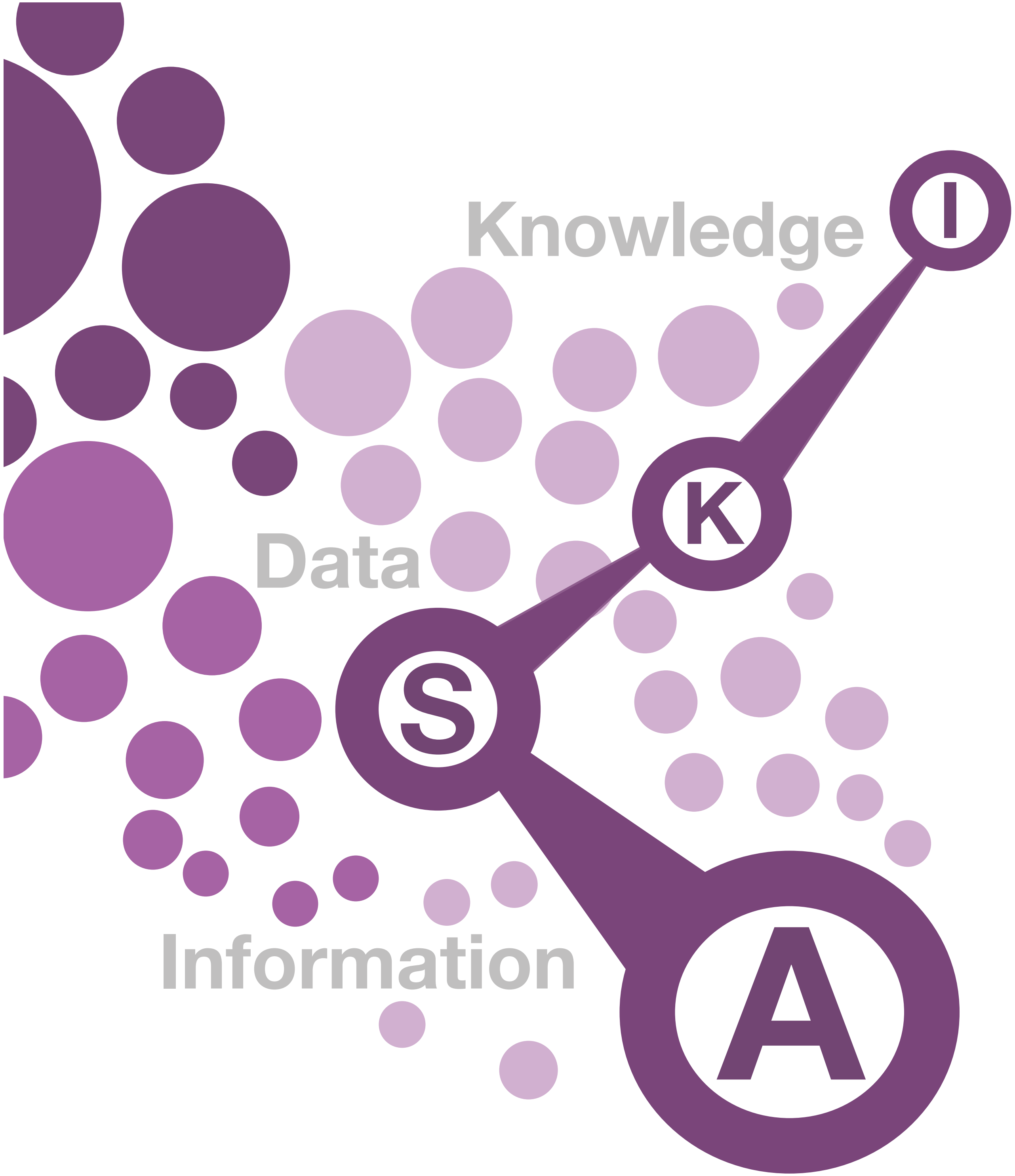


Towards a Spatial Knowledge Infrastructure

White Paper
March 2017

Australia and New Zealand
Cooperative Research Centre
for Spatial Information (CRCSI)

Production Notes:

All currency included in this white paper is Australian dollars unless stated otherwise.

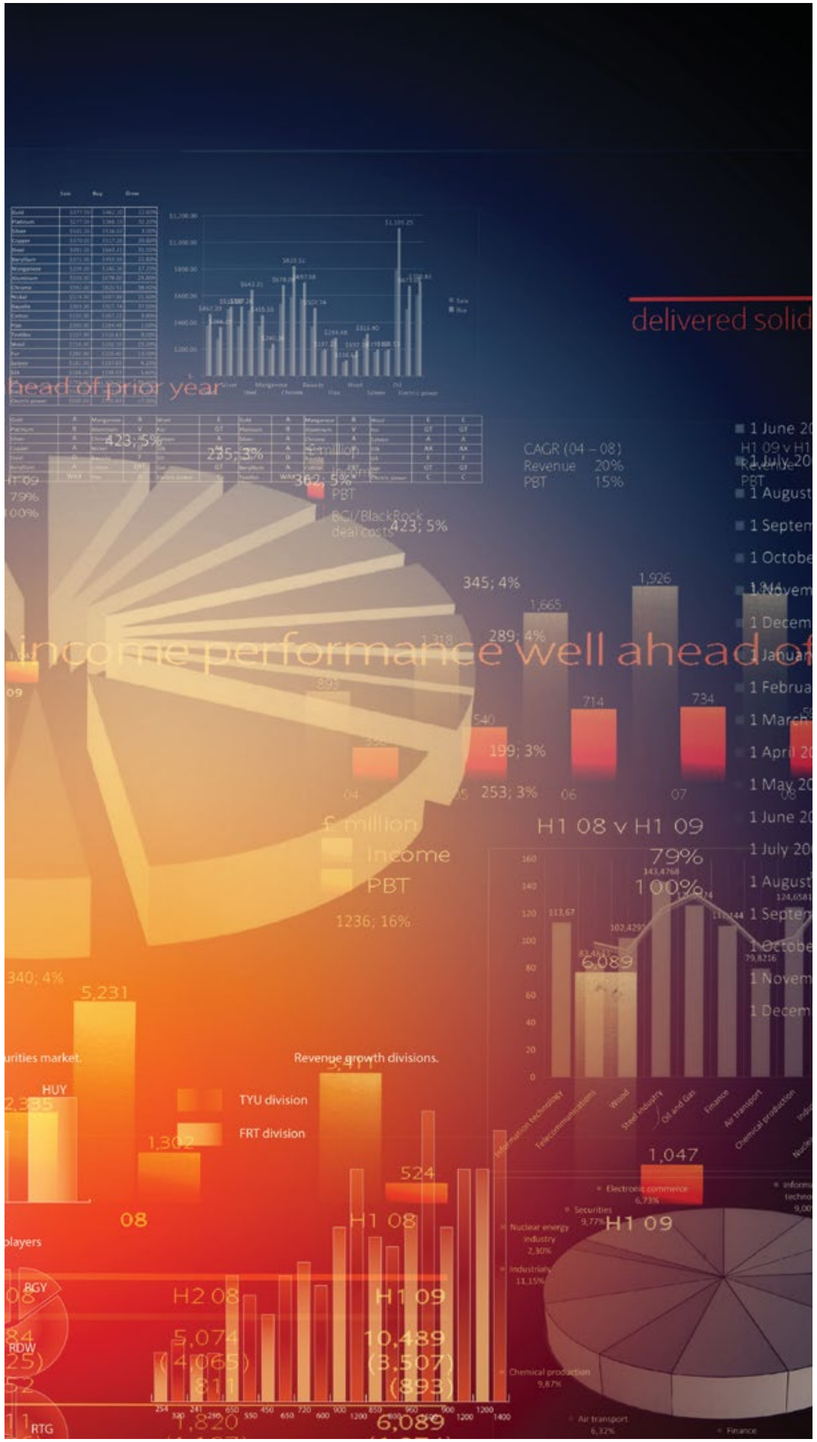
Authors:
Prof Matthew Duckham, RMIT University
Dr Lesley Arnold, Curtin University
Ms Kylie Armstrong, CRCSI
Dr David McMeekin, Curtin University
Mr Darren Mottolini, CRCSI

Australian Government
Department of Industry, Innovation and Science

Business Cooperative Research Centres Programme

crc•si

Contents



Executive Summary

The CRCSI is proposing a Next Generation Spatial Knowledge Infrastructure (SKI) that moves the agenda from more traditional Spatial Data Infrastructure concepts, to automatically creating, sharing, curating, delivering, and using knowledge (not just data or information) in support of the emerging digital economy and the rise of spatially-aware and equipped citizens.

The transition will require innovation and new practices in government and the private sector in order to capture the power of emerging technologies and to meet the future demands of users. There will be a number of converging themes; moving the focus of government away from the supply of data towards a diverse and more collaborative information management environment working in partnership with many data providers, increasingly utilising automated data sharing capabilities, developing rich open analytics capabilities, and progressively moving to the provision of a knowledge-focussed environment for customised and real-time decision-making.

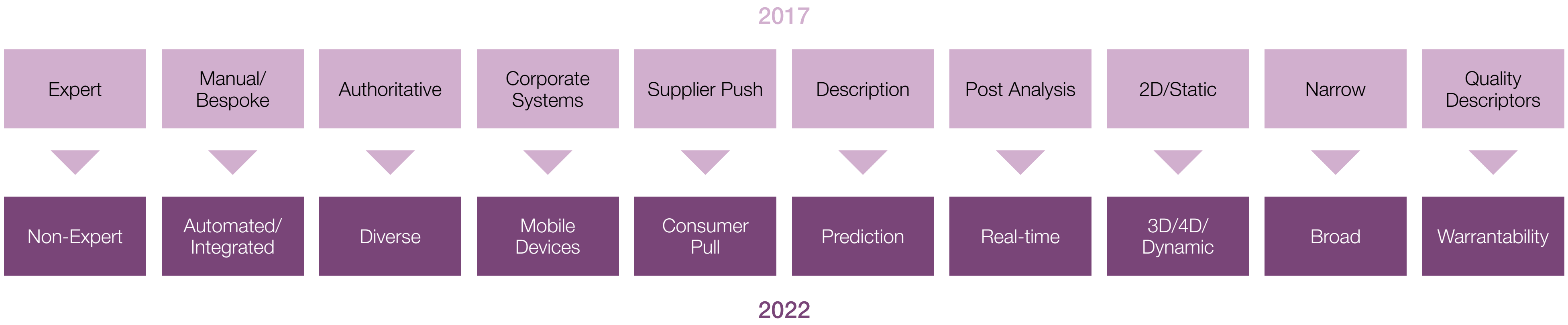
The SKI in combination with new semantic web technologies will assist in connecting, integrating and analysing data and, as a consequence, drive new knowledge-based activities such as smarter transportation networks, responsive and resilient cities, and intelligent infrastructure planning. The common thread required for these knowledge-based solutions is the delivery of data and information in real-time using machine to machine communications and on-the-fly predictive analytics.

The paper proposes a definition for Next Generation Spatial Knowledge Infrastructure:

> **A network of data, analytics, expertise and policies that assist people, whether individually or in collaboration, to integrate in real-time spatial knowledge into everyday decision-making and problem solving.**

The changes that can be expected over the next five years or so in the context of the move from a SDI dominated environment to one characterised by an SKI can be summarised in the following diagram:

The paper outlines the value proposition that justifies the move to develop and embrace the Next Generation SKI concept. It sets out the capabilities needed and charts the course that will help realise the creation of the SKI.



Introduction

Digital disruption of the spatial industry is well under way. An ever increasing reliance on mobile devices, cloud computing and open data is being driven by consumer demand for more data, accessible when required at little to no cost. Adding to this drive is the move globally towards automation, impacting the way products and services are delivered and how businesses engage with their customers.

Spatial data is fundamental to this disruption. Australia has an exemplary record in the development and delivery of Spatial Data Infrastructures within a complex federated system of government. Under this system, the responsibility for that infrastructure is distributed across a large number of federal, state and local government agencies. Currently, the majority of states and territories are delivering foundational open spatial data via a range of accessible web services, visual interfaces and other systems.

New Zealand's work to implement a national Geospatial Strategy[1] is well advanced and has resulted in the release of high-value spatial data, under open licence and in open data formats and web services; work that is supported by the New Zealand Government open data policy.

Through ANZLIC[2], the governments of Australia and New Zealand have developed the Foundation Spatial Data Framework (FSDF) to improve and coordinate access to critical and authoritative spatial information with national coverage[3]. Initiatives to improve access to open data, new addressing services, software and platforms as services and linked data initiatives are starting to be used within the spatial industry.

There remain, however, significant challenges that are impeding the impact of spatial data and related products on innovation and industry growth into the future. These include:

- Enhancing the ability of organisations and people to share and use the increasing diversity of data becoming available;
- Enabling non-domain experts to create information and apply analytics to data; and
- Reducing the duplication in supply chains to promote collaborative knowledge creation, including prediction and exploratory analytics.

The CRCSI and Spatial Industry Business Association (SIBA) have been reviewing these challenges in the context of the spatial industry[4] This review addresses specifically what is required to reduce the impediments that have impacted the spatial industry's ability to innovate and nurture new technologies.

> **It is estimated that within the next five years, new technologies and growing user demands will render current approaches to spatial data infrastructures inadequate.**

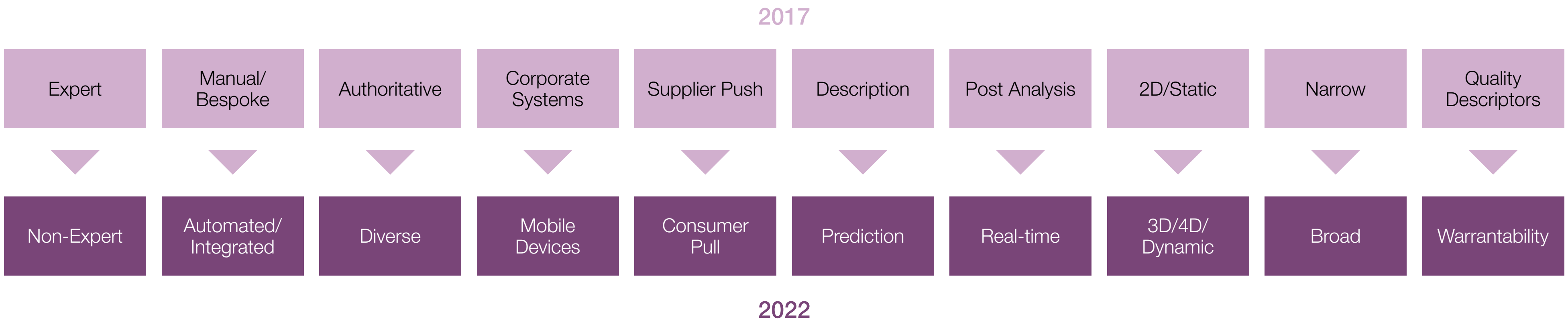
The need to evolve beyond existing unidirectional flows of predefined data products from traditional creators and custodians is critical to meet the growing demands for spatial products, analytics and knowledge by an ever increasing user base.

The challenge facing the spatial industry is to take a leading role in growing the availability and use of new and diverse data sources (including those from the Internet of Things, mobile devices, 3D and 4D data, Remotely Piloted Aerial Systems (RPAS) and humans). Collaborative analysis tools and improved automation will provide users with the capability to confidently and efficiently access the information they need, when they need it; supporting myriad applications and improved decision-making.

The CRCSI is proposing a Next Generation Spatial Knowledge Infrastructure that moves the agenda from traditional SDI concepts, to automatically create, share, curate, deliver and use knowledge (not just data or information) in support of the emerging digital economy and the rise of spatially-aware and equipped citizens.

This white paper outlines the value proposition that justifies the move to develop and embrace the Next Generation SKI concept and maps out the capabilities needed for it to be realised. Figure 1 summarises the broader technology trends and user expectations that will drive capability change over the next five years.

Figure 1: Summary of capability changes as we move towards a Spatial Knowledge Infrastructure



Case for Change

The Changing Government Landscape

Government has played a significant role in the creation and evolution of the spatial technology sector. It is a key player in shaping our future data infrastructures, including simplifying access to spatial data and driving technology standards that will enable future innovation[5].

However, the government's role is likely to change over the next decade, with all levels of government shifting to a procurement approach with a greater emphasis on external service organisations rather than in-house technology and expertise.

> **The rapid pace of technological advancement, the changing role of industry in the knowledge economy and the evolving policy priorities in the context of the Australian Government Innovation Agenda, all point to further changes to government roles in the future.**

These drivers are leading Australia and New Zealand governments to revisit the vision of an ANZ Spatial Data Infrastructure (ANZSDI), which is now two-decades old.

Embracing a New 'Knowledge-Based' Future

This white paper looks critically at what will be required of SDIs in the coming decade and charts a course of action from current SDIs to a distributed Spatial Knowledge Infrastructure (SKI).

The transition will require innovation and new practices in a number of key areas in order to meet future demands and challenges. These key areas share a common theme: moving the focus away from the supply of *data*; towards more collaborative *information* management, *automated data sharing*, and analytics, and onwards to the creation of *knowledge* for decision-making.

Information and analytics underpin the knowledge from which the majority of decisions are made. The SKI in combination with new semantic web technologies will assist in connecting, integrating and analysing data and, as a consequence, drive new knowledge-based activities, such as smarter transportation networks[5], responsive and resilient cities, and intelligent infrastructure planning. The common thread required for these knowledge-based solutions is the delivery of data and information in real-time using machine to machine communications and on-the-fly predictive analytics.

Spatial information, analytics and semantic web technologies are central to these new innovations. Globally, location based services have rapidly grown into a multi-billion dollar market. Companies like Google, Apple and Nokia compete on a global scale for the latest navigation and location-based data and innovations. They are already automating and analysing their own services to deliver better outcomes for their clients.

Technological innovations, such as cloud computing, social media, the Internet of Things, increased automation (called “Industry 4.0”[4]), and now RPAS’s, are estimated globally to generate cost reductions of $555 billion annually and business revenue increases of $650 billion per annum[6]. The Australian space and spatial industry is estimated to contribute $10 billion to Australian GDP by 2023[7].

The spatial industry has grown steadily over the past decade developing an international competitive advantage in some technology areas[7]; yet the potential return on investment achieved with real-time spatial knowledge services is still unrealised.

Value Proposition

To understand where improvements are required, we need to understand the value along the spatial data supply chain in order to move beyond current information sharing and coordination strategies and consider the value activities and partnerships necessary to maximise the benefits of spatial information more broadly (Figure 2).

Three value propositions – primary, secondary and tertiary – are discussed below from a data, information and knowledge perspective.

Data Perspective

The *primary value proposition* is aimed at internal business practices. The value activities required involve the collection and/or sourcing of data and its refinement for internal business processes. For example, cadastral survey plan lodgement, road naming and property street addressing are examples of processes contributing to the business of land administration and thus the primary value proposal. Improved value will arise from more streamlined and timelier service delivery, and where cost savings can be achieved for both producer and consumer.

Information Perspective

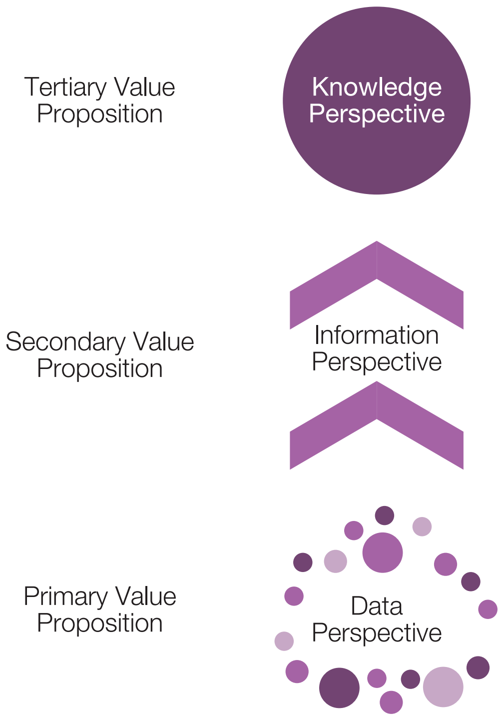
The *secondary value proposition* is aimed at improvements to external business users, often spatial data specialists who download information for planning and analysis and further value-adding. For example, the digital cadastre, essentially a by-product of the land administration process, has supplementary value as an aggregated and integrated information product. The secondary value proposition will stem from access to more information than is currently available, enhanced spatial analytics tools, improved visualisation capabilities and the ability to consistently manage datum epochs dynamically across information themes.

Improving the secondary value proposal will require spatial information to be integrated with other information themes. This will allow data to be more versatile for decision-making. In the future, linked data will become the new norm, as will domain ontologies that capture relationships and meaning between features in disparate datasets.

Knowledge Perspective

The tertiary value proposition is directed at decision makers. It is the point at which knowledge is derived from information and has subsequent value to determine the best course of action. For example, property buyers' value knowledge from integrated land and property information, as it takes the guesswork out of their purchase. Improving the tertiary value proposition will require enhanced query capabilities, real-time spatial analytics and the ability to communicate warrantability and fitness for purpose.

Figure 2 The Primary, Secondary and Tertiary Value Proposition Points in spatial infrastructures



Value Activities

To maximise the benefits of spatial knowledge, current strategies must extend current SDI activity to include four key areas to achieve a successful transition to an SKI. These activities are sharing, versatility, process and usability (Table 1).

- **Sharing:** Significant advances have been made in recent years in sharing spatial data. However, the future holds the potential to share information more intelligently, collaborate on data collection, and embed spatial analytics in innumerable tools and applications. This means making spatial data far more accessible through improved exposure, discoverability and interrogation capabilities; not only to spatial data experts, but to users across the decision-making spectrum.

- **Versatility:** Today, maps, map-layers and authoritative datasets dominate how we use spatial data. Future users will need the full range of metric 3D and 4D spatiotemporal data presented more as interactive models than as predefined map layers. Integration with a diverse range of data sources needs to be enabled, including integration with sensor, social media and crowdsourced data.

- **Process:** An SKI requires a shift in focus away from delivering data products, towards the use of products and user generated knowledge. Today, we focus on creating data products, often using workflows that are in part significantly manual. To support the evolution towards knowledge creation, our supply chains, including how we trust and warrant data and analytics and assess the fitness for use of data, must evolve to become simpler and increasingly machine to machine processable.

- **Usability:** While excellent multimedia and multidimensional visualisations exist today, all too often it is locked away in specialised tools. Future productivity requires more usable analytics. This includes easily generated scenarios and predictions and improved capabilities for non-experts; the ability to automatically locate the best spatial data and spatial analytics procedures and seamlessly integrate spatial data with non-spatial data as needed. To be truly usable, these capabilities need to be available through various online, mobile and embedded devices that are sensitive to the user's context.

These four key activity areas resonate with the Industry 4.0 agenda and its four pillars: *interoperability* (connections between machines, devices, sensors and people); *transparency* (systems that create a virtual copy of the physical world through sensed data); *assistance* (systems that support humans in making decisions and solving problems); and *automation* (cyber-physical systems become more autonomous, making simple decisions on their own)[8].

Table 1: Transforming the ASDI to an SKI: Benefits of Success and Improved Value Proposition

	Today (2017)	Benefits of success (2022)	Value proposition
Sharing	Spatial-experts dominate use and analysis of spatial data	Non-experts and domain-experts dominate spatial data use and analytics	Significant time and effort saved through improved access, sharing and collaboration on data curation; analytics; broader inclusion of domain experts in collaborative teams leads to more effective use of spatial data; reliance on spatial data increased, driving increased productivity.
	Data is shared and reused, but analysis and data fusion procedures are bespoke	Government and industry rely on automated fusion and routinely share and adapt analytics processes	
	Data analytics largely done in desktop GIS or isolated web portals	Spatial analytics easy to automatically embed in a myriad of cloud-based, distributed, and mobile tools and applications	
	Collaboration on analytics only within co-located and established groups	Broad collaborative teams with diverse expertise solve problems	
Versatility	Spatial data and analytics typically 2D “flat-land”	Seamless analytics of 2D, 3D and 4D metric data	Comprehensive spatial data available for decisions across all areas of government and industry analytics, including incorporation of 3D and 4D, dynamic, sensor-based, multi-source imagery, IoT data reflecting physical measurements and crowd-sourced data intimating human judgments and views.
	Significant duplication of data within government and wider industries that is manually collected and combined	Tools to deliver consistent and seamless datasets, with data fit for analytics purpose drawn from a variety of sources (federated)	
	Underlying reference framework is static	Underlying reference based on dynamic datum	
	Spatial data derived from relatively narrow range of authoritative data sources	Spatial data routinely from IoT, RPAS’s, sensors, crowd sourcing and social media, and mobile devices	
Process	Domination of suppliers providing users with data and describing how they can use the data	Users using the data they want, when they want and how they want it with automated understanding of use parameters and machine readable guidelines associated with usage	Increased integration of analytics and business workflows; protection from adverse effects of data misapplication; increased confidence in data and analytics; range and use of spatial data in the marketplace increased. Increased confidence in automated information and knowledge creation.
	Data quality based largely on provider reputation and known uses	Machine generated documentation of uses, production and provenance of data that can be understood by non-spatial specialists	
	Undocumented or bespoke analytics run on trusted foundational spatial data	Warrantability and trust of data, enabling scrutiny and replication of analytics from a broad range of data sources supported by fitness for purpose statements (from accuracy statements to caveat emptor)	
Usability	Data visualisation tools patchy, mutually incompatible and largely desktop-driven	Intuitive visualisation and analytics that adapt to a user’s expertise, context and devices, in open and online environment	More real-time usable, mobile, graphical and natural language interfaces; increase user base for spatial data, thus increasing efficiency; evidence-based decision-making supported by data and predictive analytics; time and costs of searching for data and using sub-optimal data and analytics reduced. Fast, efficient and cost effective spatial processes incorporated into workflows.
	Difficulty in locating most appropriate spatial data for specific applications	Intelligent search capabilities leveraging natural language eases the task of finding the most appropriate data from a diversity of options, while multidimensional ranking provides increased relevancy, supported by both text and geographic search capabilities	
	Limited and costly support for data exploration and “what if?” hypothesis testing	Ability to plan based on “what is there” and “what might happen”	
	Lack of ability to find appropriate, cost effective processes	Discovery and use of appropriate process standards with spatial workflows using plain language querying from any source	

Sharing

Sharing—of data, tools and analytics—can save significant time and effort, increase collaboration and efficient use of expertise, diversify the uses of spatial data and analytics, and ultimately drive increased productivity.

Sharing and data

Many authorities in Australia and New Zealand, and around the world have embraced open data. Today, data is more widely available than ever before. While the basic capabilities for creating and sharing data exist, data is often only accessible through government portals and cannot be interrelated and interrogated on-the-fly.

> **Data must be exposed, discoverable and able to be interrogated in the future Spatial Knowledge Infrastructure. Access to interpretable information is the key motivator as opposed to physically sharing content, which is the current norm.**

Simply sharing data physically is no longer adequate. A multiplicity of challenges exists in creating, finding, linking, integrating, processing and visualising spatial data across increasingly complex and user-focussed supply chains. Being able to integrate and process spatial data in real-time, to create new information layers and insights, will drive future innovation.

Sharing and information

While sharing of spatial data is today common, sharing of spatial algorithms and the analytics to transform data into information is not. Sharing the processes and expertise for creating information from data relies on new and emerging capabilities in at least three related areas, including:

- Simplifying access to data: Standards for data service delivery have been around for over a decade yet the uptake is still not comprehensive and the focus is on pushing predefined product data to the end user. Creating simpler data structures will make it easier for users to source the information they require. Currently, users typically only have access to predefined products. Linked data will enable information in the user's context to be drawn from multiple data sources on-the-fly. There is also a flow-on benefit to suppliers of fundamental data as the need for customised product versions, with low economies of scale, is reduced.

- Mechanisms for capturing and sharing spatial analytics: Although open-source software is widely used, the logic, data structures and algorithms behind this software are largely inaccessible except to software developers and domain specialists. Opening up this logic requires new tools and techniques for explicit and machine-readable representation of analytical processes and workflows[9]. Automatically orchestrating workflows and models to create new information layers will assist in deriving new knowledge on-the-fly.

- Cloud-based efficient platforms: Cloud-based systems are well-suited to sharing spatial analytics, and are already being used for the purpose. Research into automated cloud-based deployment and efficient parallel processing of spatial analytics and algorithms, for example, is continuing to strengthen this area[10].

Sharing and knowledge

Today's problems increasingly depend on solutions designed and grounded in the sharing and validation of human knowledge. Full automation of human spatial tasks by artificial intelligence remains a longer-term goal, but even today automation is significantly impacting many activities and domains[11].

Increasingly explicit representation and automated reasoning, using semantic web technologies will capture and leverage domain and process knowledge[12]. These techniques can be used to automatically derive the provenance of data and gauge user purpose, and process data accordingly. In combination with data captured about the user's specific context, such as previous searches, downloads, analysis types and visualisation possibilities, automation holds the potential to identify the best datasets and in the right format to meet needs.

Summary

Key capabilities for sharing spatial data and analytics for an SKI either already exist (such as data sharing infrastructures) or are in advanced stages of research, such as scientific workflows, cloud-computing and the semantic web. It seems realistic that additional research effort, targeted towards automation and integration of key critical support areas, will achieve significant impact in the next five years.

Versatility

Today, spatial information infrastructures are almost exclusively focused on a two-dimensional, static “flatland” view of the world: in short, “maps” or “layers”. Taking advantage of the breadth of new and emerging data sources, including new physical sensors, connected computing devices (such as wireless sensor networks, mobile devices and the Internet of Things), crowdsourced data and social media (humans-as-sensors), opens up the potential for extending the evidence base for decision-making into areas that do not traditionally fall under mapped information. These future areas of knowledge include real-time movement, human sentiment, three-dimensional and internal building information, and situational awareness.

Versatility and data

The diverse range of new technologies is generating a flood of new data about our geographic environment and this presents significant challenges. The research areas addressing these challenges include:

- **Three dimensional data:** Recent years have seen significant advances in the data structures and analytics required for three-dimensional data, such as building information models (BIMs). However, seamless storage, data querying, and analysis of fully metric and topologically structured 3D data, remain active research challenges[13].
- **Moving objects:** There have been significant advances in efficiently storing and analysing data about moving objects, such as GNSS trajectories of vehicles or people. Some research translation is now required to make these tools and techniques more widely available and integrated.
- **Event-based models:** Many applications require more sophisticated temporal capabilities than simply tracking moving objects. A significant body of foundational research exists to represent and query processes and events, such as urbanisation or traffic jams[14], however these are not easily combined with other tools, workflows and contexts for integrated decision-making.
- **Crowdsourced data:** A wide range of data mining and analytics tools have already been proposed for making sense of spatially referenced social media and crowdsourced data. Challenges remain in selecting the most useful and informative analytics, for example, through automation of trust models, and involving the crowd with more than simply the capture of spatial data[15].

- **Dynamic datum:** Finally, there is an urgent need to be able to manage the dynamic datum in terms of consistent management of datum epochs across information themes and the continuous translation of datum dynamics for real-time applications such as mobile mapping and automated guidance vehicles[16]. While the basic concepts and mathematics are understood, there are still a range of research questions connected with the technical and logistical issues surrounding efficient and transparent use of a dynamic datum.

Versatility and information

Creating information from diverse data sources relies on being able to interrelate data.

- **Fusion:** Conflation and fusion are amongst the most long-standing research topics in spatial data, with that work becoming all the more important for an SKI. The more diverse the data, the harder it becomes to interrelate the data. Research is already addressing semi-automated and even automated conflation and fusion approaches, with the semantic web a component of contemporary solutions[17].

- **Distributed processing:** Diversity implies not only a range of spatial datasets, but a range of devices and systems for creating information from those datasets. A range of algorithms and protocols already exist for leveraging the computational power of distributed devices, such as phones, tablets and sensor nodes[18]. Current research is developing techniques to find distributed processes to (semi-) automatically include in analytics workflows.

Versatility and knowledge

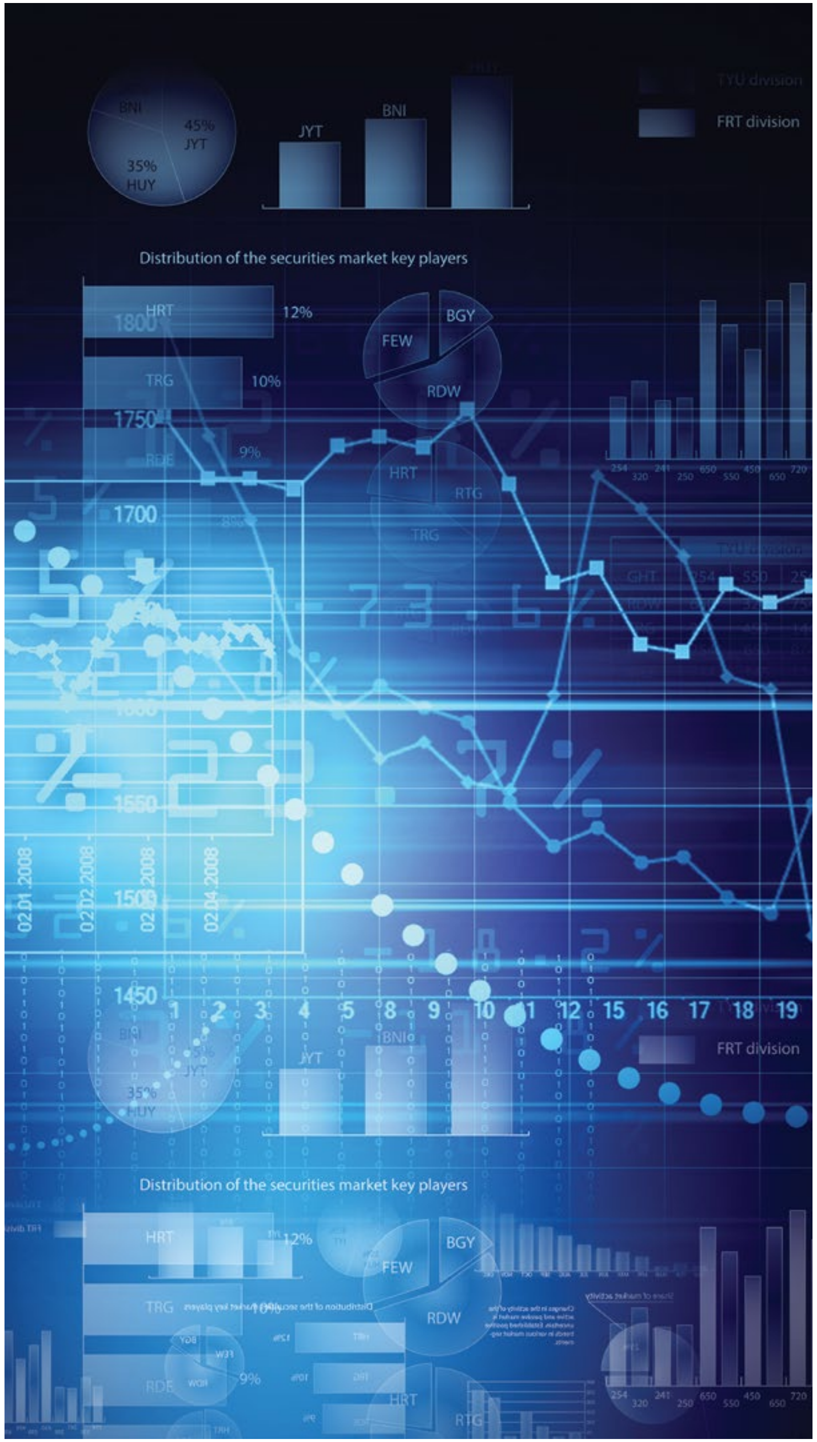
An effective SKI must support collaboration between diverse domain experts. In turn, such cooperation requires a rethink of mechanisms to enable the capture of collaborative expertise including:

- **Asking questions, not executing operations:** Today's spatial computing technologies are focused primarily on executing operations on data (e.g. buffers, overlays, transformations) rather than asking questions of data (e.g. "What locations are most suitable? Which regions are changing fastest?"). A growing body of research is exploring the development of core concepts and visual languages for framing spatial questions in order to form a commonsense common language for collaboration, free of spatial jargon and technical terms and considerations[19].

- **Non-spatial experts:** Spatial analytics is no longer the preserve of spatial experts. Significant innovation in the use and analysis of spatial data is already appearing in applications where spatial is just one component of a larger system from another domain. Research is already broadening the user-base of spatial analytics, as an integral part of everyday business intelligence[20].

Summary

The explosion in the variety of available data and the attendant increases in requirements for versatile analytics that can use this data, make this one of the most active areas for research in spatial. Advances are being made in several directions, including capturing, analysing and making sense of data from the full range of today's physical sensors and humans as sensors. Businesses will be able to automatically create warrantable knowledge to support decisions.



Processes

Authoritative spatial data can be expensive to produce; by contrast the diversity of new data sources (discussed in versatility) is generating opportunities for creating value from less authoritative data. In all cases it is the process of use, and the supply chains that support that process, that are responsible for the value created.

Processes and data

Data within an SKI will move away from being delivered as products to being published in a much rawer form and dynamically linked to metadata, rules and associated processes. Data will also be created, curated and evolved as the result of a range of dynamic processes. Two significant process changes in focus are:

- **Shift from push-to-pull data:** Current spatial data supply chains are geared towards data provision, but not discovering new knowledge, planning and decision-making. Organisations commonly lack information about why and how spatial data is used and what knowledge end users are seeking. Yet the value of spatial data lies in the knowledge that is extracted for decision-making. Current research is focused on reducing the difficulty in finding, downloading, reworking, validating and analysing data to support rapid pull of fit-for-purpose data by consumers.

- **Communicating data quality:** Understanding the quality of spatial data is essential to reasonable and effective use of that data. Experience has shown that quality is not immutable; instead the idea of fitness for use means that data suited to one use might not be suited to another. However, capturing and documenting relevant information about data quality is notoriously difficult, one of the oldest research topics in the field[21] and one that many organisations are working to resolve over the next few years.

Processes and information

The true power and value of spatial analytics is becoming more apparent and accessible as it is increasingly liberated from traditional, siloed spatial applications.

- **Embedded tools and apps:** We are already moving away from portals and desktop. Research is needed to enable this transition, making available spatial functions outside monolithic GIS, with (automated) composition of tools that allow analytic processes to be easily generated and automatically accessed.

- **Integrating spatial and non-spatial functions:** Just as we need to make spatial analytics available outside of GIS and spatial databases (above), we also need to make it easier to seamlessly integrate non-spatial analytics, data mining and machine learning.

Processes and knowledge

Moving beyond data quality, and even sharing spatial analytics, automatic access to and understanding of the processes required to process spatial data is a highly active area of research today.

- **Warrantability:** Industry and government is today hampered by an inability to provide warranties for data. Warrantability involves more than just a statement of the quality of the data (i.e. metadata). It requires a knowledge of the provenance (i.e. metadata plus lineage) of data at each step along the processing and usage chains. Making provenance available in a machine readable format enables automation of analytics workflows, a highly active area of current research[22].

- **Trust:** The increased diversity of datasets and processes will challenge data quality norms particularly in the government sector e.g. data coming from government will likely be sourced from a range of origins including government, industry and individuals. Trust takes on two forms: trusting the information coming into and then out of the supply chain. Future mechanisms to give confidence on the trustworthiness of datasets and data providers might include reputation and user review mechanisms[15].

- **Integration with business rules:** Automation of knowledge will be a major disruptor for organisations over the next 10 years. Pressure to automate domain knowledge activities such as policy and rule interpretation is expected to play an increasingly dominant role to speed up service delivery. New automated systems will give employees more time to focus on higher-level design, audit transactions and create more effective solutions and policies. Blending of scientific and business workflows is one promising example of how this may be achieved, while automating spatial transactions is another.

Summary

The framework for moving beyond simple annotation of data with quality indicators is well advanced. However, research is needed to turn this framework into practical advances for automatically capturing, tracing, reasoning and extracting the provenance of spatial data as it is processed along the entire supply chain.

In addition, the ability to automatically profile and match a user's purpose will ultimately deliver greater information, and thus knowledge that is relevant. Data provenance in combination with a user's purpose has potential to enable future automatic ranking (relevance) and rating (fitness for purpose) of information for a particular purpose.

Usability

Valuable data can lie unused if it is difficult to understand, import, find or otherwise include in decision-making processes. Making data more usable increases efficiency, ensuring the best evidence is used.

Usability and data

Making data more usable primarily involves assisting users in their search for the data they need. Recent years have witnessed significant improvements in the infrastructure to assist people in finding and accessing the right data for their needs. Research is still needed however to enable smart search capabilities, such as improved natural language search, user profiling and ranking based on past consumer behaviour. In this manner, users are assisted to discover new data relevant to their purpose, not simply access data they may have already known about. Further, the ease of capturing and storing spatial data also tends to lead to increasing duplication of data, in particular across different organisations. Supporting more efficient and intelligent search and discoverability functions is fundamental to reducing redundancy and duplication.

Usability and information

There exists a long history of innovation and expertise in mapping, user interfaces, and human-computer interaction in the geospatial sciences. As a result the map remains a key mechanism for communicating spatial information (for example, through digital globes), and there exists an ever-increasing range of different ways humans can interact with spatial data.

- **Digital globe:** The recent improvements in data search have also been matched by improvements in digital globes: single, simple interfaces to multiple dataset access, query and even analytics (for example, the National Map, the Queensland Globe and the AURIN Portal). Research is now focusing on smarter searching and a more personal experience, leveraging individual user profiles and context to provide each user with more relevant information and analytics.

- **Mobile, augmented reality, virtual reality, and non-desktop users:** Just as the analytics themselves are moving away from the desktop, so the interfaces that support decision-making with spatial data must support the full range of devices, such as smart phones, TVs, in-vehicle navigation systems, augmented and virtual reality, and many other devices and technologies. This evolution is consistent with the goal of ubiquitous computing: as our information systems become more mature, the less visible they become, instead becoming "unseen assistants" embedded in everyday objects and activities.

Usability and knowledge

As might be expected, the most active research area today in spatial data usability the creation of knowledge for decision-making. In moving towards an SKI, three areas of current research are:

- **Scenario exploration:** An SKI needs to enable people to work as easily with predictions about the future, as data about today or the past. This means being able to pose "What if?" questions, exploring the likely outcomes in the future and all different possibilities today. Integration of predictive models, data assimilation and scenario exploration are highly active areas of research with direct implications for future spatial data use[23].

- **Collaborative decision support:** An SKI should be a focus for collaboration. Infrastructure must increasingly support collaboration, both co-located and remote, as today's challenges frequently require teams with diverse expertise to work together.

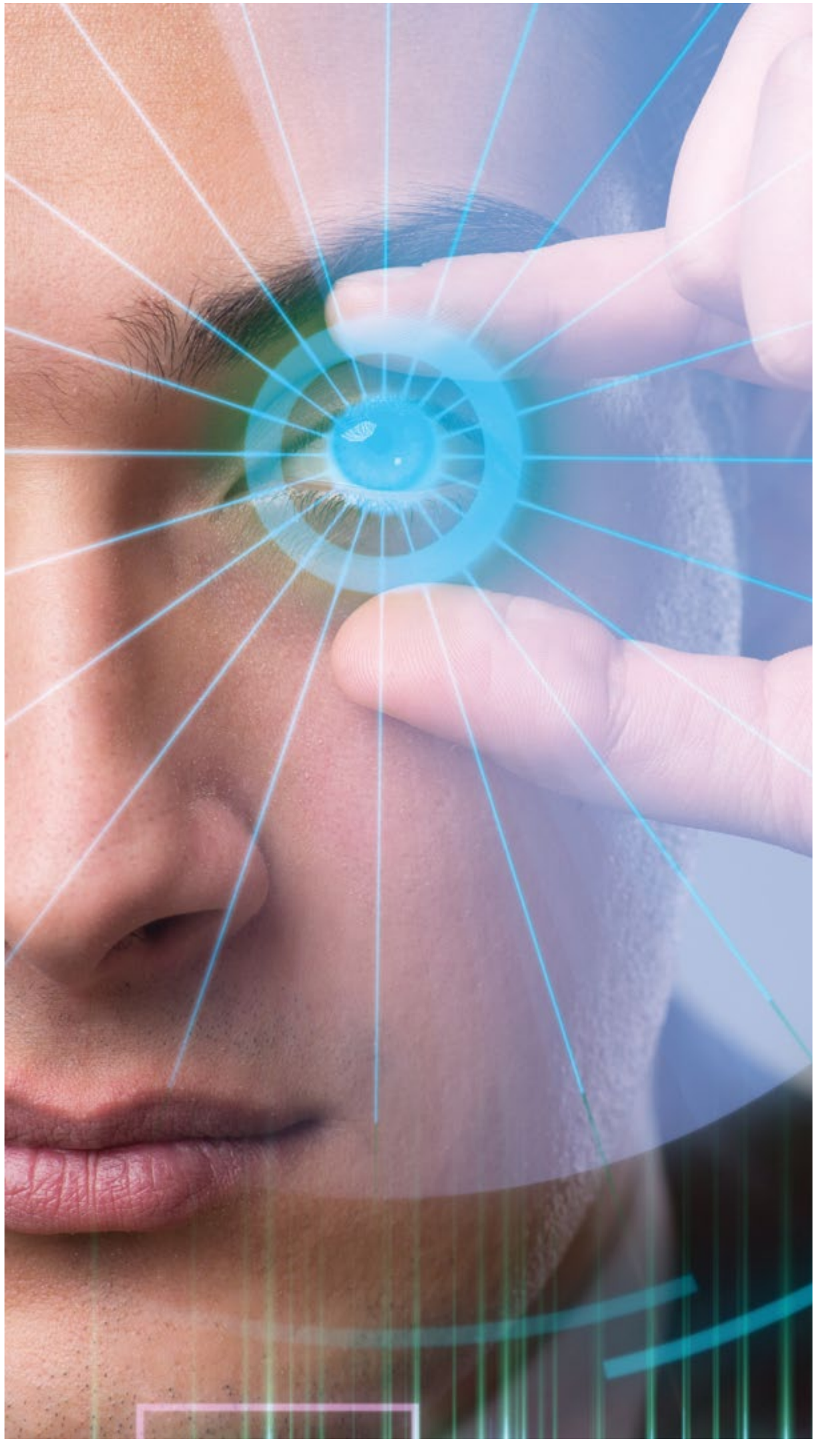
- **Rapid feedback:** Rapid feedback is fundamental to usability. Even though spatial data is getting "bigger", spatial analytics tools must be efficient enough (e.g. see Sharing and Information, cloud-based platforms) to provide rapid feedback to users. Research in this area is looking at interface design and the underlying algorithms, for example providing rapid feedback on coarse-grained approximations while more exhaustive computation continues behind-the-scene.

Summary

Spatial data query interfaces and analytics are progressively being refined as new technologies permit improved processing of big data and analytical workflows.

> **Ultimately, these processes need to be automated and integrated with all parts of a user's business if we are to achieve real-time decision-making capabilities with information about the current state of our world.**

In addition, new research is required to integrate modelling and prediction capabilities into everyday business workflows so users can achieve a better understanding of the likely future outcomes resulting from the decisions they make.



Conclusion

A shift in focus is underway in the spatial domain today: from the creation and maintenance of *data*, to the creation and maintenance of *knowledge* as the primary source of value. This shift offers the promise of a step change in efficiency and value capture. Moving the emphasis away from data products and towards knowledge creation holds the potential for increased sharing and collaboration of a wider range of data sources; for promoting engagement of all stakeholders in decision-making processes, including those without IT or spatial expertise; and reducing duplication and latency in the supply chains that underpin the knowledge economy.

This state-of-play leads us to propose a tentative definition for a Spatial Knowledge Infrastructure as:

> **A network of data, analytics, systems, expertise and policies that assist people, whether individually or in collaboration, to integrate in real-time spatial knowledge into everyday decision-making and problem solving.**

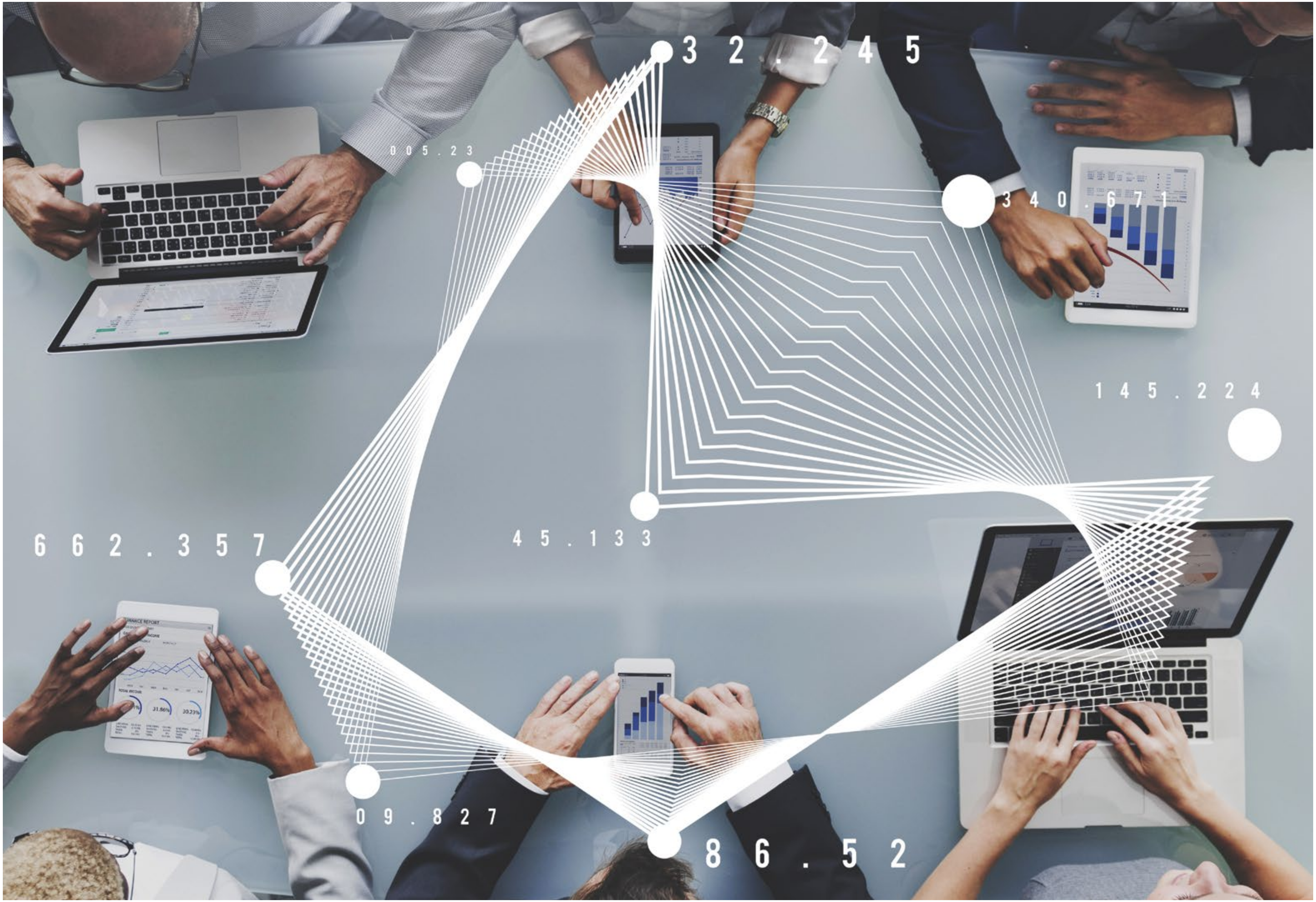
The white paper has set out a framework for understanding the types of activities needed to deliver efficiency gains and improve sharing and collaboration along the supply chains. It also considers the increasing diversity of data and the need for improved data integration capabilities and spatial analytics, and managing the whole knowledge-creation process, from data to decisions in terms of increasing value to end users and delivering knowledge in real-time.

Many of the capabilities for supporting change already exist and the white paper has provided a structure for understanding the priority areas for new research and innovation (summarised in Appendix 1), including increased automation, improving integration of a wider range of data sources, shifting to user-pull supply chains, supporting non-expert users and collaborative teams, moving beyond description to prediction, and improving capabilities to better manage the roles of suppliers and producers in broader supply chain partnerships.

Further work will need to consider the technology, policy, governance frameworks and roles of key sector players with which the future SKI will operate. Arising within such a framework is an opportunity to make sense out of the complex network of data, processes and knowledge, integrating content from applications and systems across the entire web, to create national enabling knowledge infrastructures, such as an integrated property fabric. Such a fabric has the potential to generate new insights, spin off new value added businesses and improve risk management in areas such as insurance.

In proposing the Next Generation SKI model, the CRCSI has reviewed national and international research, policy and current initiatives within SDIs and within the context of broader information and innovation initiatives. We have extended the current SDI and SKI thinking and set out a plan to extend the model to include all the necessary components to create knowledge for users.

This paper is the first in a series that will set the strategic framework and implementation components to support the delivery and use of spatial information into the future.


32.245
005.23
340.671
145.224
45.133
662.357
09.827
86.52

References

1. Land Information New Zealand (2014) New Zealand Geospatial Strategy, Available at http://www.linz.govt.nz/about-linz/our-location-strategy/geospatial-strategy-for-spatial-data-infrastructure, accessed December 2016.
2. ANZLIC (2017) The Spatial Information Council, [Online] Available at http://www.anzlic.gov.au.
3. ANZLIC (2017) *ANZ Foundation Data Framework*, [Online] Available at http://spatial.gov.au/fsdf-themes-datasets.
4. CRCSI and SIBA (2016) 2026 Spatial Industry Transformation and Growth Agenda, http://2026agenda.com/
5. Department of Prime Minister and Cabinet (2016) *Smart Cities Plan*, Commonwealth of Australia, https://cities.dpmc.gov.au/smart-cities-plan
6. PwC (2016) *Industry 4.0: Building the digital enterprise*, PricewaterhouseCoopers http://www.pwc.com/gx/en/industries/industry-4.0.html
7. PwC and Department of Industry, Innovation, Climate Change, Science, Research and Tertiary Education, (2013) Industry Sectors: Analysis and Forecasting, http://www.industry.gov.au/industry/OtherReportsandStudies/Documents/PwC-Report-on-15-Industry-Sectors.pdf
8. Hermann M, Pentek T, and Otto B. (2016) *Design Principles for Industrie 4.0 Scenarios*, Proc. 49th Hawaii International Conference on System Sciences (HICSS), doi:10.1109/HICSS.2016.488
9. Altintas I, Wang J, Crawl D, Li W. (2012) *Challenges and approaches for distributed workflow-driven analysis of large-scale biological data: vision paper*. In Proc. 2012 Joint EDBT/ICDT Workshops, 2012.
10. Wang S, Anselin L, Bhaduri B, Crosby C, Goodchild MF, Liu Y, Nyerges TL. (2013) *CyberGIS software: a synthetic review and integration roadmap*. International Journal Of Geographical Information Science 27(11): 2122-2145
11. Frey CB, Osborne MA. (2013) *The future of employment: How susceptible are jobs to computerization?* Oxford Martin School, September.
12. McMeekin DA, West G. (2012) Spatial data infrastructures and the semantic web of spatial things in Australia: Research opportunities in SDI and the semantic web. Proc. 5th International Conference on Human System Interactions, IEEE, 197-201.
13. Barbeito A, Painho M, Cabral P, O'Neill JG, (2016) *Exploring the human body space: A geographical information system based anatomical atlas*. Journal of Spatial Information Science 12: 87-111.
14. Hornsby KS, Cole S. (2007), *Modeling moving geospatial objects from an event-based perspective*. Transactions in GIS. 11: 555–573. doi:10.1111/j.1467-9671.2007.01060.x
15. Bishr M, Mantelas L. (2008) *A trust and reputation model for filtering and classifying knowledge about urban growth*. GeoJournal 72: 229. doi:10.1007/s10708-008-9182-4
16. ICSM (2017) Frequently Asked Questions on Australia Dynamic Datum, [Online] Available at http//:www.icsm.gov.au/gda2020/faqs-2.html
17. Yu F, West G, Arnold L, McMeekin DA, Moncrieff S. (2016) *Automatic geospatial data conflation using semantic web technologies*. In Proc. Australasian Computer Science Week (ACSW), doi:10.1145/2843043.2843375
18. Duckham M. (2014) *Decentralized Spatial Computing: Foundations of Geosensor Networks*. Springer, Berlin.
19. Kuhn W, Ballatore A, 2015. *Designing a language for spatial computing*. In AGILE 2015 (pp. 309-326). Springer, Berlin.
20. ESRI (2016) *Insights for ArcGIS*. http://www.esri.com/products/arcgis-capabilities/insights
21. Shi W, Fisher P, Goodchild MF. (2003) *Spatial data quality*. CRC Press.
22. Car N, Cox S, Fitch P. (2015) *Associating uncertainty with datasets using Linked Data and allowing propagation via provenance chains*. EGU General Assembly Conference Abstracts. Vol. 17.
23. Trubka R, Glackin S, Lade O, Pettit C. (2015) *A web-based 3D visualisation and assessment system for urban precinct scenario modelling*. ISPRS Journal of Photogrammetry and Remote Sensing, http://dx.doi.org/10.1016/j.isprsjprs.2015.12.003

Appendix 1: Summary of Key Research Areas

SPATIAL 'DATA' INFRASTRUCTURE TO SPATIAL 'KNOWLEDGE' INFRASTRUCTURE			
KEY RESEARCH AND INNOVATION AREAS			
	Data Perspective	**Information Perspective**	**Knowledge Perspective**
Sharing	• Open data principles • Spatial transactioning • Data warehouses	• Mechanisms for capturing and sharing spatial analytics • Cloud-based platforms	• Automation of human tasks • Encapsulating and sharing knowledge using domain ontologies
Versatility	• 3-D, 4-D moving objects and event-based models • Crowd-sourced and social media data • Dynamic datum transformations	• Automated or semi-automated data conflation and fusion • Distributed and decentralised processing	• Responding to questions via visual and natural query languages • Responding to questions via visual and natural query languages
Process	• Value activities that contribute to 'fit for purpose' data • Automated capture and use of data quality • Communizing 'fit for purpose'	• Ubiquitous access to analytical tool sets • Automatic orchestration of scientific workflows • Tighter integration of spatial and non-spatial analytics	• Scenario exploration • Knowledge-service/ interface suitable for the masses • Trustworthiness: Automatic extraction of provenance and trust modelling
Usability	• Removing supply chain duplication and redundancy • Smart search: Find information in distributed supply chains	• Innovative mapping platforms • Multi-platform access, including virtual reality, augmented reality, mobile users	• Scenario exploration, predictive models, data assimilation • Collaborative decision support • Rapid feedback

CRC for Spatial Information
Level 5, 204 Lygon Street
Carlton, Victoria, 3053
Australia

+61 3 8344 9200

JourneytoSKI@crcsi.com.au

www.crcsi.com.au

Australian Government
Department of Industry, Innovation and Science

Business
Cooperative Research Centres Programme

crc•si