today, knowing how to use it properly is crucial. Collecting data alone isn't enough; we need to extract meaningful information from it.

This principle is evident in my eHealth work, where we collected large amounts of data for the Australian Schizophrenia Research Bank. In agriculture, combining disparate datasets intelligently can improve decision-making for farmers. Farmers are experts at risk management, but providing them with useful information can help them make even better decisions. This is what ASKBILL was all about.

In sports science, data helps us determine the best ways to train athletes, improve performance, reduce injuries, and create a better environment for everyone. In professional sports, watching top athletes is more enjoyable for fans and brings in significant revenue. If we can reduce injuries and keep star athletes in the game, it benefits everyone.

Conclusion



Pathways

In conclusion, I haven't really had a fixed research direction. I've stumbled into opportunities as they came along, and if they seemed useful, I pursued them. I believe I've accomplished some fairly useful things. For example, the Australian Schizophrenia Research Bank (ASRB) started around 2008 or 2009 and continued until 2022. Unfortunately, it's often easier to get government funding to build new systems rather than maintain existing, useful ones, but we had over a decade of providing valuable information to researchers.

My computer science skills can be applied to virtually any area. However, I have found it challenging to publish in cross-disciplinary areas. For instance, with the ASRB, health venues found our work too technical, while computer science venues found it not technical enough. Similar issues have arisen in other areas as well. Despite this, we achieve useful results and make significant contributions, which I believe is important.

I'm always open to exploring new areas because I haven't had a fixed research direction. I've worked on various topics like technology acceptance and software-defined networking, often with PhD students. I'm open to working in disciplines I never thought I would, and I'm willing to help wherever needed.

Thank you. If you have any questions, please feel free to ask them now.