**Centre for WHS – iLab information 2022**

**Can you introduce yourself?**

* The Centre for WHS mission is to deliver solutions that prevent harm in the workplace. It’s as simple as that – and we do this in a number of ways:
  + the traditional research process partnering with academics,
  + the less traditional R&D pathway of partnering with developers and innovators,
  + and through data science, thanks to a team of experienced data scientists we have in-house, who will certainly provide support to the students during their time with us.
* If you check our website <https://www.centreforwhs.nsw.gov.au/> you will see a number of safety innovations we have delivered in the last few years such as:
  + the world-first real time silica detector

<https://www.centreforwhs.nsw.gov.au/Projects/current-projects/silica-detection>

* + guidelines for BIM

<https://www.centreforwhs.nsw.gov.au/Projects/completed-projects/building-information-modelling-for-whs-management>

* + AI scorecard - a tool to identify and assess risks of AI systems in the workplace.

<https://www.centreforwhs.nsw.gov.au/tool-box/ethical-use-of-artificial-intelligence-in-the-workplace-final-report>

* ‘Safety innovations’ are our end goal with everything we do. We identify a need and bring together industry, government, academic, and development partners to come up with the best way to fill the gap and ultimately prevent harm.

**What is the challenge you have put forward for iLabs?**

* Simply put, we provide the students with a data science problem, and we expect them to come back to us with a solution.
* The challenge is part of a new project, which aims to provide business intelligence to SafeWork NSW (the regulator for WHS in NSW).
* The need here is for the regulator to improve its use of data intelligence to better inform decisions at a strategic and operational level.
* We have already allocated a whole team towards this project, and at the moment, we are exploring the different ways we can collect and consume relevant data to inform this business intel function; what intelligence can we get from:
  + Existing databases we have access to
  + Data insights reports published by other agencies nationally and internationally
  + Social media
  + Operational data – what our inspector are seeing on the ground
  + Grey and academic literature published nationally and internationally.
* That is the main focus of this student challenge we are proposing:
  + As a stating point, we are providing you with a real life dataset. It is a list of abstracts from academic literature that we collected from searching a scholarly database: it is gross, it is unstructured, it is difficult to analyse and interpret.
  + Our expectation with this student challenge is that 1) you will find out an efficient way to analyse this data and create relevant intelligence and 2) potentially, find a way to improve the way we collect this type of data in a first place.
* What does success look like? There are 2 key words when defining success here:
  + Quality = we need to be able to collect the relevant info out there, and analyse it to create meaningful intelligence.
  + Sustainability = resource friendly – here we would love to see a process of data collection and analysis that is less manual, and easier to achieve.
* While the student challenge focuses on something very specific – we will also be welcoming any contributions they would like to make, for instance in identifying opportunities to explore other data sources.
* You will see this with the example dataset we have provided, which represents academic literature in the WHS field, drawn from the publicly available PubMed database (https://pubmed.ncbi.nlm.nih.gov/).
* See “ the data” section for information about the datasets.

**Why is this important?**

* The student challenge is central in this business intelligence function we are setting up for SafeWork NSW.
* The expectation is that through this BI function, Safework will be better informed to make decisions at the strategic/exec level (e.g., what types of harms should we focus next? What types of practices should we introduced our inspectors to?) and at an operational level (new standards or codes of practice published in Australia or elsewhere).
* If Safework is better informed, it can better anticipate the new and emerging ways of working as well as future work practices. If SafeWork can get on top of those new and emerging practices in time, it can regulate them better, or raise awareness about how to prevent potential harms, before those practices become standard in the industry.
* If SafeWork is better informed, it can allocate resources to high-risk industries and businesses more efficiently.
* This student challenge, ultimately, will help to save lives of the workers in NSW, by improving the way WHS is regulated in NSW, and doing so, by improving the prevention of harm.

**Do you have potential directions or inspirations for student teams to take up?**

* All we are providing to the students is a starting point (with a dataset we have created through a data collection process) and expectations.
* What we would love is for them to take this away, bring their innovative and challenge thinking into it, obviously their data science skills and surprise us with something 1) that is of quality = relevant = insightful and 2) sustainable.

**Do you have specific resources or domain knowledge that you would like to share with students?**

* Students can get a feel of this by looking into the Work Health and Safety Roadmap published by SafeWork NSW (<https://www.safework.nsw.gov.au/roadmap#home>) and also the Centre for WHS Blueprint (<https://www.centreforwhs.nsw.gov.au/about-us/our-approach/research-blueprint>) which outlines the priority areas we are after .
* There is a whole project team that will be working on other part of this BI function in parallel so there is an opportunity for the students to get involved in that work if they are interested.
* Students can also have access to the data science work we have done in the last few years, for instance we developed a 5 star compliance rating for all businesses in NSW, which we computed using predictive analytics and machine learning, and gives SafeWork NSW a mean to better allocate resources to businesses at most risks of WHS compliance.
* Students can learn more about the Centre’s projects on our website. [www.centreforwhs.nsw.gov.au](http://www.centreforwhs.nsw.gov.au) .

**Are you looking for specific skills and attributes in student teams that work on your project?**

* We’re eager to work with students who are open to trying new methods and challenging the status quo. They should be excited by new possibilities and not afraid to ask questions and seek answers to tricky problems.
* Most of all, we want to work with students who aren’t afraid to suggest smarter approaches, beyond the limitations of existing tools and technologies.
* We pride ourselves on taking a multi-disciplinary, collaborative approach in the Centre, so this project will suit students who will thrive while working across cross-functional teams.

**What skills can MDSI students develop during the project? How do you think it can help prepare students for their careers/ how does it apply to the wider data science community?**

* Students will experience what data science means in the real world – the fun part is the analytics part, but as a data scientist, you are also often best placed to drive and improve data collection processes, and improve the quality and format of data you will analyse.
* They will explore real-world data which are a lot messier and less structured than data you typically get through a curriculum.
* They’ll also have the opportunity to develop a range of practical hard and soft skills that comes with the role of data scientist in any organisations
  + Data sourcing and processing
  + Data analysis, modelling, and visualisation
  + Use of statistical software and data tools and programming languages
  + Git, which is a source code management tool
  + Problem solving
  + Critical thinking
  + Collaboration.

**Anything else you would like to add?**

* At the Centre for WHS, we really appreciate the value of the unique perspectives students bring.
* This year, we launched our own Academic Studies Program, which brings funds, resources and opportunities to increase collaborations with students and bring their innovative and challenging thinking into WHS.
* We are hosting a National WHS Colloquium in Sydney this December. We would love for all the students that have partnered with us to have the opportunity to present their work at this forum, if this is something they want.
* There will also be networking opportunities with other students, as well as research and data science professionals from across Australia.

**The data**

Students are provided with the following files:

|  |  |  |
| --- | --- | --- |
| **File** | **Description** | **Note** |
| Abstract-Home care service.txt | A txt file resulting from a PubMed1 search following the conditions laid out under **Aged / Disability care** in the *Literature Search Strategy* document. This content is aligned with a program of work including in-home aged and disability care, the harms to which in-home care workers are exposed, and the potential solutions that seek to prevent harm. | This is the output from the search conducted on 14th July 2022. |
| Abstract-MHW last 10 years.txt | A txt file resulting from a PubMed search following the conditions laid out under **wellbeing / mental health** in the *Literature Search Strategy* document. This content is aligned with a program of work including mental health, wellbeing, and linguistic analysis within the Centre for WHS. This includes the harms to mental health and wellbeing to which workers are exposed, and the potential solutions that seek to prevent harm. | This is the output from the search conducted on 14th July 2022. |
| Literature Search Strategy | A document that outlines the literature search strategy for each of the relevant topic areas, being:   * Wellbeing / mental health * Aged/Disability care * Scaffold safety / construction * Culturally and linguistically diverse workers’ WHS | You will note that aligned datasets have been provided for the topics of Wellbeing / mental health AND Aged/Disability care only.  The remaining two potential datasets may be developed by students using literature databases excluding PubMed, due to the topic areas being non-clinical. Other sources may include Scopus2, ResearchGate3, Google Scholar4 etc. The literature search strategy should be used as a guide to source grey and academic literature and inform the outputs of these literature searches. |
| Centre for WHS information | A document that contains updated copy of the pitch provided by Dr Gregory Zelic, and instructional information regarding the datasets provided. |  |

*1 PubMed is a free resource and search engine for biomedical and life science literature launched to the internet by the US National Library of Medicine. The PubMed database provides more than 34 million citations and abstracts of biomedical literature.*

*2 Scopus is subscription-based database providing access to abstract and citation across a number of disciplines, not only in the Health Sciences and Life Sciences, but also from the Physical Sciences, Social Sciences and Humanities. Scopus includes records from PubMed and EMBASE (European equivalent to PubMed) and other sources. Scopus has more than double the number of records in PubMed. Scopus does not only provide search for articles, but also book, book chapters and conference abstracts.*

*3 ResearchGate is a social networking site for scientists and researchers to share papers, ask and answer questions and collaborate. Its search functionality is limited to keyword, title and author.*

*4 Google scholar is a free web search engine with a focus on scholarly literature across many publishing format and disciplines (covering Biology, life sciences, environmental sciences, business, engineering to social sciences, arts and humanities).*