Collaborative Spatial Analysis and Modelling in a Research Environment

Andries Naude[1] and Graeme McFerren[2]

Abstract

This paper deals with the results of CSIR work undertaken to establish and undertake initial application and testing of a platform for Collaborative Spatial Analysis and Modelling (referred to as CoSAMP). The platform is aimed both at improving intra-organisational interoperability, collaboration, knowledge management and capability building around geospatial analysis and modelling (within the CSIR research environment) and at improving the inter-organisational inter-operability of livelihood, geo-economic and ecosystem modelling (within the broader South African context). The activities undertaken as part of this project include the development of an open-source geoportal and geospatial content management framework (adapted for low-bandwidth environments), customisable spatial analysis workbenches (providing guidance and tools for geoprocesses such as spatial disaggregation) and the formulation of common or unified geoframes.

[1] CSIR Built Environment
[2] Meraka Institute

1. Introduction

1.1. GIS and Spatial Analysis in the CSIR research environment

Seen in terms of its mandate, inter alia, to develop and deploy technologies in response to national socio-economic development, infrastructure management, service delivery and environmental management challenges, CSIR has been involved in numerous geospatial projects and initiatives. These include projects and initiatives to standardise and improve the exchange of spatial information (e.g. the SA-ISIS project); provide information products based on earth observation data (e.g. the South African Land Cover database), undertake environmental impact predictions and assessments (e.g. the SAFARI project); as well as develop and deploy spatially enabled Management Information Systems (e.g. the IDP Nerve Centre) and GIS-based decision support systems (e.g. Gauteng's e-land system).

Despite all this, and having one of the largest geomatics and spatial analysis capabilities in South Africa, a long-standing concern has been that this capacity is not fully realised, neither in efficiency terms (reduced cost of delivering the same service, through economies of scale and work-load sharing); nor in terms of supporting all the CSIR's scientific enquiries and decision support activities that could potentially benefit from it.

1.2. Project objective and derived research problems

Against this background, the COSAMP project is focused on the following overall objective:

> *The establishment of an enhanced spatial analysis platform in support of advanced scientific enquiry and high-level decision support.*

The core research problems that underpin this research goal and form the rationale for the project can be elaborated as follows:

1. Core "external/ application-related" problem (relating to CSIR's mandate and the associated imperative to respond to the national development agenda):

   Underutilisation and misapplication of spatial analysis as a tool for integrated development planning, scientific enquiry (especially of sustainable development and service delivery issues), knowledge integration, intervention targeting and impact appraisal.

2. Core "internal/ capacity-related' problem (relating to limitations of CSIR's and the wider SA geospatial community's capacities to redress the situation above):

   A series of "collaboration divides and barriers", including hard, technical barriers (such as the ongoing bandwidth constraints and associated difficulties of transmitting geospatial data) as well as softer barriers (such as inadequate geospatial knowledge management and poorly linked geospatial resources and processes, leading to duplication and resource wastage on data assembly, pre-processing and other operations that could potentially *be shared, or significantly streamlined).*

Figure 1 provides a diagrammatic outline of seven “collaboration divides”, which have been used as basis for deriving many of the specific Co-SAMP R&D themes and required deliverables.

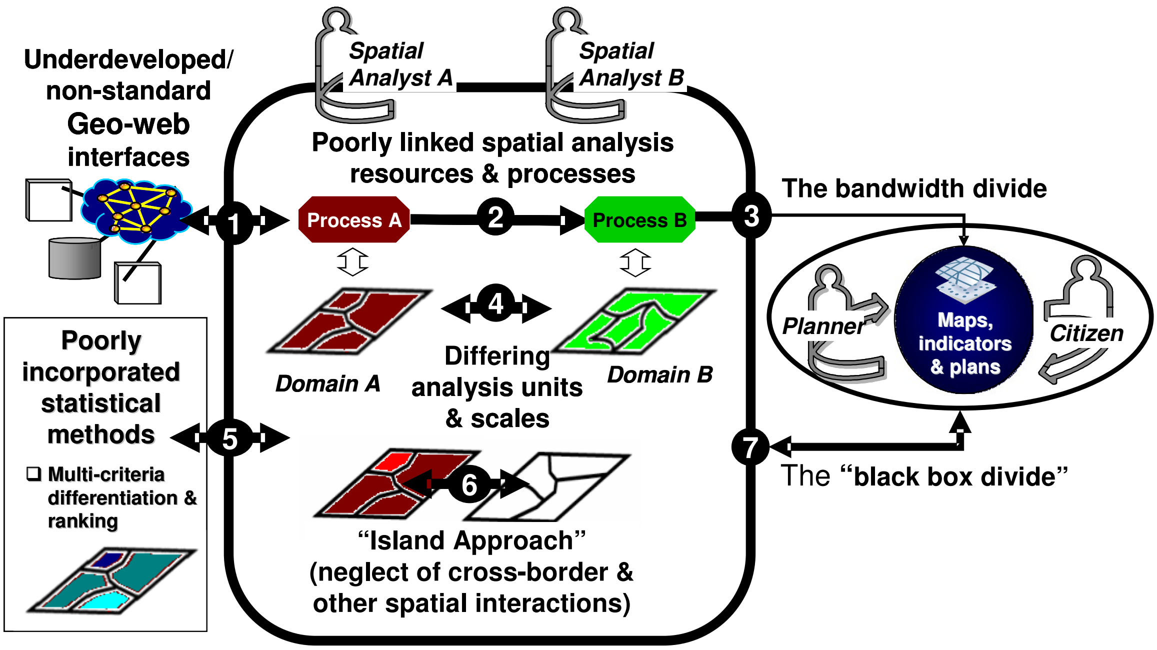

Figure 1: Collaboration Divides

1.3. Outline of paper

The following two sections describe the work that has been undertaken to develop requirements for, and develop an open source I.T. platform for addressing the first two problems outlined in Figure 1 – the problem of underdeveloped or non-standard interfaces to geo-information; and the problem of poorly linked spatial analysis resources and processes. In these two sections, the concept of a knowledge geoportal is introduced. A knowledge geoportal includes the notion of customisable workbenches, aimed at addressing the other key problems seen in Figure 1.

Further background on these problems and how CoSAMP could help to address them is given in Section 4, where two “use case scenarios” are discussed: firstly where a typical researcher or spatial analyst is confronted with complex problems of sourcing, incorporating, relating, integrating, interpreting, sharing and supplementing geo-information; and secondly, where a policy analyst or planner needs to gain deeper insights than those provided from coarse large area statistical datasets and in ‘black-box’ spatial analysis processes, where the policy analyst/ planner cannot place faith in the outcome.

2. Rationale and Requirements for a Geo-spatial

interoperability and collaboration support system

2.1. A Framework for Spatial Thinking

Over the last few years, a complementary advancement of the science and systems of Geographic Information (GI) has been taking place, driven by the increased use of spatial thinking as a framework for describing natural and social patterns and processes that take place on earth. GI science provides the theoretical underpinnings of these frameworks, while GI systems provide the tools to implement geographic thinking (Longley, Goodchild, Maguire and Rhind, 2001).

The Co-SAMP project is conceptually in this space – providing a framework of methods , data and tools to enhance spatial thinking about the South African economic, social and environmental context. The Co-SAMP project team considered it necessary to develop an Information Technology Platform to support such spatial thinking, to deal with the derived research problems outlined above and aim to achieve interoperability.

It should be noticed that the described platform is a 'knowledge' platform. The concept of knowledge is well defined by Tiwana (2002), emphasising the ability of a person or organisation to respond to new experiences or information through entraining a “fluid mix of framed experience, values, contextual information, expert insight, and grounded intuition”.

It is well understood (Nonaka,1995; Nonaka,2000; Cormican 2001) that organisational knowledge is mainly built on personal interaction, networking and collaboration. Such knowledge is often considered to be tacit to individuals or teams. Some of this knowledge can be made explicit through various processes. Outputs could be termed knowledge artefacts. Artefacts of this 'fluid mix' can often be found inside documents and repositories, organisational routines and processes, norms and practices (especially good practices). Any I.T. intervention in knowledge management must be based on this understanding. Essentially, an I.T. platform can do two things: allow knowledge artefacts to be made available for use; and as Dooley (2000) notes, provide tools that allow the “joint construction and distribution of experiences and insights and enable creation of social networks”

A broad set of requirements for a geo-spatial interoperability and collaboration support system or platform were identified, based on the understanding of knowledge elucidated above. The system is concerned with linking data and knowledge artefacts while simultaneously providing mechanisms for the involvement of and collaboration between people in the pursuit of solutions to real world problems. This is illustrated in Figure 2 below.

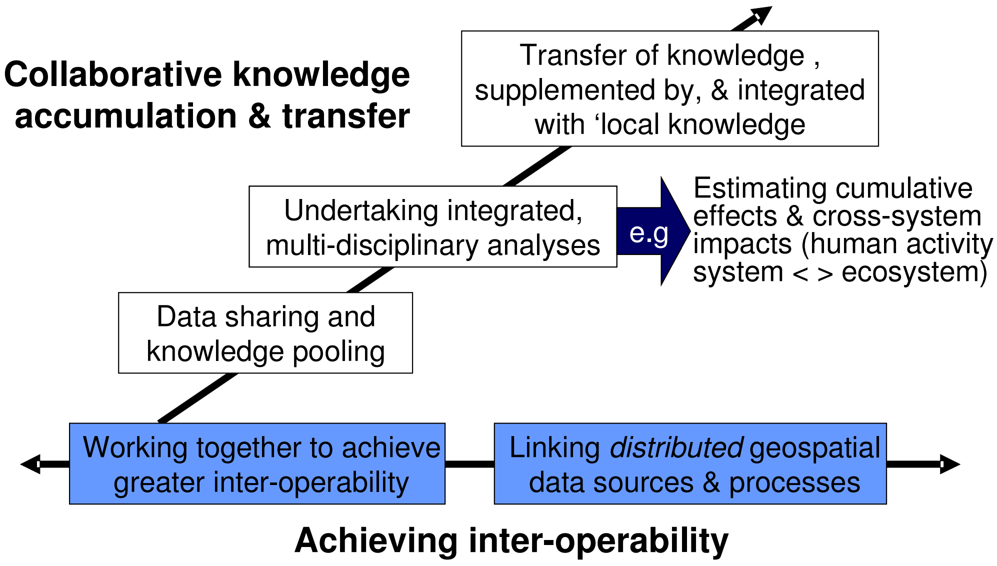

Figure 2: Geo-spatial interoperability and collaboration support system

2.2. Rationale for a Geo-spatial interoperability and collaboration support system

A rationale for the system or platform was agreed on by project participants. It was believed that investing in a system would be beneficial if it enabled CSIR staff, partner organisations and stakeholders concerned with geospatial analysis and modelling to:

2.2.1. Access spatial data

Users of the platform should have *enhanced access* to spatial data and metadata. Enhanced access means access to multiple datasets distributed across the organisation and its partners instead of access only to data available locally to a user.

2.2.2. Share good practice

Sharing of good practice allows for quicker embracing of sound methodologies that have been proven to work, whilst avoiding anti-patterns, traps, errors and wrong assumptions that have hindered people in the past. This should allow people to work faster and better, using appropriate tools with more confidence.

2.2.3. Avoid duplication of effort

Certain practices around geospatial information custodianship can reduce duplication of efforts. These would include clearer descriptions of the useful purpose of data, information and services and the reasonable use thereof. If efforts are increased to make such data information and services more accessible, duplication can also be avoided.

2.2.4. Enhance re-use of knowledge

Knowledge artefacts can be re-used and re-purposed if they are well described, can be discovered easily (often through good metadata records) and can be utilised by a broad base of applications or users. This last point highlights the importance of standard ways of describing, sharing and accessing knowledge artefacts where possible. It should be noted that codification of knowledge into artefacts is important, but such artefacts often cannot be divorced from the "knower" or person who generated them. Re-use of knowledge is more likely if that person is accessible too.

2.2.5. Create awareness of bodies of knowledge

An I.T. platform can act as a visible presence of a body of knowledge. Various information push and pull techniques can be utilised to ensure that potential users are aware of such a body of knowledge.

2.2.6. Reach more users and provide a good experience

If access to the artefacts and sometimes the creators of a body of knowledge is made available through a single I.T. platform, this platform can act as a rich source of knowledge for many users spread throughout the organisation and amongst its stakeholders. The user experience should be one of dealing with the problem at hand rather than struggling to make parts of an answer fit together.

2.3. Requirements for an I.T. platform

Based on the above rationale, a set of requirements for an I.T. platform could be elucidated. These requirements have guided the development of the Co-SAMP I.T. platform.

2.3.1. Access to spatial data

Access to spatial data needs to be provided in a way that allows for multiple users to search for, discover and consume these data, utilising potentially many different clients/ user interfaces.

The platform itself is not focussed on data storage, rather it is concerned with accessing data stored at distributed nodes on reliable and powerful database engines. The platform is not concerned with physically serving the data from such databases. Instead it aims to consume data served through standard interfaces by data service providers distributed across and outside of the organisation. Data are to be described by these providers using rich, standardised metadata sets which can also be accessed through the I.T. Platform.

2.3.2. Process Frameworks

The concept of a knowledge artefacts has been discussed above. A key form of knowledge artefact that has been identified in the Co-SAMP project is 'process knowledge'. In the Co-SAMP domain, emphasis has been placed on instructional, statistical, spatial reasoning and problem framing processes. Process artefacts need to be described and access to them provided.

Process description or a process ontology must be flexible enough to describe an atomic process adequately, but generic enough that similar processes can be grouped. Processes descriptions must not be onerous to create – information requirements must be as simple as is useful while mechanisms need to be provided to capture information.

Instances of processes such as models, algorithms, software codes, manuals and guidelines must be made accessible on the platform. In some cases, the process would simply be perused, but in other cases, accessible processes would be executed or incorporated into larger processes to deliver an output, in either a manual or automated form.

2.3.3. Content Management

It can be seen that the I.T. platform is concerned with content – data, metadata, processes and other knowledge artefacts such as documents, discussions and contact details, for example. Challenges lie in managing, integrating, presenting and utilising these sets of content. Logical (if not physical) structure needs to be placed on the content. This structure needs to be understandable and useful, allowing for easy access to content and searchability of content. Rich search functionality is seen as an important cog in the I.T. platform, itself enabled by metadata.

2.3.4. Composition and re-use

The I.T. platform should allow these data and knowledge artefacts to be utilised in different ways, perhaps unexpected. Such utilisation would often take the form of 'chaining' whereby processes and data are linked together to perform an analysis or build an application, for example. This point is not necessarily about software automation, but includes the idea that people can chain processes together to solve problems. A degree of flexibility needs to be intrinsic to the platform to make it possible to service unforeseen needs.

3. The Interoperability Challenge

Co-SAMP is an ambitious project that aims to draw together disparate parts into a greater whole. The *parts* are the various skills, capabilities and infrastructure items related to spatial problem solving, analysis and modelling that are found at various locations in the organisation. These are the knowledge flows in Tiwana's “fluid mix”. The *whole* is an enhanced ability on the part of the organisation to harness the *parts* into a platform for advanced spatial analysis and modelling, for itself and on behalf of partners, stakeholders and clients.

In an Open Geospatial Consortium white paper, Reichardt (2004) argues that non-interoperability has the potential to lead to havoc – defined as “great confusion and disorder” , realised in applications not working very well together. Two words in the title of the project hint at the need for interoperability, namely 'collaborative' and 'platform'. They both point to the idea of composition described above – the need to draw together the resources and knowledge of people, supported by the power of various I.T. elements to achieve a state of 'working together'. Now, it is possible that composition could occur without interoperability, and certain tasks completed successfully, but the overheads are high (data, software and hardware costs can be larger than necessary) and risk can alo be increased (technologies may not deliver expected benefits, leading to consequences for technology users) (ibid.). The benefits of interoperability can be noted, almost intuitively, when the definitions of the concept are explored.

3.1. Interoperability Definitions

There are two main levels of definitions of interoperability that concern Co-SAMP:

3.1.1. Technical – concerned with I.T.

“The ability of information systems to operate in conjunction with each other encompassing communication protocols, hardware, software, application, and data compatibility layers.” - www.ichnet.org/glossary.htm

“The ability of multiple databases to share digital objects across domains.” -

www.nrcan.gc.ca/cfs-scf/science/prodserv/kmglossary_e.html

“Ability of different types of databases, applications, operating systems, and platforms to function in an integrated manner.” - www.dddmag.com/scripts/glossary.asp

These definitions emphasize the concepts of sharing of infrastructure and data. The ability to function in an integrated manner is enhanced when interoperability is considered.

3.1.2. General – concerned with knowledge

The technically oriented descriptions miss out on the point that a collaborative platform – in the Co-SAMP case – involves knowledge nodes, that is, people. Knowledge systems have a decidedly human element to them , as previously discussed. These general descriptions allow for a human role in systems.

“The ability of content, a subsystem or system to seamlessly work with other systems, subsystems or content via the use of agreed specifications/standards.” - www.tasi.ac.uk/glossary/glossary_technical.html

“The ability of systems, units or forces to provide services to and accept services from other systems, units or forces and to use the services so exchanged to enable them to operate effectively together.” - www.lsc.co.uk/training/glossary.html

The themes of 'agreement' and 'ability to provide and accept' are important here. Crucial too, is the notion of using standards and specifications. Without them, it becomes harder to share more widely. They allow for disparate systems and agents (human and machine) to begin to share understandings.

The definition below is powerful, if a bit technical, and should be given some attention as it alludes to a fluid, functional world.

“The ability of different types of computers, networks, operating systems, and applications to work together effectively, without prior communication, in order to exchange information in a useful and meaningful manner. There are three aspects of interoperability: semantic, structural and syntactical” - library.csun.edu/mwoodley/dublincore/glossary.html

The technical themes are apparent again, but coming through strongly is the concept of re-purpose. If meaning(semantics), the ordering of this meaning (syntax) and the structures to house this meaning are shared, the domain of discourse is a made clearer: data and processes are not ambiguous and can be used in different ways, for different purposes, at different moments, with more confidence.

3.1.3. Beyond definitions

The above definitions are workable, but it must be pointed out that they cannot stand alone. They give no picture of the methods for and difficulties of achieving interoperability.

Willingness

Interoperability is unlikely to be achieved if there is no willingness for parties to share and work together. Political buy-in from organisations is important. Individuals need to make commitments of time and energy to make interoperability real. Organisations and individuals have to be financially willing to interoperate. There are costs in terms of resources and time that cannot be discounted. Willingness to embrace shared understandings is crucial. This could manifest in willingness to abide by agreements, to use certain standards, to aim for certain objectives or to make use of certain processes, for example.

Barriers

Interoperability is not achieved instantaneously. Barriers of cost, understanding of requirements, lack of willingness and incomplete infrastructure can all help to derail interoperability initiatives. Interoperability will be limited if a shared meaning and shared collaborative space cannot be created. The process of building interoperability needs to account for all these factors and can thus be time consuming.

3.2. Co-SAMP interoperability research challenge

The Co-SAMP project will require that interoperability is dealt with on several levels:

- Systems need to be constructed from various components and need to integrate with other systems – *System Interoperability*
- Processes need to work together, perhaps in a chained fashion – *Process Interoperability*
- Models need to slot into processes – *Model Interoperability*
- Data need to merge with other data and to fit various models and processes – *Data Interoperability*

These states cannot be achieved without shared semantics, syntax and structure, expressed, for example, in standards.

A further point requires noting. These states of interoperability need to be based on realistic requirements and are limited by what data, models, processes and systems are available and useful. Effort needs to be placed into understanding which components should interoperate and with which other components.

Finally, interoperability exists between people. Ideas, concepts and discussions need to be shared to allow knowledge to be built and passed around. If collaboration is absent, then interoperability initiatives will likely be weak at best.

4. Co-SAMP as a Knowledge Geoportal

The rationale and requirements described above closely resemble the concept of a geoportal. This section outlines the geoportal concept and how geoportals are considered to be gateways to spatial data infrastructures. The focus then turns to the limitations of geoportals in addressing the Co-SAMP project objectives and proposes the idea of a knowledge geoportal.

4.1. Spatial Data Infrastructures (SDI)

Before developing an understanding of geoportals, it is necessary to discuss some of their underpinnings, particularly the concept of spatial (or geospatial) data infrastructures (SDI).

Philosophically, an SDI aims to ensure that the “policies, organisational remits, data, technologies, standard delivery mechanisms, financial and human resources” are in place such that people working in the geospatial arena are not impeded in acquiring, processing, distributing, using, maintaining and preserving spatial data (Bernard, 2005; www.fgdc.gov, 2004). An SDI focusses on processes to provide a base for data custodians and users to from relationships that facilitate data sharing (www.fgdc.gov, 2004). The concept of SDI has a legal undertone to it – often SDI's are signed off by executive orders (Maguire and Longley, 2005) or initiated by policy setting bodies (Bernard, 2005). The point must not be lost, however, that SDI's are created and used at many scales, from local (organisation) to global (European SDI, US National SDI).

Geographic information systems have been on a steady evolution since their emergence as workstation systems focussed on projects or specific tasks. GIS is now firmly embedded in enterprise wide systems characterised by geographic distribution, large and powerful databases, complex and numerous applications , driven by industrial strength hardware and software (Maguire and Longley, 2005) . The frontier that is now being pushed and breached is that of GIS on the internet, leading to inter-organisational GIS. The concept of SDI is a top-down attempt to put order onto the GIS landscape, where bottom up development of infrastructures has led to multiple technology standards, schemas and financial models proliferating, which need to be reconciled(ibid.).

Metadata, particularly that concerning provenance, ownership, quality, age, purpose, fitness for use and restrictions on use is a crucial part of SDI's. Metadata is so important that standards have been developed and significant effort placed into building metadata clearinghouses, where catalogues of metadata can be searched according to spatial, temporal and attribution criteria(ibid.).

It can be seen that SDI's are ideally technically driven solutions that have a strong policy and governance basis. They are data-centric solutions for achieving 'connectivity' between suppliers and users of geospatial data. SDI's place no instruction on how they should be utilised. There can be multiple entry points and many ways of accessing data from an SDI. Geoportals have lately entered the SDI space to cater to the need for a single entry point to an SDI, thereby simplifying its usage.

4.2. Geoportals

According to Tait(2005), geoportals are web sites that act as entry points to web-based geographic content, where such content can be discovered. Geoportals, in the view of Maguire and Longley (2005) are “gateways that organise content and services such as directories, search tools, community information, support resources, data and applications”. In the last few years, demand has increased for the dissemination and leveraging of geographic knowledge, capabilities and content (Tait, 2005). Tait argues that geoportals are the response to this need. More generally, portals provide web environments for an organisation or community of information providers and users to: 1)aggregate and share content and information flows; and 2) build consensus (Maguire and Longley, 2005)

Maguire and Longley (2005) formulate a classification of geoportals, which is extended in this paper. This classification is easily conceptualised in diagrammatic form in Figure 3:

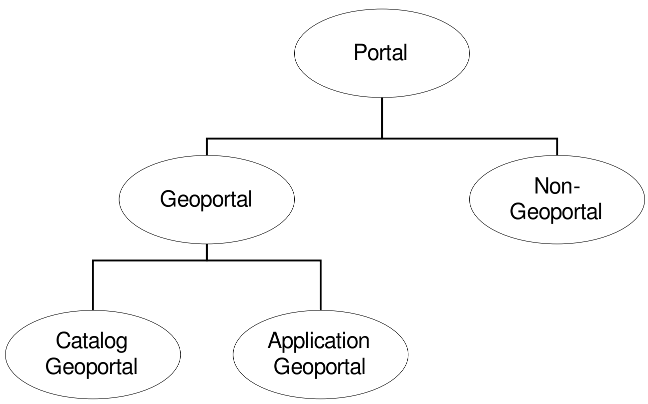

Figure 3: Geoportal classification

In this conception, most geoportals have a cataloguing function, concerned with organising geospatial data and providing access to it. Capabilities would focus on querying metadata records of data services and then linking directly to the data services themselves. Application geoportals, in addition to a cataloguing capability, provide on-line, dynamic access to web services such as routing, geocoding and mapping services, for example.

The model of how a geoportal is utilised is well described in Maguire and Longley's diagram in Figure 4. Geographic information service providers publish services and their metadata to a geoportal. Users or clients of the geoportal can then search the metadata to discover services. Once discovered, such services can be consumed (datasets added or services executed). Ideally, a geoportal should support interaction with an assortment of clients, ranging from lightweight clients such as web browsers, to full-featured GIS packages that can assimilate data and services from a number of providers to a geoportal and indeed from a number of different geoportals. It can be seen that a geoportal in this conception is an internet gateway to an SDI.

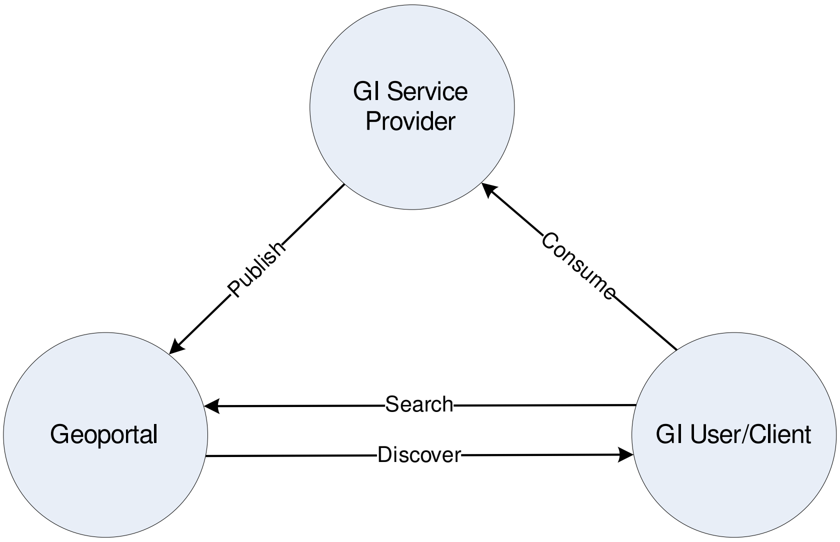

Figure 4: Geoportal Usage

Geoportals are distantly removed from the original closed world GI systems. Now, due to the Internet based nature of geoportals, users of GI can access up to date information at any time and are not bound to the desktop or workstation environment. Standard interfaces to metadata, data, information and services can allow them to be thoughtfully integrated, aided by metadata sets including fit-for-purpose and quality statements. Access to GI has been simplified with the result that more people and organisations can include geographic concepts, databases, techniques and models in their work (Maguire and Longley, 2005)

4.3. Co-SAMP as a Knowledge Geoportal

It can be seen how the notion of a geoportal as developed by Maguire and Longley, can be closely matched to the Co-SAMP I.T. platform rationale and requirements and how some of the interoperability challenges can be met. The Co-SAMP project is not primarily an SDI and geoportal development exercise, however. A geoportal is a useful aspect of the greater project, which is aimed at enhancing spatial analysis and modelling skills, performing tasks in better ways and helping to solve geospatial problems in ways that are appropriate to southern African physical and cultural conditions. The I.T. platform needs to support these use cases and indeed, this is where the notion of a knowledge geoportal originates from.

The first basis for this difference is as follows: catalogue geoportals help to discover and link to geospatial data and services, application geoportals provide various forms of access to geospatial web services in addition, but these portals do not necessarily make the link between a user (human or machine) and the problem that user is trying to solve. This situation is particularly important in situations where users of the portal are untrained in GI concepts, principles and practices. Consider that geoportals may not explain, for example:

- why a user would perform a particular task in a geospatial operation,

- what the pitfalls of a using a particular methodology or process are,
- the different possible approaches to solving a problem,
- who the people (knowledge nodes) are that can assist in formulating a solution,
- the semantics of concepts, terms and their relationships.

In addition, GI technologies do not exist in a vacuum – they are used as tools of verification, inspiration and persuasion (Berry, 1995) when dealing with colleagues, clients, stakeholders and decision makers. So-called 'black-box solutions' may not achieve these aims. The Co-SAMP project recognises that spatial reasoning and problem solving processes may require deep involvement of such actors. Clear, available, discoverable and re-usable knowledge artefacts, used in conjunction with supportive tools may help in the exploration and communication of perspectives and building of trust. This is what Berry (ibid.) refers to as “humane GIS” .

The Co-SAMP knowledge geoportal aims to address these challenges by providing a set of patterns (templates) for working with knowledge artefacts. Structured or explicit knowledge and unstructured, even tacit knowledge can be dealt with in these patterns. There is a strong reliance on metadata to build these patterns. The concept of a pattern is perhaps best illustrated by describing two types of pattern, namely a workbench and a process description framework (Naude, 1999).

4.3.1. Workbenches

The first pattern to be implemented on the Co-SAMP I.T. platform is that of a workbench. A workbench is a virtual space on the platform where groups of related knowledge artefacts can be assembled. An alternative conception of a workbench is a container for related knowledge artefacts. It would primarily combine guidelines, processes, methods, data and software tools relevant to a particular domain of discourse into a composite content type that could be accessed as a single entity from the I.T. platform. Functionally, workbenches can be divided into four main components: data management and processing components; components supporting the model selection and construction process; visualisation components; and components supporting the planning/ model application process (naude reference?).

Some examples of workbenches in the Co-SAMP project would include:

- *Geo-framing & assembly workbench* – pertaining to the creation of a small areas dataset(s) suitable for addressing social, economic and environmental development needs
- *Spatial statistical workbench* - concerned with assisting users in following a structured approach to analysing quantitative spatial data. Its aim is not to give a “shopping list” of techniques, but rather it tries to bring across a structured and holistic approach to using such data.
- *Spatial profiling workbench*
- *Location & spatial interaction modelling workbench*

- *Sketch planning workbench*

A workbench can be constructed, adapted or copied for specific purposes and is a discoverable artefact in its own right. It would be possible to compose a workbench that included another workbench.

The main research areas concerning workbenches are twofold: developing a workbench description framework (metadata attached to platform artefacts that can allow a workbench to be composed); and creating the software structures that provide a user interface to a workbench.

4.3.2. Process Description Frameworks

The Co-SAMP project recognises that geospatial research work to support the requirements of stakeholders and clients should build on the available body of knowledge, tools, good practice and guidelines related to the creation and application of geo-information. A thread running through the Co-SAMP project is that knowledge can sometimes be made explicit, and this could often be in the form of a described process, or 'hard process', which could even be codified into a software component. It is also recognised that 'soft processes', perhaps the more humane side of GI referred to above, are equally if not more important than the 'hard processes'. There are multiple efforts to describe and model hard process – inter alia engineering process modelling, business process modelling and software process modelling, but description of soft process is more challenging and perhaps less conducive to generic modelling approaches.

An important research challenge for the Co-SAMP project is to produce a process description framework that can bridge between soft and hard processes. To this end, the Co-SAMP project has explored the Open Process Framework (http://www.donald-firesmith.com/Overview/WebsiteOverview.html) meta-model. It is a "general purpose management and engineering process framework that is primarily intended for the object-oriented, component-based development of software-intensive systems" . The Open Process Framework seems reasonably useful, but is software life cycle focussed. Alternative frameworks may be the Process Specification Language (PSL), an ISO standard (http://www.mel.nist.gov/psl/rationale.html) and even the World Wide Web Consortium (w3C) OWL-S model, a reusable, OWL (Web Ontology Language) based representation of services which details the concept of process.

The PSL is designed as a modular, extensible ontology, capturing concepts required for process specification. It is seen as an enabler for creating a "process representation that is common to all manufacturing applications, generic enough to be decoupled from any given application, and robust enough to be able to represent the necessary process information for any given application". Importantly, the core of PSL has been incorporated into SUMO, the Suggested Upper Merged Ontology of the IEEE (http://suo.ieee.org/SUO/SUMO/index.html), which provides definitions for general-purpose terms and acts as a foundation for more specific domain ontologies/ knowledge bases (Niles and Pease, 2001). SUMO is aimed at knowledge re-use and encourages systems to be modelled such that they can be re-purposed or extended without radical re-engineering (Pease, et. al., 2002). The advantage that SUMO and PSL have over the Open Process Framework is that they represent axiomatic, coherent, machine processable knowledge bases with extensive user bases.

The OWL-S model, is concerned with software service modelling and specifies a process model as one of its service concepts. This process model describes how a service performs its tasks, including “information about inputs, outputs (including a specification of the conditions under which various outputs will occur), preconditions (circumstances that must hold before a service can be used), and results (changes brought about by a service).
The process model differentiates between composite, atomic, and simple processes. For a composite process, the process model shows how it breaks down into simpler component processes, and the flow of control and data between them... Atomic processes are essentially 'black boxes' of functionality, and simple processes are abstract process descriptions that can relate to other composite or atomic processes” (Elenius, et. al., 2005).

A major difficulty noted of implementing a process description framework in a knowledge environment is that of retrofitting artefacts into the framework. Documents, for example, are a common way to describe process, but it can be difficult to discover the knowledge inside a document automatically and harness it into a process. In addition, they are not typically written with a framework in mind.

The knowledge geoportal usage model is illustrated in Figure 5, and can be seen as an extension of the Maguire model.

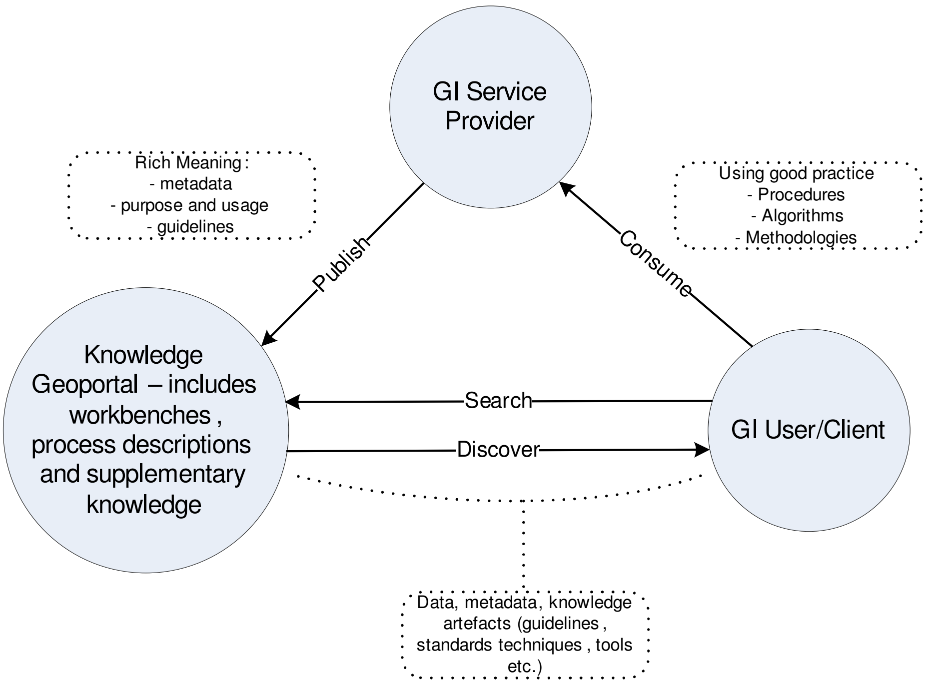

Figure 5: Knowledge Geoportal Usage

5. Co-SAMP in use: overcoming the collaboration divides

The potential uses and impacts of the Co-SAMP knowledge geoportal can be understood best by grounding it in a use case story, focusing on a hypothetical CSIR researcher/ spatial analyst, set within a context characterised by all the collaboration divides mentioned in the introduction. As part of this story, two use case scenarios will be sketched – one *with* and one *without* the use of CoSAMP and related of linked geoportals – and specific examples will be given to elucidate the implications of the various collaboration divides and related knowledge gaps.

5.1 General application context

The general application context for exploring the potential uses and impacts of CoSAMP can be briefly summarised as follows:

a) *A general policy or decision support context*, where several government departments (including the Presidency, the dti, the Department of Housing and the Mpumulanga Provincial Government) have recently requested CSIR researchers, among others, to obtain sufficiently detailed or spatially disaggregated data, and then assist them with the profiling and ranking of different local areas in terms of indicators of economic development potential, livelihood potential, service backlogs, accessibility to markets and ecosystem impacts;

b) *A high-level decision support, research and development context* (including CSIR, other South African research councils and the universities), characterised by shared basic GIS knowledge and competencies, but distinctly different and insufficiently shared bodies of application-related knowledge, causing – among others – a general divide and lack of interoperability between socio-economic and bio-physical spatial profiling and modelling activities;

c) *Simplistic "island thinking" approaches to the use and interpretation of GIS*, referring to the tendency to profile and plan interventions for sub-national territories as if they are isolated, internally homogenous "islands", thus ignoring important cross-boundary interdependencies and impacts; as well as the possible existence of significant internal heterogeneity (i.e. "structurally different" types of local development environments);

d) *Unevenly developed interoperability infrastructure and agreements*, characterised on the one hand by emerging networks of web-linked portals and information systems (such as LG-NET or Local Government Network, a Development Bank of Southern Africa funded initiative to supply bandwidth to local government, and then utilise this backbone to deliver information technology services, including geospatial information services), whilst, on the other, many organisations are still subject to bandwidth constraints, confidentiality constraints and/or competitive systems development and data accumulation agendas, causing them to operate as relatively independent "information islands";

e) *Differing analysis units* – implying that different models cannot easily talk to one another, or that it is very difficult to construct composite indicators, say of the potential for sustainable resource-based economic activity *and different or wrong scales of data* – in most cases too aggregated.

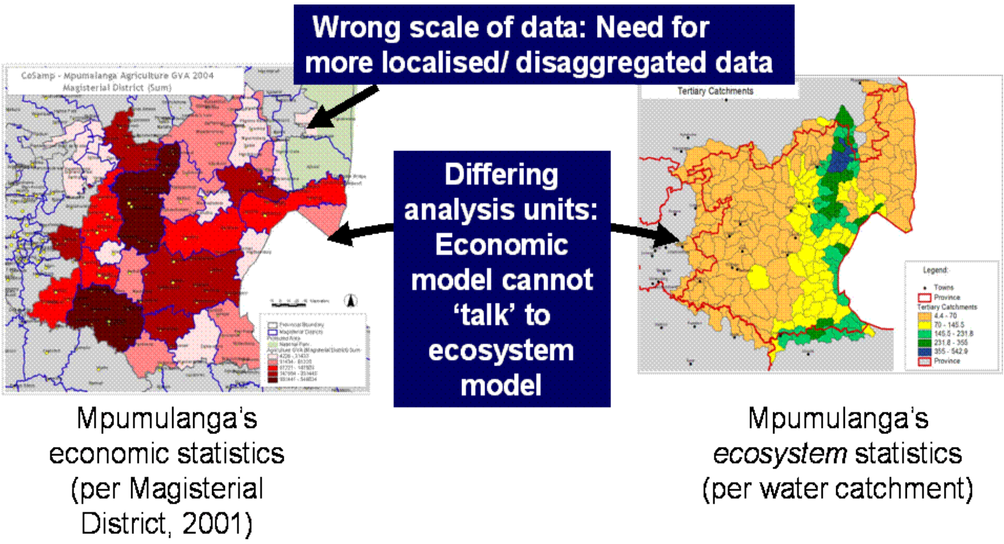

Figure 6: Illustration of wrong scales and differing analysis units

5.2 Specific use case: service access and livelihood profiling

Seen within this context, our hypothetical use case story involves a spatial analyst or researcher who has been approached – as part of a integrated planning team also involving planners and policy analysts from national, provincial and district government agencies – to source relevant geospatial data and to develop *service access and livelihood profiles.* For the purpose of the use case story, let us further assume that the provincial context is Mpumulanga, and the specific focus of the integrated planning is the coordinated targeting of housing subsidies, population-serving facilities (such as clinics) and related infrastructure investments within the Ekangala District Municipality.

To develop the service access profiles, the researcher have to consider both of the following categories:

- *Service Access Category 1*: Access to housing-related services such shelter, clean water, sanitation, and energy;
- *Service Access Category 2*: Access to facility-based services – which are services that are mainly provided from facilities such as clinics, schools, tele-centres, police stations, libraries and multi-purpose centres.

A combined service access index then needs to be constructed, and combined with other (initially undefined) indicators of the potential for sustainable livelihoods. Spatial targeting priorities will then be determined both with reference to service backlogs (or maps of poorly served areas) and the ranking of areas in terms of their relative livelihood potential.

5.3 Low collaboration and knowledge access scenario

In terms of the first scenario, the researcher obtains data electronically only via the intranet or a standard web-browser (without access to something such as the CoSAMP geoportal). As an accomplished spatial planning researcher with considerable GIS skills, she relies mainly on her accumulated experience. But, important for our story, this does not include any familiarity with spatial linkage or interaction modelling (neither the potential applications nor the execution thereof).

Starting off, due to the large size of many spatial datasets and the bandwidth limitations, she expends considerable time and effort trying to locate and download datasets, some of which were not fit for the purposes they seemed to fulfil.

After the information has been sourced, the researcher also has to spend considerable time: a) performing data cleaning operations; b) disaggregating and converting different types of data into a compatible dataset for the same set of analysis units[3].

This leads the researcher into several traps: in an attempt to disaggregate some ancilliary statistical information, the researcher does not notice that she is hiding some important clusters; furthermore, inaccuracy is introduced by the researcher neglecting to apply a suitable projection to the different datasets used in the disaggregation analysis process.

Using census statistics and information on recent housing infrastructure projects the researcher is eventually able to develop quite good indicators of access to housing related services (i.e. Service Access Category 1, as defined above). But these sources prove unreliable and outdated for the purpose of developing indicators of access to schools and other facility-based services, partly because of considerable changes in the number of facilities that have been built since the last census. To address this shortcoming, the researcher then sources and succeeds to obtain recent information about the location and sizes of all the important population serving facilities. Realising that aggregate availability indicators such as shown in Figure 7 (see Diagram A) would be too coarse for the purpose of identifying poorly served areas (and targeting priorities), the researcher re-estimates these for smaller analysis zones, nested within the boundaries of the larger areas as indicated in Diagram A. But even

[3] Di and McDonald (1999) note that in many cases, the temporal and spatial coverages and resolution, origination, format, and map projections of datasets are incompatible. These authors point out that data users spend considerable time on assembling the data and information into a ready-to analyze state, estimating that more than 50% of users' time is spent on the geoquery and geo-assembly steps of the geospatial knowledge discovery process.

with this refinement (not shown in the diagram), the researcher is unable to develop a robust indication of the poorly served areas.

The reason for this is illustrated by Diagram B in Figure 7. Although the calculated facility availability indices highlight Area 2 as the neediest area, the extent of cross-boundary movements[4] to nearby facilities in Area 1 means that the population in Area 2 could actually be quite well served. And whilst preparing disaggregated availability indices might highlight certain poorly served pockets (such as shown to be in Area 1), the results might also be subject to arbitrary boundary effects.

Dia
Ca

Figure 7:

(Morojele and Naude, 2003)

Having spent time unproductively at geo-information assembly and other tasks at the beginning stages of the geo-information value chain, the researcher now has very little time to apply her mind to the design of a logical set of indicators for the last part of her assignment: constructing livelihood indicators. She simply selects the well-known "five capitals livelihood model" (in terms of which sustainable livelihoods are deemed to depend on social, human, natural, financial and physical "capitals"), assembles readily available information that somehow relates to this, and "crunches" this together to provide a composite map of spatial variations in livelihoods. In the process, she does not have any time to:

[4] This is a specific example of what Maguire defines as *boundary problems*. Formally stated, these occur because geographical study areas are usually bounded in ways that do not correspond with the effects of spatial processes (Maguire,1995).

- Consider or model the local livelihood effects of jobs that are or could be accessed through long-distance commuting (e.g. in Northern Gauteng);
- Interact with local stakeholders, and supplement the statistically derived indicators with local knowledge about recent projects or unrecorded or unanticipated factors (e.g. the livelihood effects of high levels of air pollution in the Witbank area).

The researcher delivers his work to the team, but is disappointed when the team questions the results, and ultimately rejects the work, noting that they could not trust the reliability of the outputs, and also points out that some local knowledge is missing from the analysis. The researcher is asked to conduct the work again.

5.4 High collaboration and knowledge access scenario

In the case of the second scenario, the researcher spends some time discovering resources in the organisation devoted to geo-information, and is made aware of the existence of a collaborative spatial analysis and modelling platform (CoSAMP), with a geoportal as the entrance point.

She discovers that CoSAMP provides easily accessible and user-modifiable guidelines (workbenches) to explore and address the complex range of problems associated with data/ **sourcing**/ assembly, **incorporation** of data or indicators from different domains, **disaggregation** (but then incurring boundary problems)**, relating** local scale features or zone properties with relevant wide area or cross-boundary entities, **integrating and interpreting** the information (e.g. by constructing composite indicators, or providing inputs in an integrated assessment process); **sharing,** and **supplementing** (as part of an interactive or participative process).

Following on the earlier description and diagrammatic illustration of the concept of a knowledge geoportal, Figure 8 provides an overview of how the researcher, together with other members of the hypothetical planning team (indicated in the box: *policy analysts/ local planners*) could utilise: a) the CoSAMP Geoportal, as well as b) a (possible) derivative geoportal (indicated as *LGNet-linked Geoportal on Geospatial Knowledge Services*) for the required service access and livelihood profiling process.

By implication, this scenario assumes that the CoSAMP geoportal will be deployed as part of a network group of geoportals that exist under the umbrella of a broad information delivery service. The service illustrated is LG-NET or Local Government Network, a Development Bank of Southern Africa funded initiative to supply bandwidth to local government, and then utilise this backbone to deliver information technology services, including geospatial information services. Co-SAMP is shown as a node in this network, making certain of its data, metadata and other items available to the broader network. Other geoportals would make similar provisions, to create a situation where distributed information is accessible to a user of LG-NET.

Figure 8: Hypothetical Use Case of Co-SAMP portal

The diagram also indicates that it could be possible to design the profiling process in such a way that it forms part of a general value chain of which some of work on the pre-process or supplementary process stages could be shared and/or farmed out, saving costs and leaving the researcher to devote most of her attention on the core project activities and deliverables. Hence, for the purpose of this scenario, one can assume that:

- Pre-processing work on data assembly and disaggregation is farmed out to a geo-assembly specialist/ GI service; and
- A sketch planning service provider is involved in supplementary processes aimed at capturing local knowledge, and/or supporting the planning team to plan and assess different spatial targeting options.

Assisted by the guidelines, tools, geoframes[5] and other knowledge artefacts that are contained in a *geo-assembly workbench*, the geo-assembly specialist accesses the geoportal, searches for and discovers relevant data, metadata and knowledge artefacts. Figure 9 provides further details about on the specific geoquery/ geoassembly processes that this specialist could undertake (see processes 1a and 1b). Under the ideal scenario, the information that will be accessed will be "formatted" in terms of the standard geospatial data service specifications of the Open Geospatial Consortium (OGC), which include Web Map Services (WMS) and Web Feature Services (WFS).

[5] Geoframes, in this context, refer to geo-referencing frameworks for "area data". By implication, geoframes can then be seen as a specification of the boundaries of the areas to which the area data is referenced.

Assuming the existence of a "Geo-link" spatial profiling workbench, our researcher then concentrates on utilising the guidelines, tools and other knowledge artefacts that this workbench provides for addressing most of the disaggregation and boundary problems that were highlighted in the description of the first scenario. Processes 2a and 2b in Figure 9 are examples of what this could comprise.

Gene
categ

Figure 9: Process chains for geo-assembly and access profiling
(based partly on Li, 2004)

Having undertaken the necessary geocomputation processes (see process 2b), our researcher is now in a position to produce much more robust and detailed service access profiles and maps, clearly indicating the specific areas or zones where access to facility-based services are below specified access time or distance thresholds, or where the accessible facilities do not have sufficient capacity (Morojele and Naudé, 2002) – see Figure 10. Calculating the numbers of people in these areas, she could then also produce several derived service access indicators, such as the total under-serviced households per municipal area, or for any other relevant larger area. Finally, it would also be possible to calculate distance or travel time-based service access statistics for each zone in the "hexframe", and combine it with other service access indicators that might have been assembled for these zones.

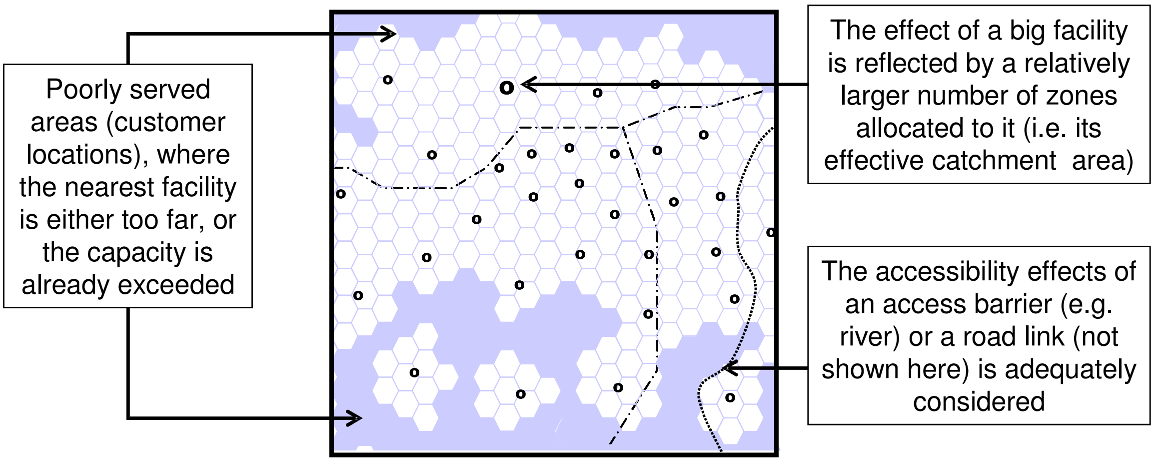

Figure 10.

Moving on to the estimation of livelihood profiles, the researcher finds that the same basic workbench that she used to develop service access profiles, can now also be used to estimate accessibility to job opportunities. Using a province-wide demarcation of meso-level zones (a so-called "mesoframe") and associated data on employment per zone, as well as inter-zonal travel distances, she develops a weighted index of the total employment opportunities within successively wider distance bands. Overlaying this with the service access information that was earlier calculated, she is then able to construct and map a composite index of livelihoods (see Figure 11).

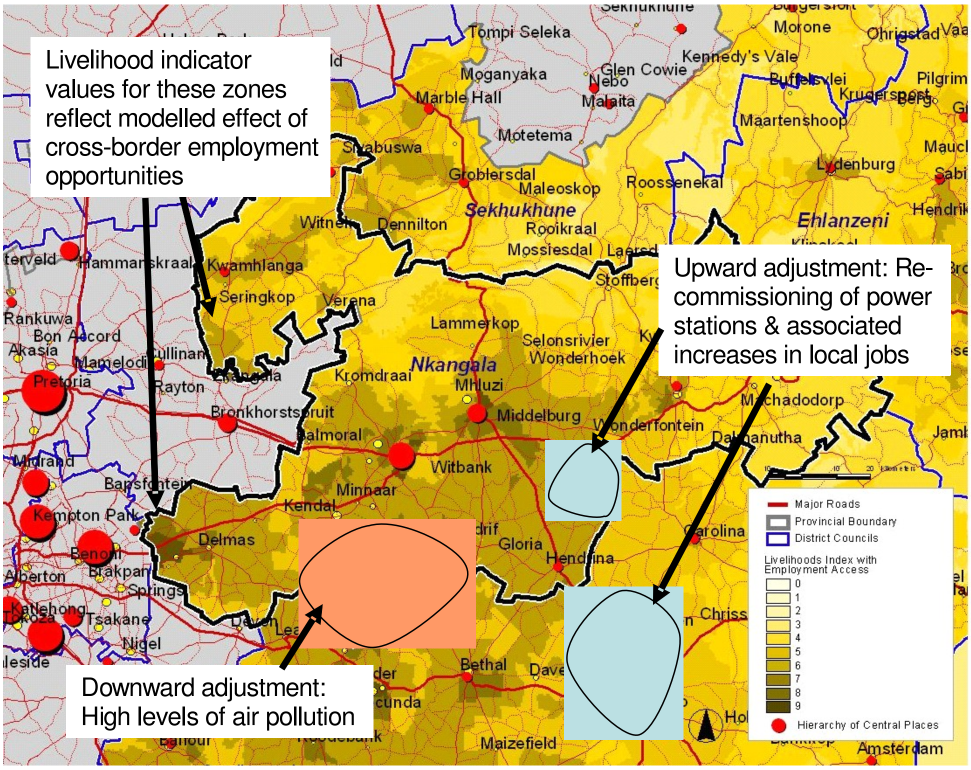

Figure 11: Output example: simulated sketch planning exercise based on disaggregated livelihood indicators for Nkangala District Municipality

Emboldened by the results that were obtained, the researcher presents the results to the planning team, and suggests that they supplement it by hosting an interactive sketch planning[6] session, facilitated by a sketch planning service provider. She motivates this by referring to the need to obtain the inputs of local stakeholders and other persons with good local knowledge on factors that – in their opinion – are, or could be having a significant future influence on sustainable livelihood opportunities.

The leader of the planning team then accesses the Geoportal on Geospatial Knowledge Services (see Figure 8) and discovers a local planning consultant who has been trained as a sketch planner (and happens to keep up to-date on the evolving guidelines and best practices in this field through a link to the Co-SAMP sketch planning workbench).The team commissions this service provider and, using an interactive SMARTBoard to display the livelihood results and zoomed-in images of selected areas (e.g. Google Earth), he assists them with making "upward" and "downward" adjustments of the livelihood indicators (see annotations on Figure 11).

5. Conclusion

It is envisaged that the Co-SAMP platform become a useful place where geospatial data, information and knowledge is generated, extended, shared, described, re-used and re-purposed. Agility, good practice and quality of work is the desired outcome of the I.T. Platform, leading it is hoped to better research and servicing of the needs of researchers, partners, stakeholders and clients.

[6] According to Hopkins (1998), sketch planning does not only require a capability for initial broad-brush planning, but also the means for relatively effortless elaboration and specification of options so that semi-automated calculation processes can be applied to estimate their impacts.

References

Bernard, L., Kanellopoulos, I.,Annoni, A., Smits, P. 2005. “The European geoportal – one step towards the establishment of a European Spatial Data Infrastructure”. Computers, Environment and Urban Systems, 29 15-31

Berry, J.K., 1995. Spatial Reasoning for Effective GIS. Fort Collins, GIS World Books

Cormican, K.T. 2001. “Product Innovation Management for Networked Organisations.” Computer Integrated Manufacturing Research Unit, Engineering. Galway, National University of Ireland

Di, L., 2004. “Distributed Geospatial Information Services-Architectures, Standards, and Research Issues”. The International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences. (7 pages. Invited).

Dooley, L., Cormican, K., O'Sullivan, D., Wreath, S. 2000. “Supporting Systems Innovation”. International Journal of Innovation Management, 4(3) 277-297

Elenius, D., Denker, G., Martin, D., Gilham, F., Khouri, J., Sadaati, S., Senanayake, R. 2005. “The OWL-S Editor – A Development Tool for Semantic Web Services”, Lecture Notes in Computer Science, Volume 3532, Jan 2005, Pages 78 - 92

FGDC. 2004. “Geospatial one-stop: encouraging partnerships to enhance access to geospatial information”.
Available www.fgdc.gov/publications/documents/geninfo/gos.pdf

Hopkins L.D. 1998. “From sketch planning through commitment: progress toward a Planning Support System for urban development”. Computers in Urban Planning and Urban Management: Volume 2. Narosa, Bombay, India, 679-691

Longley, P.A., Goodchild, M.F., Maguire, D.J. and Rhind, D.W. 2001. Geographic Information Systems and Science. Chichester, Wiley

Maguire, D.J. 1995. “Implementing spatial analysis and GIS applications for business and service planning”. In Longley P and Clark G.(eds.) *GIS for Business and Service Planning*. Geo-information International. Cambridge

Maguire, D.J., Longley, P.A. 2005. “The emergence of geoportals and their role in spatial data infrastructures”. Computers, Environment and Urban Systems, 29 3-14

Morojele, NI, Naudé, AH and Green, CA. 2003. “Planning for Equitable and Sustainable Access to Facility-based Services: The Use of GIS-based Auditing and Planning Support Systems”. Paper presented at the 9th International Winelands Conference: 10-12 September 2003: Innovative Service Delivery in the Context of Change. Stellenbosch, September 2003.

Naudé, A.H. 1999. The Essential Elements of a Planning Support System; *Proceedings of the 6th international conference: Computers in Urban Planning and Urban Management - CUPUM99*, FrancoAngeli; Venice, 8-10 September, 1999

Niles, I., and Pease, A. 2001. “Towards a Standard Upper Ontology”. In Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001), Chris Welty and Barry Smith, eds, Ogunquit, Maine, October 17-19, 2001

Nonaka, I., Takeuchi, H. 1995. The Knowledge Creating Company. New York, Oxford University Press.

Nonaka, I., Ichijo, K., Von Krogh, G. 2000. Enabling Knowledge Creation. New York, Oxford University Press.

Pease, A., Niles, I., Li, J. 2002. “The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications”. Proceedings of the AAAI-2002 Workshop on Ontologies and the Semantic Web, July 28 - August 1, Edmonton, Alberta, Canada. Located at http://projects.teknowledge.com/AAAI-2002/Pease.doc

Reichardt, M. 2004. “The Havoc of Non-Interoperability”. Open Geospatial Consortium Whitepaper, 10 December 2004 . www.opengis.org

Tait, M.G. 2005. “Implementing geoportals: applications of distributed GIS”. Computers, Environment and Urban Systems, 29 33-47

Tiwana, A. 2002. The Knowledge Management Toolkit: Orchestrating I.T., Strategy, and Knowledge Platforms. Second Edition. Upper Saddle River, New Jersey, Prentice Hall PTR