**Proposal Transcript**

Hello I’m Ian and first let me say welcome to my presentation hopefully what we have planned will push forward the knowledge in the area of Covid19 Data and Data presentation. There is a famous quote

“Research is seeing what everybody else has seen and thinking what nobody else has thought.”- Albert Szent-Györgyi

In this presentation will be charting my thinking on the research topic while defining the research proposal and how these fits into the wider understanding of the research area. I will also clarify the area of understanding that this research will study along with the scope key objectives and deliverables that will be derived from this study.

The deficit we plan to investigate will cover areas such as the problems with the existing solutions, identifying the areas that currently have a weakness that our research can address. We will also look at factors that could limit the scope of the proposed research and we will also address any ethical framework’s that will need to be taken into account when designing our research proposal. Finally, we will look at a project timeline on how our research can be delivered in an effective way maximising outcomes along with the key literature that our research will use as its starting point.

**Problems with Current Solutions**

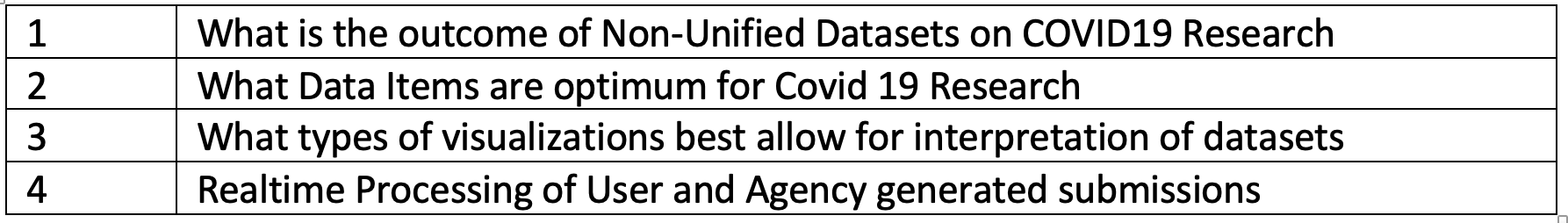
Covid19 research was not an area of research that had a well established track record or research plan with research either being born out of necessity or existing research programs being recast and re-engineered to study covid 19. Because of this combined with the world wide nature of the Covid 19 some of the problems I observed while carrying out the review of available literature were

As the supply of data crossed national and international boundaries there is not one unified way of collecting the data for analysis some data is pre-processed and supplied to the agency’s carrying out the data analysis as flat files (CSV) while others use other transport mechanisms such as SOAP & JSOSN APIs. Other issues identified with the current solutions is in regard to the data being captured. There is not a common agreed dataset that all parties provide this makes doing any meaningful analysis between different jurisdictions challenging. By having a common data specification, it allows easier and more detailed analysis to be undertaken between entities. Examples of this could be seen through the pandemic where the public were told to disregard excess death, infection comparisons between countries as the data were not calculated in the same way.

Another issue identified is the lag between a subject or responsible health authority reporting and the data being available for study research and inclusion in COVID 19 dashboards. A Real time data processing mechanism allowing for the streamline ingest of data would allow for a more Realtime view of the current data and this could then be used to inform Realtime decision making.

**Research Topics**

There are a number of topics that we plan to research



First of these is What is the outcome of Non-Unified Datasets on COVID19 Research is the use of non-standardized datasets having an adverse effect on COVID19 research. Along with the ability of data visualizations to show true comparisons between datasets, this topic of study will allow to see if having a single common dataset has practical advantages over the current situation where each jurisdiction dataset may differ in what data is captured. The result of this study could inform how the international community handles future international data sharing in areas such as pandemics and other fields of global impact such as the impact of climate change.

The second topic we will look at is what data items are the key ones for a covid 19 dataset this ties in with topic one by looking at what is the minimal viable set of data items needed to be included for users and researches to use the dataset for both understanding Covid19 and for any research requirements the outcome of this would be a set of data that would make up a MVP (Minimal Viable product) for capturing reporting on and studying these two topics when combined would add to the global knowledge of data required to support covid 19 reporting for both the general public and research realms.

The third area of study as part of this research proposal is to look at the types of visualizations and which types are best for displaying and relaying information. Throughout the initial phases of the Covid19 pandemic a number of different infographics where used to describe the data being captured for this part of the proposal we intend to take a more analytical look at these to see if there are any lessons that can be learned and what best practice looks like so this can be applied to future situations where complex datasets need to be presented to all parts of society with vastly different backgrounds both in terms of usability to understanding and interpret the data and their ability to draw reasoned conclusions from the data being presented.

The final area of study in this proposal looks at the value of real time data processing does the ability to have data in close to real time enhance the usefulness of covid19 and dashboards the answer we hope to get out of this study is to see if the time from collection to delivery should be prioritized over data items and data processing. Or if the effect of latency of data being provided to the dashboard is less important than the quality of the data being provided. This a vital question as this can be used to shape future research work to maximise the impact of data being collected.

**Documentation and Research**

The project will take account of the key pieces of research currently in the field

**Dong, E., Ratcliff, J., Goyea, T. D., Katz, A., Lau, R., Ng, T. K., Garcia, B., Bolt, E., Prata, S., Zhang, D., Murray, R. C., Blake, M. R., Du, H., Ganjkhanloo, F., Ahmadi, F., Williams, J., Choudhury, S., & Gardner, L. M. (n.d.). The Johns Hopkins University Center for Systems Science and Engineering COVID-19 Dashboard: data collection process, challenges faced, and lessons learned.**

This study lays out the ground work for the current state of covid 19 dashboard development and the challenges faced when creating COVID19 dashboards.

***Falisse, J.-B., & McAteer, B. (2022). Visualising policy responses during health emergencies. Learning from the COVID-19 policy trackers. Convergence, 28(1), 35–51.***

Describe how data and data visualizations can be used during a pandemic and the use of visualizations during the pandemic these two papers will be the foundation of our proposal that we will build on with our own research.

**What do we want to achieve & deliverables?**

Our primary aim is to improve the pool of knowledge in the area of COVID19 data and the use of this data for display and research

Almost all data has a Geography component in “William Huxhold’s 1991 book ‘An Introduction to Urban Geographic Information Systems’ reported the results of a needs analysis for an urban geographic information system (GIS) eighty to ninety percent of all the information collected and used was related to geography.’”

We want to leverage this fact as part of our project and to use all the tools available we hope to create a data source that has GIS data as part of the dataset. As far as physical deliverables this proposal intends to produce the following assets.

* REST API (JSON , GRAPH)
* Data Cleaning procedures / techniques
* DBMS DESIGN (SQL , NOSQL)
* Reports & Dashboard Designs
* Platform Architecture Design
* HOTS (Hand Over to Support) Document

Or first deliverable is an API to allow the automated submission of data automation is a key part of the research the more we can automate the process the more efficient the process will become and the close we will get to having a near real time system. The second item that we want to deliver is a set of processes to clean data being submitted data quality is of vital importance when dealing with medical data as in the case of COVID19 so as part of our proposal we intend to create processes to clean and validate the data as part of the ingest pipeline to ensure that the data that is used to generate dashboards and used for analysis is of the highest quality and can be trusted as a true representation.

Our next deliverable will be a DBMS design that optimizes the collection and analysis of data we will look at traditional relational DBMS systems and NOSQL systems to determine the optimum configuration for the storage and retrieval of COVID19 data. Combined with this we will ensure that any proposed solutions architecture can scale to meet the demands that could be imposed if the system was to be used to manage a future worldwide pandemic.

The final deliverable will be a HOTS (Hand over to support) document as the ultimate aim of the research is to have a practical real world application and this will allow for a smooth transition from the research realm to a real implementation.

**Ethical considerations and risk assessment**

While undertaking this research there are a number of ethical considerations, we need to take account of. Being based in Europe I will be looking at these from a viewpoint of a European research project if however, this project was to be conducted in another jurisdiction or cross international boundaries consideration will also be required for those areas.

The first thing we need to consider is GDPR for any data we process proper consideration needs to be given to the data we hold and if it complies with GDPR (2018) rules and regulations. We also need to take into account UK DPA (Data Protection Regulations 2018) to ensure that data is collected in a way that is consistent with rules on data collection.

We also need to ensure that any system built meets current international security guidelines and is designed in a way that data is protected though out the data processing pipeline. To Ensure this is the case an Information governance (IG) review will be carried out post implementation to ensure all data meets these requirements and we comply with all legislation.

**Project Limitations**

Limitations that we could foresee with the proposed research / project. One of these would be getting the Data Protection Impact Assessment in place to allow sharing of raw data with primary care trusts there is some distrust of the European / US centric research industry by countries outside of the 1st World. This is however not unsurmountable but would require some education to partners on what we hope to achieve and how we would respect their submissions to the project. With COVID19 there are also different priorities between countries in regard to data analysis with what we hope to achieve hopefully going forward once our outputs are released the benefit of what we are trying to achieve is seen any resistance will be removed.

**Methodology**

We will be using a number of different Methodology’s these will include the following

1. Reviews of existing processes
2. Dialogues with SMEs in the field
3. Focus Groups where appropriate
4. Trials to determine optimum solution
5. Secondary Data Examination

From my initial design of the research proposal these are the methods that if used correctly will give the best outcomes, first of all taking into account the current landscape in regard to data capture and use and by getting input from experts in the field so we capture what is actually needed rather than just a theory of what is required. This will be reenforced by the use of trials runs in a SDLC feedback loop to arrive at the optimum solution. This method of running the project does have a larger resource requirement but we feel this is justified by the expected superior outcomes.

During this process I will also be using quantitative techniques to gather accurate and independently validated data that can be used during the investigation and implementation of this project.

**TimeLine’s**

Having looked at the requirements of the research and weighed up the options. The research currently being proposed would have a small team of researchers, the approach we plan to adopt is a waterfall approach with the research being split into 5 Pillars.

1. Requirements
2. Design
3. Implementation
4. Confirmation
5. Maintenance / Reworking

Stage 1 (Requirements) would be assigned a 10 week slot this will be used to communicate with stakeholders and gather how their current data analysis teams work and what data they currently collect as part of their covid19 dashboards.

Stage 2 (Design) 15 Weeks we will take what we learned in stage one along with our own research and produce a design that allows the aims of our research project to be realised in a system / artifact this stage along with Stage 1 are the foundations of our research program which is why a majority of the time is allocated to these first two steps.

Stage 3 (Implementation) this will be assigned a 5 week slot this is where the outcome of Stage 1 and Stage 2 will be turned into a tangible product made up of an Architecture Design Database ERDs, API Specification and Security Review of proposed deployment. It is also where infrastructure will be stood up so allow real world demos and UAT of our platform design to take place.

Stage 4 (Confirmation) 2 Weeks we will confirm the system operation with the system up and running we will try to answer our original research questions using our managed dataset.

Stage 5 (Maintenance) if anything has come out of Stage 4 it will be remedied, we also will carry out a lesson learnt exercise to study what lessons have been learnt during the project. This is also the stage where formal documentation will be produced documenting the project outcomes.

**References**

Dong, E., Ratcliff, J., Goyea, T. D., Katz, A., Lau, R., Ng, T. K., Garcia, B., Bolt, E., Prata, S., Zhang, D., Murray, R. C., Blake, M. R., Du, H., Ganjkhanloo, F., Ahmadi, F., Williams, J., Choudhury, S., & Gardner, L. M. (n.d.). The Johns Hopkins University Center for Systems Science and Engineering COVID-19 Dashboard: data collection process, challenges faced, and lessons learned.

Falisse, J.-B., & McAteer, B. (2022). Visualising policy responses during health emergencies. Learning from the COVID-19 policy trackers. Convergence, 28(1), 35–51. <https://doi.org/10.1177/13548565211048972>

Guide to the UK General Data Protection Regulation (UK GDPR) | ICO. (n.d.). Retrieved November 27, 2022, from https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/

Data Protection Act 2018. (n.d.). Retrieved November 27, 2022, from https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted

About the DPA 2018 | ICO. (n.d.). Retrieved November 27, 2022, from https://ico.org.uk/for-organisations/guide-to-data-protection/introduction-to-dpa-2018/about-the-dpa-2018/

Huxhold, William E.. “An introduction to urban geographic information systems.” (1991).

1957, Bioenergetics by Albert Szent Györgyi, (Epigraph for Part II: Biological Structures and Functions), Quote Page 57, Published by Academic Press, New York.