CRC for SPATIAL INFORMATION

Productivity Sector Commission Inquiry into Data Availability and Use Across the Public and Private Sectors
Terms of Reference

Submission by the Cooperative Research Centre for Spatial Information

Introduction

The Australia and New Zealand Cooperative Research Centre for Spatial Information (CRCSI) specialises in the development of spatial information and related technologies to generate economic growth in key industry areas. These technologies include global navigation satellite systems, remote sensing and earth observation systems and spatial data infrastructures and the associated analytics. Priority areas for our research include; health, agriculture, natural resource management, climate change, urban planning, and defence. The CRCSI has about 100 partnering organisations from government, the research sector and the private sector.

The CRCSI specialises in conducting research that maximises the use of data that is spatially or location enabled (technically known as geocoding), delivered through supply chains (processes) for the generation of new analysis methods or creation of valuable information products.

A priority for us, and our partners, is to ensure that data is 'location enabled'. 'Location' can be described as a precise geographic coordinate or less precisely as a place. The dynamic use of location manifests itself in navigation and tracking. This data, in static or dynamic form, is an important linking element for other data and information about people, the environment, built assets. In reality this covers a vast range of public and private activities and needs for the economy. Data that are enabled in this way are commonly referred to as 'spatial' or 'geospatial'.

A typical example in the health sector is a person's health record. When it is coupled with their location of residence (address) it is spatially enabled. This record can be geographically co-registered with a wide range of other data to unlock insights and correlate with environmental data, socio-demographic data, provision of health services and much more.

Key points

Our key points and recommendations are:

1. CRCSI research tells us that there is great benefit to be derived through improved linkages across different data types and that location enablement can effectively provide this linkage. Many important data streams in Australia are not spatially enabled yet.

2. Geocoding (spatial enablement) of data should be applied at point of capture where-ever possible to avoid additional costs later on. Data that is spatially enabled in this way has improved provenance (point of origin) and lineage (tracking how it is used and changes through time and with analysis) thereby improving both the elements of trust and value.

3. Data made openly available from public agencies should be supported by a strong policy of provisioning it through web services such as Application Programming Interface (API's'), increasingly as a continuous stream where-ever possible, rather than as a once-off or periodic supply that is rapidly out-of-date. Amongst other benefits is the stimulation of the web services industries and most importantly analytics and tools, without which the data has little chance of adding value to users.

4. Given the exponential improvements in technologies, and especially spatial technologies (including earth observation from satellites and airborne remote piloted vehicles, global navigation satellites systems and so on) value chains are enhanced when the linkages between the research community and users are optimised. Government policy and supporting regulations should be framed to ensure that research opportunities are enhanced and not impeded.

Research supporting the case for increasing data availability and use

The CRCSI is managing two core research programs that directly support the generation of infrastructure and systems that maximise data sharing, accessibility use and value-adding.

The first relates to the creation of 'Next generation spatial data infrastructures' (Program 3) focussed on semantic web enablement of supply chains. The second research program 'Rapid Spatial Analytics' (Program 2), is focused on improving workflows, rules and systems that can be utilised and reused with real-time data and other information sources to speed up the provision of information and outputs supporting day to day business decisions.

Program 3 – Next Generation Spatial Infrastructures: The aim of this research program is to move from a manual data supply approach (supplier push model) to a consumer driven, data demand (user pull model) approach. This requires substantial change in how government information is organised and a move towards semantic 'graph' data models providing rich self-description interrelations of data, enabling machine to machine linkage. The output of this research will be modern infrastructure, data and web models that will make the sharing of data less complex for organisations.

Some examples of the current outputs of this research relevant to this inquiry:

- Addressing: In 2015 the CRCSI commissioned Business Aspect and Mercury Project Solution to review geocoded addressing in Australia and to provide advice on how to optimise this supply chain. The report found that "*The supply chain is complex, non-linear and in many aspects convoluted, which creates contradictory evidence in applying confidence levels to address verification and geocoding processes using reference address files compiled from similar sources.*" The study highlights the complexity of coordinating inputs across three tiers of government for both public sector use and private use of a fundamentally important dataset – addressing.

  The report noted that only 11% of Australia's landmass is properly addressed although there is a wide range of use-cases where no addresses exist, e.g. in indigenous communities, greenfield developments, mine sites and marine environments. Until recently, the only way to reference locations in these situations was to use elaborate natural language descriptions, or X,Y coordinates which do not enhance end user usage. Modern semantic and spatial approaches to addressing can close this disparity.

- Spatial Data Management Ecosystem: New thinking has generated a revised framework that takes a supply chain view of all key elements of the data ecosystem.

  This framework outlined the method to capture and understand provenance of data as it moves through a supply chain informing the end user of the quality and trust. Provenance understanding provides a measure of trust to the data being accessed in providing evidence in decision making processes. This is critical to ensure that policies put into practice are not just supported by data, but the understanding of how that data was generated is also known.

CRCSI Submission to Productivity Sector Commission Inquiry into Data Availability and Use Across the Public and Private Sectors

Page 2 of 5

- Spatial Data Supply Chain and End User Framework (Research Paper) identifying drivers that support a move towards demand driven data access.

- National Data Grid**:** Summary of a modelling tool to integrate and spatially link data from multiple sources to support commonly required queries, analytics and modelling tasks.

Our research is working towards addressing the same recurring data-related themes identified in the commission's report:

- *Insufficient data sharing between agencies*; a problem that can be addressed by adoption of new 'semantic' tools and models. It is particularly important that agencies responsible for data storage and provisioning understand and adopt the means to improve machine to machine learning.

- *Insufficient data linkage*; demonstrating that 'location' enabled data provides new insights and methods for linkages based on location and the shared geographic relationship of otherwise disparate data. This can be especially useful where data is not commonly linked, for example a person's health records can be linked to the proximity to wetlands mosquito breeding areas.

Program 2 – Rapid Spatial Analytics: This program is conducting research that improves the ability and efficiency of government and industry to rapidly create valued information products using mobile and cloud infrastructure. The research builds on the outputs of Program 3 to realise an improved ecosystem of online data accessibility through automated systems. Current activities include:

- State of the Environment reporting**:** Cloud-based queries, analysis and visualisation of the current and historical state of environment reporting for routine strategic planning and decision-making in government.

- RAISE: Enabling land valuers, city councils, state government policy and industry to collaboratively explore and test hypotheses connected with the likely causes of land valuation changes in relation to infrastructure decisions.

The outputs of this research will enable Government to address;

- *Missed potential for stronger evidence-based policy*: By enabling a rapid and customised workflow for location-enabled data that can produce high value environmental information for statutory monitoring and reporting or more rapidly assessing the impact of infrastructure changes on commuter travel through web enabled sensors and core administration data.

The CRCSI is continuing to develop new methodologies to optimise data linkages through spatial techniques in the Health and Environmental fields.

CRCSI Submission to Productivity Sector Commission Inquiry into Data Availability and Use Across the Public and Private Sectors

Page 3 of 5

Examples of Applying Spatial Improvement within the Health Sector

HealthTracks™ & Epiphanee™
The CRCSI partnered with the Department of Health WA (DoHWA) to transform the way data was stored and made accessible for reporting tasks. This involved facilitating data linkages through reworked data management processes to enable the delivery of a standardised querying tool (HealthTracks) for the department. Starting from a base of half a dozen users, the Healthtracks system now has over 1000 users generating thousands of reports, critical for the operation of the Department.

HealthTracks identified a need to innovate how privacy was managed in Government to allow more flexibility while maintaining privacy. This led to Epiphanee, a data linkage system that employs a 'probabilistic method' to determine the risk of identification of an individual based on the user query parameters. The true value of Epiphanee that distinguishes it from other analytical tools is its novel approach to maintaining confidentiality protocols essential for assuring individual privacy. This allows agencies to mine sensitive data to a richer, deeper understanding than conventional privacy approaches without compromise to patient anonymity.

> **Epiphanee is on-the-fly reporting relating to the amount of disease counts identified in a specific cohort in a particular location.**

Epiphanee is an innovative example of how government can improve data access for both research use and evidence based policy by allowing users to query all available data whilst protecting the rights of individuals to anonymity through automated privacy protection checks.

Spatial Modelling
Another example of the use of spatial methodologies and the value of spatially enabled data was demonstrated through the Queensland Cancer Atlas project. The CRCSI, in partnership with the Queensland Cancer Council (QCC) and Queensland University of Technology (QUT), developed innovative statistical tools for spatial modelling of patterns in disease, with an emphasis on cancer outcomes (incidence, mortality and survival) providing significant insights into how health service catchment areas vary across time and focused on:

- Small area disease mapping
- Risk factor modelling and its impact on spatial inequalities
- Spatial models for health service utilisation and its relationship with cancer outcomes.
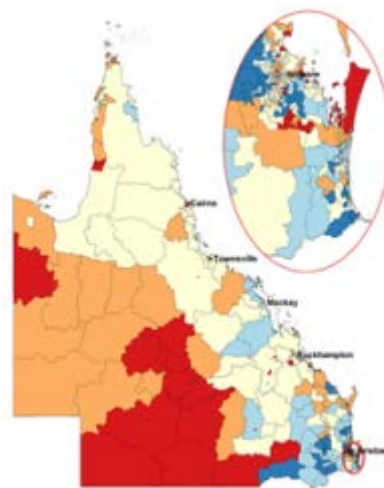
Application of this methodology to mammography data from Breast Screen Queensland provided significant insight into the demand and capacities of different services and how uptake varies spatially, both regionally and by individual service. It examined the influence of cancer stage at diagnosis, distance to treatment facilities and area-disadvantage on spatial survival inequalities for breast and colorectal cancer, and estimated the number of premature deaths due to non-diagnostic related spatial survival inequalities after adjusting for cancer stage at diagnosis.

CRCSI Submission to Productivity Sector Commission Inquiry into Data Availability and Use Across the Public and Private Sectors

Page 4 of 5

This work has underpinned significant changes to cancer services and public policy over the last four and a half years including substantial increased public funding for schemes to assist regional patients travel to access cancer treatment.

In addition, our research findings helped inform the Queensland Government's 2014 state-wide health service strategy, which includes an objective focused on reducing inequalities through specified actions as a priority for the state.

Due to the success of the Queensland Cancer Atlas, the research team at QUT and QCC, the CRCSI, the National Health Performance Authority, Cancer Council Australia and the Australian Institute of Health & Welfare have now embarked on the development of a national digital cancer atlas that we believe will provide significant insights into cancer incidence and survivability on a national scale.

The references below help demonstrate new thinking in how data infrastructures, systems, processes, data, security, access and visualisation assist the end user in maximising data for their purpose.

Thank you for the opportunity to submit this response.

Contact:

Dr Peter Woodgate
Chief Executive Officer
Cooperative Research Centre for Spatial Information

References:

- Program 3 – Next Generation Spatial Infrastructures. Overview of the core research topics: http://www.crcsi.com.au/research/3-spatial-infrastructures/
- Program 2 – Rapid Spatial Analytics: http://www.crcsi.com.au/research/2-rapid-spatial-analytics/
- Introducing Semantic Graph Data: http://www.linkeddatatools.com/introducing-rdf
- Optimising the supply Chain for Geocoded Addressing in Australia – Current State Supply Chain: http://www.crcsi.com.au/assets/Resources/Geocoded-Address-Supply-Chain-Review-Final-V6.0-Final.pdf
- Tracking Data Provenance in Spatial Data Supply Chains: http://www.crcsi.com.au/assets/Resources/Project-3.02-Tracking-Data-Provenance-in-Spatial-Data.pdf
- Epiphanee Project Overview: http://www.crcsi.com.au/impact/visualisation-and-analysis/

CRCSI Submission to Productivity Sector Commission Inquiry into Data Availability and Use Across the Public and Private Sectors

Page 5 of 5