

# Policy and design issues affecting the development of an Information Architecture for a Government Resource Discovery Service

Lloyd Sokvitne  
Manager (Information Systems Design)  
State Library of Tasmania  
Email: [lloyd.sokvitne@central.tased.edu.au](mailto:lloyd.sokvitne@central.tased.edu.au)

**Abstract:**

*This paper discusses the nature and characteristics of a Government Resource Discovery Service, and uses, where appropriate, experience and research from Tasmania Online and ServiceTasmania Online. A user-centric model is proposed to form the basis for a Government Resource Discovery service that is based on the relationship between user needs, government content, and the behaviours that users exhibit online to satisfy those needs. Within this model, three key user outcomes are identified: known item, known resource, and known topic. The methodology to provide these outcomes is defined as the Information Architecture of the Government Resource Discovery Service. Components of this Information Architecture include search engines, subject directories, special indices, and alphabetical listings. Each of these components is reviewed in terms of the specific adaptations needed to satisfy user outcomes. All such components need to provide result sets with high levels of relevance and precision, meaning that content selection and metadata quality are critical components. A surrogate-based model for content selection and management is developed and experience at ServiceTasmania Online described. The nature of a business case to support such a service is discussed, largely based on the quantification of those costs as a proportion of overall Government online expenditure and as necessary to realise that expenditure.*

## Introduction

National and state governments in Australia are increasingly involved in the provision of online information and services to their constituents. This process is often enunciated in formal policy objectives and is supported by specific initiatives and standards to promote the creation of online resources and to maintain a quality online environment (Office for Government Online, 2000b). To complement this, most jurisdictions provide a single location on the World Wide Web where their constituents can gain access to the full range of online services.

The provision of a single entry point on the Web for access to government services, together with policies that coordinate government online services and content, have the potential to allow governments to develop powerful and effective Web sites. Most importantly, these Web sites can realise the government expenditure in online services by ensuring that government content is easy to find and use. This paper will argue that an effective Government Resource Discovery Service (GRDS) has unique characteristics and requires a specialised approach. This includes a specific Information Architecture and adequate resource provision.

The rapid development of the Web over the past five years has meant that governments lack highly developed and tested conceptual models or empirical research to guide GRDS development. Most of the research and empirical components of this paper are derived from the development and operation of *Tasmania Online* (<http://www.tas.gov.au>) and *ServiceTasmania Online* (<http://www.service.tas.gov.au>). The conceptual models proposed and used in this paper are also based on that experience.

*Tasmania Online* provides a single access point to Tasmanian content on the World Wide Web that includes all Tasmanian community, commercial and government resources. *ServiceTasmania Online* provides focused access to government services through a single Web site. The *ServiceTasmania* project in Tasmania has adopted a three-part approach to simplify access to government services for the Tasmanian community. Government services have been consolidated and delivered physically via *ServiceTasmania* shopfronts; over the telephone using a common interactive voice response (IVR) phone service and the provision of a single number to access government; and over the Internet using *ServiceTasmania Online*.

## Background

The development of a single and simplified approach to government information and services on the World Wide Web understandably reflects the issues and problems experienced on the World Wide Web generally. Although this is an environment where content continues to grow rapidly and where there is little or no control over content development and quality, major strategies have developed to enable the user to locate or discover specific World Wide Web content.

There are three basic types of discovery services on the Web:

- Open Web portals - provide a variety of value-added services to identify and locate information of all types across the World Wide Web, e.g. Yahoo, Altavista
- eCommerce portals - sites offering commercial, sales, or fee-for-access services, - these portals can be horizontal - offering access to a variety of commercial enterprises/services e.g. a shopping mall, or vertical - offering access to a single company or service e.g. Amazon.com, eBay
- Subject portals - sites related to a specific topic area - e.g. SeniorsNet, Information Gateways are portals with high levels of content assessment, provider collaboration, and quality control (Campbell, 2000), e.g. MetaChem, EdNA

Open Web portals attempt to cover as much of the World Wide Web as possible and the breadth of that coverage is seen as a major competitive advantage for most Open Web portals. As a corollary and as a characteristic of the World Wide Web itself, Open Web portals assume that the content is openly

available, but uncontrolled. The content of an Open Web portal is based on open access, inclusive coverage, and uncontrolled content production.

eCommerce portals reflect a tightly controlled resource environment, where resource sharing is limited and determined by the market value or competitive advantage of that information. The content on an eCommerce site is therefore tightly controlled in terms of how it is created and presented and what is included. The content of an eCommerce site is closed, exclusive, highly controlled.

Subject portals assume that Web resources are to be made openly available, but limit the resources provided through their scope definition and an appropriate selection process. Subject portals do not have control over the development and quality of resources on the Web, but can include or exclude content on the basis of its appropriateness and quality.

Portal type	Resource availability	Range of content	Resource content control
Open Web	Open	Open	No
eCommerce	Closed	Restricted	Yes
Subject portals	Open	Restricted	No

An Information Gateway is a specific type of subject portal where the resources are open but selected for relevance and quality, where portal and discovery standards have been developed, and where high levels of collaboration exist between content providers and the discovery service. It is this type of subject portal that best reflects the need of both government and its constituency, and that should provide the model for a GRDS.

Governments have, however, tended to provide entry points that use the software and management process associated with Open Web portals. This is not surprising given the dominance and importance of Open Web portals to the economics of software development and the overall culture of the Web. It is also easy to assume that government content is uncontrolled and that content restrictions within government are impractical.

### A user-based model for government resource discovery

In order to determine the basic characteristics of a GRDS, the following model is proposed. It has been adapted from a number of different components used to define information-seeking behaviour in a research environment (Erdelez 1999), but simplifies those components to reflect an environment where the user has already decided on the scope of their search.

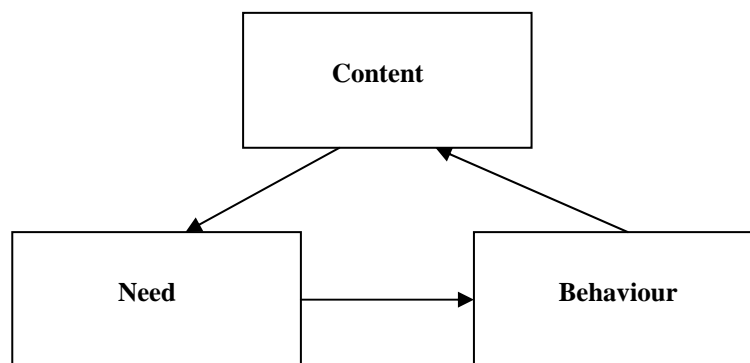


Figure 1 The three components of user interaction with a Government Resources Discovery Service

In this user-centric model the user is attempting to satisfy a need through access to government content. Government content is seen by the user as a simple homogeneous whole and the user will exhibit a variety of discovery strategies (behaviours) depending on the nature of that need and the type of knowledge/skill that the user possesses.

The client's behaviour is linked to their internalised understanding of their need and to their perceptions as to the type of content that will meet that need. The defining requirement for a GRDS is that it conforms to the behaviour shown by the user in their endeavors to satisfy their need. It is not appropriate or possible for the GRDS to modify the user's internal understanding of their own need, nor is it appropriate to modify the user's need so that it conforms to content structure.

From this model, the objective of a GRDS is to provide an easy way for users to locate what they need that is based on their own understanding of that need and where it can be satisfied. The GRDS in other words must match the inherent behaviour(s) of the user. However, many GRDSs often implicitly or explicitly expect users to adapt their behaviour to meet the structure of both government and the GRDS.

The process of building a Web site that allows normal or intuitive user behaviour to find content requires the formal development and implementation of an Information Architecture (IA). The importance of an Information Architecture as a formal discipline was highlighted by a recent summit (American Society for Information Science 2000). A useful definition of Information Architecture, provided by Louis Rosenfeld (2000 p. 002.htm), is that it:

"involves the design of organization, labeling, navigation, and searching systems to help people find and manage information more successfully."

A GRDS Information Architecture includes user needs, content, and user behaviour and cannot be effectively developed without due reference to all three components. Nor can an Information Architecture taken from other sites (e.g. Open Web portals) be simply adopted for a GRDS.

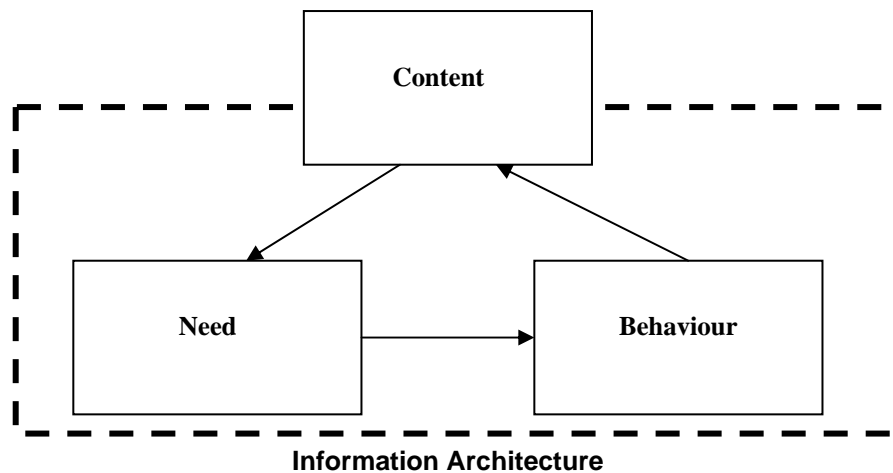


Figure 2 The role of an Information Architecture for a GRDS

### The identification of user needs

Although the need to adequately identify user requirements is a key requirement for an Information Architecture (Head 2000), it should be relatively clear that a GRDS has an easier task in defining user needs and IA objectives than other portal types. This is because the type of needs to be satisfied and the resources needed are relatively well defined by government itself.

The needs of the community regarding online government information and services is dictated by their specific individual requirements, socio-economic characteristics, and most importantly, the types of services provided by government. It is not within the scope of a GRDS IA to influence or develop those user needs, this comes within the broader sphere of government policy and activities, and is pursued at the specific service level by the government agency concerned.

The GRDS, through effective and efficient operation, can create a climate where the user's perception of the range of needs that can be satisfied through an online environment will increase. The GRDS is in fact a major tool for government to promote its online services in a Whole of Government sense, and to realise the individual expenditure made across specific agencies.

Specific online services can and should be promoted and advertised to ensure that the community are aware of their availability. The challenges and issues related to content promotion and advertising at the agency level are beyond the scope of this paper, but should be considered when GRDS roles and functions across government are considered.

Although the GRDS acts as an intermediary in the online discovery process, the user does not see the GRDS as a separate entity when it comes to the delivery of online services. Experience with both *Tasmania Online* and *ServiceTasmania Online* has shown that the GRDS may be asked to assist when a certain service is not available, or is not operational. This is despite the fact that the service is provided by a different government agency.

In such cases, the issue should be passed on to the appropriate agency for action. Such a process may or may not fit within the online content creation and maintenance strategies of a given agency, and it is usually up to the agency concerned to interact with the user to achieve a satisfactory conclusion. However, there will be situations where the number of recurring problems or the nature of the unsatisfied needs become apparent to the GRDS and even detrimental to the performance and credibility of the GRDS. In such cases the GRDS should have the capacity to influence strategies at the agency level. This is a policy issue that cuts across the normal autonomy practiced by government agencies, but is one that must be addressed if a GRDS is to meet client needs effectively and to reflect Whole of Government objectives on behalf of government.

## User needs and discovery outcomes

The search outcomes expected by the user are a major factor in developing the GRDS Information Architecture. The behaviour of users of the *Tasmania Online* and *ServiceTasmania Online* web sites has been analyzed for the past three months in order to identify expected outcomes. This analysis has included the discovery tools most commonly adopted, the free-text terms most commonly entered on the search engine, and the subject or directory structures and content most frequently visited. This has been supplemented by experience in managing a GRDS help desk (email-based).

The outcomes were originally tested against a three-fold information-seeking behaviour schema of known item, unknown item and serendipitous browsing (Gordon 2000). However the types of search outcomes that were identified led to the refinement of known and unknown item into the following three categories:

- 1) **Known item searching:** the user is looking for a specific item that they know exists online, e.g. the home page for the consumer affairs office, the Tasmanian Dangerous Goods Act of 1997, an application form for a business license, etc.; users will expect to find the specific item easily in their search
- 2) **Known resource searching:** the user is looking for a limited set of specific resources which they expect to find online if they exist: e.g. consumer affairs support, legislation related to dangerous chemicals, business license information; users will expect to find all relevant resources easily in their search
- 3) **Known topic searching:** the user is looking for information on a particular subject but has little knowledge as to the breadth and nature of resources that exist: e.g. consumer affairs, legal issues, establishing a business; users will expect to find all relevant resources easily in their search but will also expect to contribute to the process of selecting and evaluating relevant resources. The difference between Known Resource and Known Topic searching is largely the specificity of the search, but this distinction is important because it affects user behaviour.

These three outcomes are indicative of users who have already made a choice to seek resources that they know or expect to be provided by government. They reflect a need for the GRDS to provide

precise and accurate result sets within an Information Architecture that caters for a variety of approaches and content.

The other major type of search behaviour, that of serendipitous searching, has been discounted here because it was difficult to identify, and because it does not at this stage have significant relevance to the development of a GRDS IA. Navigation structures and site design that support serendipitous searching have their greatest value on entertainment sites and eCommerce portals where browsing and impulse retrieval is important (Gordon 2000).

### **Government content from the users perspective**

On the open Web, the scope of the content being sought by the user is only limited by the imagination of the user. Most users also accept that open Web content is not part of a managed publishing environment with reliable quality control processes. The user of a GRDS will, however, expect that the Government content accessed through the GRDS is accurate and current.

The expected content of a GRDS is relatively easy to scope, at least in general terms: all that comes under the aegis of government. Problems of scope definition do arise, however, when the different content across government jurisdictions is raised, as is content that comes from quasi-government agencies, or from activities and services that are privatised, out-sourced, or mixed.

If the user-centric model is to be properly applied by a GRDS, then the GRDS IA should not require the user to understand how government is structured internally or externally. The GRDS may need to access content that comes from organisations and entities that may not technically be government bodies and that may even be in the commercial sector. Certainly the GRDS should provide seamless access to resources across all of its own government agencies, but it should also include appropriate access to content in other jurisdictions or levels of government. This cross-jurisdictional approach is a key element to the *Governet* (Office of Government Online 2000a) initiative currently being pursued by the Commonwealth and States.

Tight or narrow content definitions are not likely to be effective in modifying user behaviour no matter how well promoted or explained. An analysis of common search terms on the *Tasmania Online* search engine that yielded no results over three months (May-July 2000) showed 18% of the terms entered were outside the clearly defined scope of the site. This was despite the prominent explanation of the type of content to be found on *Tasmania Online*.

### **User behaviour - the key to an effective Information Architecture**

"The user wants resources bundled in terms of their own interests and needs, not determined by the constraints and capabilities of the supplier, or by arbitrary historical practices." - (Dempsey 2000)

Established research into information-seeking behaviour has focused on traditional information seeking environments such as libraries, research organisations, and academic institutions. The processes and information retrieval techniques developed from such an environment tend to emphasize specialised services appropriate to the subject area, skilled intermediaries, and appropriate training for the information seeker.

The challenge that the World Wide Web introduces is that the individual information-seeker has become a generic and nebulous individual. The user is now the ordinary person, usually lacking training, possibly with minimal skills or aptitude for search formulation, but with high expectations that are fueled by the hype of search engines and the Internet. Yet discovery tools rarely reflect this.

For example a study of Internet searching behaviour found that the use of search engines is not intuitive to the average user (Pollock 1997). Most users will not take the time to understand the concepts that underlie effective searching, or the processes and specific strategies that apply to a given search engine.

Discovery tools that reflect and accommodate intuitive or natural user behaviour are likely to be more effective than the provision of advanced search options with explanatory text or help pages. At *Tasmania Online*, a 3-month analysis of the search engine help page showed that the help page was consulted 0.01 % of the time. This is despite the help option being placed in an obvious location next to the input box of the search engine.

The need to accommodate user behaviour is made more challenging because users do not simply exhibit a single behaviour type. Their behaviours will reflect different states of knowledge in a given subject area, different assumptions about content, and the different social and emotional contexts that define specific user needs.

The solution to this issue is to provide a variety of discovery tools and to design the web site so as to maximise their availability. A great deal of knowledge has developed over the last five years on how best to structure a Web to maximise usability (Useable Web 2000). The visual and networked characteristics of the Web entail a complex relationship between function, content, design, and layout. The principles of good design and effective layout are beyond the scope of this paper, but must be included in the IA of a GRDS.

The main discovery tools found on the Web include: tables of contents, menus, navigation bars or graphics, alphabetical lists, site maps, special indices, subject directories, and text-query search engines. Of greatest interest to this discussion of a GRDS are those tools that go beyond the structure of the Web site itself. These are the search engine, subject directory, special indices, and alphabetical listings.

## **Appropriate discovery tool design for a GRDS**

### **Search engines**

Search engines were developed to provide access to an unstructured information environment where the textual content of the Web pages could be indexed simply and easily. This type of search is not a keyword search as it is often mistakenly labeled, but instead relies on the serendipitous use of words by the content creator.

Retrieval on the basis of these words is also serendipitous, made somewhat effective by the huge range of resources available, and the likelihood of common terms being used by both the searcher and the author. This can be augmented by the author through the use of specifically assigned terms added to the web page to assist searching - e.g. keywords. None the less, the major problem for search engines is that extremely large result sets are often returned that contain many resources not relevant to the query. This is high recall with low relevance.

### **Leveraging the search term context**

Given the narrow scope of GRDS content, and the known context of searches on a GRDS, the problem of low relevance and high recall can be addressed through selective result sets and tailored ranking. Because of the need for known item and known resource outcomes discussed earlier, it is appropriate to anticipate or predict those sites that should be returned with major prominence. Prominence can be achieved by either reducing the number of results returned, or by ensuring that the most probable sites are returned high in the results list.

An analysis of the top twenty search terms for Tasmania Online's search engine over three months showed that many terms had specific government applications. For example, the term 'gazette' was consistently in the top twenty search terms. It is clear from a GRDS context in Tasmania that a search for the term 'gazette' was in fact a search for government job advertisements ('gazette' is a colloquial

term for the Tasmanian State Service Notices that contains job advertisements). It should be possible to build on this knowledge to ensure that a GRDS search engine in Tasmania returns those key items (such shortcut searches are now provided on *Tasmania Online*).

The results of the searches for the top twenty search terms were assessed according to the placement in the result set of the resource considered most appropriate to satisfy that search. From this analysis, the most likely resource appeared on the first results screen for 8 of the 20 terms, a success rate of 40 %. Given the experience that many users have on the Web, this may be quite a satisfactory result.

*ServiceTasmania Online*, on the other hand, has developed a search service specifically designed for the needs of a GRDS. The same twenty terms applied to the *ServiceTasmania Online* site's search engine yielded the correct result in the top ten results in 17 out of the 20 searches, a success rate of 85%.

## **Metadata**

The need to improve resource discovery on the open Web has led to the development of a standard way to describe Web resources using metadata. Metadata can be defined as a formal description about a resource, and the most important discovery metadata standard is called Dublin Core (Dublin Core Metadata Initiative 2000). Using metadata, search engine software should be able to provide more accurate results (increased relevance and precision with reduced recall). An important objective of Dublin Core (DC) is that it be simple enough to allow content creators to easily and cheaply add that metadata to their own Web pages.

This potential value of DC rises when the online content becomes subject or domain specific. The Australian Government has formally adopted Dublin Core with specific modifications and additions to reflect Australian Government needs. This standard, the Australian Government Locator Service (AGLS), is maintained by the National Archives of Australia and has been adopted by most States as the metadata standard for government Web resources (National Archives of Australia 2000).

In terms of the GRDS context, it is clear that AGLS metadata has the potential to improve the performance of GRDS search engines and the specificity and recall of the results returned to users. However, the actual effect of AGLS metadata for improved retrieval is directly related to the quality of that metadata.

In a previous paper (Sokvitne 2000), the author looked at the existing metadata across various Australian government web sites and found that less than half of the titles used in the metadata actually reflected the obvious title from the web page. Less than one quarter of the metadata descriptions for creator/publisher used a standard form of author notation, and 79% of subject entries were too general to facilitate focused retrieval.

The *Tasmania Online* search engine is able to specifically index and search certain Dublin Core metadata fields, including subject. Recently, an analysis of the top twenty search terms over three months was conducted using the metadata option to limit the search for those terms. Unexpectedly, the number of times an acceptable resource was presented on the first page of results was 20%, as compared to 40% for free text searches that didn't rely on metadata.

Although a portion of this drop occurred because some of the expected resources did not have metadata, three out of 20 key resources had metadata that was incomplete or incorrect. Of equal concern, an investigation of the other non-relevant records presented to the user found that many had incorrect metadata which resulted in their placement high in the results set. This result again demonstrates that it is the quality of all the metadata that matters, but also illustrates that a proportion of poor metadata can distort result sets and counteract the benefits of quality metadata.

It must be stated, however that both of these analyses reflect an uncontrolled environment for distributed metadata authorship, with no established processes in place for training and promoting quality amongst the metadata creators. However, it remains a valid question as to whether any training program could adequately up-skill and motivate the content creators to spend the necessary time in

metadata creation and updating. This is still an area where standards are not easy to implement and where consistency only comes through either sophisticated toolset development or highly trained and capable human input.

*ServiceTasmania Online* has, since April 2000, created full AGLS metadata records for government resources and utilises a range of specifically developed tools to facilitate and ensure quality control over that process. This experience at *ServiceTasmania Online* has shown that consistent and accurate metadata creation is a difficult process. Experienced cataloguers (professional librarians) have themselves required up to two weeks to develop the specific skills and mind-set necessary to adequately produce accurate metadata. Nor should it be assumed that all content creators have the full range of abilities necessary to create metadata. From our experience, it is unrealistic to expect a diverse and distributed range of content creators across government to add quality metadata.

The creation of quality metadata for GRDS presents a difficult business issue for government, in that the specific resources are not necessarily available to fund the independent development and creation of metadata. This has led to an emphasis on distributed creation, typically by the authors of the content itself. Sensibly, this process is to be supported by well-developed standards and tools, and adequate training (Office for Government Online 2000b). However, such a process itself becomes costly, and one can ask whether it might not in fact be cheaper to centralise or consolidate metadata creation to centres of excellence. Roy Tennant (2000) has recently described metadata as "cataloguing done by people paid more than librarians". In other words, it may be cheaper to employ professionals.

## **Subject directories**

Another key discovery tool for a GRDS is the subject directory. Directories provide the user with a logical hierarchical structure, with the ability to start from a general concept and narrow that concept down to specific subject areas and resources. Most hierarchies and directories are based on generic taxonomies and ontologies, in that they attempt to classify the full range of knowledge. However they can also include geographic or other aspects of the content (e.g. content format: on-line forms, images, etc).

The advantages provided by hierarchies make them attractive to many users over search engines. The user can select, even browse, their way to the subject they wish. This is very important if the client is unable to verbalise the concept they are looking for, or if the concept is difficult to verbalise, especially if the obvious words are so broad as to make free text searching undesirable.

Jakob Nielsen (2000a) states that search engine based discovery behaviour accounts for slightly more than half of the discovery strategies adopted by the average user. This research is presumably based on open Web searching and its validity was tested against *ServiceTasmania Online* usage. Over the May to July period at *ServiceTasmania Online*, search engine searching accounted for 34% of use, subject directories and indices for 62%, and alphabetical listings for 4%. This suggests that in the GRDS environment a well presented and maintained directory is actually more important to the users than a search engine.

Like search engines, the concepts of precision, recall and relevance are important measures for a hierarchical-based GRDS discovery tool. An effective hierarchy is one that provides a small number of relevant results at any given level. For a GRDS this is possible because of the known scope and outcomes expected by the user.

To achieve appropriate levels of relevance and precision, the taxonomy used on a GRDS should be developed specifically for the GRDS. *ServiceTasmania Online* found that taxonomies developed in other arenas or for other purposes and employed on a GRDS did not have the specificity where required. Nor did such taxonomies group related government services together. As a result *ServiceTasmania Online* developed a specific hierarchy based on a purpose built taxonomy that relates to government functions, that provides a simple browseable path to specific resources, and that keeps the number of results at each level to a manageable number.

## Specialised Indices

Specialised indices are specifically constructed indexes that meet a specific purpose or address a special need. A specialised index does not necessarily reflect the predominant directory structure or search tool of a site, and is often the product of a manual indexing process so as to accommodate alternate views and usages of the main data resource.

In the GRDS environment, specialised indices can be provided to meet the task orientation of the client (e.g. make a payment), or by the socio-economic context of the discovery need (e.g. a parent of a newborn, a person with disabilities, a farmer, etc). The range of specialised indices provided should be designed to meet user needs and their expectations of government, and as such may change over time.

*ServiceTasmania Online* offers a number of specialised indices. Special options are available on the *ServiceTasmania Online* web site for users who are focused on the task they wish to perform, e.g. to make a payment, application, contact government, etc. Another type of listing is available that addresses the needs of identified social and economic groups, e.g. farmers, women, people with disabilities, and youth. Other options are available to reflect major events such as the birth of a child. All of these options rely on value-added indexing where government content is seen as having more than one dimension or context.

The Information Architecture of a GRDS should allow specialised indices to interact with the other discovery tools provided. For example, audience-based indices should permit free-text searching of their content (i.e. the ability to search for a given word/phrase within the content of all resources under the 'farmers' special index). Task-based indices should be combinable with topic-based searching (i.e., the ability to choose payments as a task within a subject hierarchy). This type of flexible navigation allows the user to adopt different approaches to reach the same resource and is an example of the GRDS addressing alternate user behaviours.

This type of flexible hierarchy may require special software development and an example of this type of functionality has been implemented at *ServiceTasmania Online*. For example, the user can start with the concept of 'Health' and then select 'Payments' ( from a task-based index) or 'Seniors' (from an audience-based index) as a refinement option. The search can also start with 'Seniors' and then choose a subject or task, or do a free text search within those resources.

## Alphabetical listings

Alphabetical listings support known item searches and provide fast and precise returns to the user. An alphabetical list of government departments is an example of a traditional alphabetical list. Such lists are clearly inadequate as the only type of discovery tool. But they are important as an adjunct to other retrieval options and the user will be frustrated if they are unable to access such immediate and precise results for know item searching.

Research from *ServiceTasmania Online* over the past three months has shown that access to a listing of government organisations represents 4% of total accesses on the site. *Tasmania Online* offers a search engine, directory listing, and alphabetical title listing of Web sites. For the same period, alphabetical listings accounted for 33% of site usage. This high figure shows the value of alphabetical searching for known item searches.

## Building an integrated Information Architecture around the discovery tools.

The concept of surrogates in searching has been well described by Carl Lagoze and others (Lagoze 1997, Payette 2000). A surrogate is a unit of information about the actual resource that is presented to the client as the result of a search. This surrogate allows the user to efficiently review that information to determine if the actual item should be visited or retrieved.

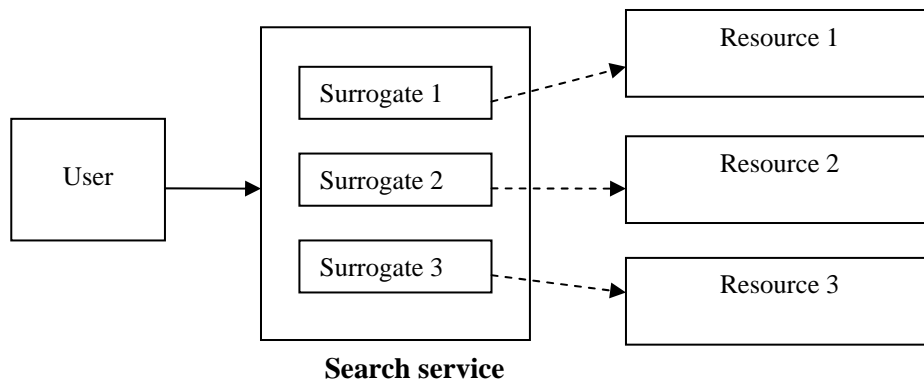


Figure 3 - Surrogates act as an intermediary between the user and the resource

Search Engine results lists use simple surrogates by providing the URL, page title and a description for each resource listed. Surrogates are typically created automatically by the search engine at the time of index creation, and are at risk of being inaccurate at the time of display (e.g. the site no longer exists when selected from the results list).

Dynamic surrogates could be created at the time of display by real-time referral to the resource itself - for example every time a resource is about to be displayed the search service updates the description from the actual Web site. However dynamic surrogate provision is rare given the current variability in network response time and the degree of collaboration and standardisation required by content producers.

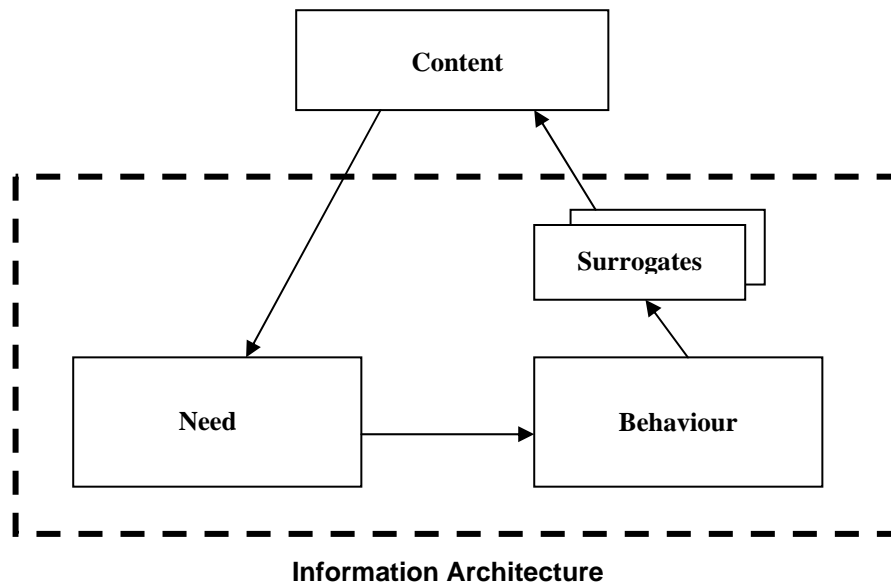


Figure 4 The role of surrogates within an Information Architecture for a GRDS

### Metadata as a building block for surrogate-based Information Architecture

Surrogates that use and display AGLS metadata could be created through the gathering process employed by most modern search engine software. The weakness of this approach is that the

surrogates depend entirely on the existence and quality of the AGLS metadata harvested. Another potential problem is that this type of surrogate production produces one surrogate for every resource harvested. The large number of surrogates produced will produce large result lists thereby increasing recall and reducing relevance. And as discussed previously, the mixture of AGLS metadata and non-metadata content found across a large dataset can be detrimental to the overall results.

However if these AGLS surrogates are developed and maintained separate to the resource, a number of significant advantages emerge. It becomes possible to describe a whole range of resources, not just text-based Web pages. The content of the surrogates becomes easy to maintain and update because the content owner is not required to change embedded page data. The metadata itself can have higher standards of quality control and can be developed and augmented against the context of the GRDS itself, providing improved consistency and relevance across the range of government resources.

Most importantly independent surrogate provision allows the GRDS to create surrogates for only those resources that require discovery by the search service. For a GRDS, the user is likely to expect a high degree of precision and relevance, and will accept and even expect that a selection process has occurred to ensure that only high quality resources are indexed and provided. Traditional search engine and harvesting software do not produce the type of selection results desirable on a GRDS. Tight and focussed result sets are feasible if the GRDS assumes that not every Web page and resource produced by government agencies must be indexed or delivered in response to a given query.

Policies are needed to determine which resources should be included in a GRDS, and the AGLS application guide offers useful assistance here in terms of those resources that should receive priority for metadata creation. However, the most important source must be the user and this will be an ongoing process as online content grows and user needs change. The use of selected surrogates requires the GRDS to be seen as a dynamic entity, constantly monitoring its performance and changing accessible content as needs dictate.

### **Surrogates provide more than one-dimensional discovery**

The real power of a surrogate-based GRDS model is that the surrogate data can be used to provide an integrated, easy to maintain, and efficient Information Architecture. Surrogate data can be used to dynamically populate a whole range of GRDS navigation and presentation structures, including search engines, directories, specialised listings and alphabetical lists. Such an approach suggests, but is not limited to, a database approach to Information Architecture.

At *ServiceTasmania Online* an Information Architecture has been developed that uses surrogate data to manage the Web site, provide dynamic navigation, context-based hierarchies with fluid structures, specialised lists (payments, farmers), alphabetical lists, as well as free-text searching. These surrogates exist as XML files and contain AGLS metadata as well as other data descriptions to populate the *ServiceTasmania Online* navigation structures and indices. These XML files can be updated and maintained independently of the resource itself. The resources covered include State, Commonwealth, and local government content.

The nature of this surrogate-based architecture allows *ServiceTasmania Online* to provide a service independent of the metadata provided or not provided by various jurisdictions. It ensures the user receives high quality, consistent, and comprehensive results for GRDS queries. To counter the problem of the separation between surrogate and resource, routine software processes verify the existence of the described resource and remove the resource from the user display if no longer available on the Web.

The *ServiceTasmania Online* hierarchy has itself been created to meet the needs of government resources and a GRDS. Text-based searching options are available that encompass the high-quality metadata within surrogates and the free-text found within the pages referred to by those surrogates. A non-surrogate based text search option is also available that covers all resources found by a recursive harvesting of government Web content.

Government resources are carefully selected for GRDS delivery and consequent inclusion in the search indexes and hierarchies. The basic policies for resource selection have been driven by demonstrated searching strategies from previous/related sites and reasonable assumptions about user needs and appropriate government content. Much however will depend on the actual analysis of site usage and special software is being developed to allow this to occur. In due course client surveys and focus groups will also be used to augment and understand user needs.

At present approximately 3000 surrogate records have been created, representing less than 5% of the estimated online resource available across government. However these 3000 records represent the demand points for online services covering all three levels of government in Tasmania, and provide comprehensive and quality results for Tasmania users. The user is presented with the best resource known that addresses the query, although at times this may be an entry point to a fuller range of detailed resources provided on a government agency Web site.

Quality control and internal standards adherence are fundamental to the operation of *ServiceTasmania Online*. Input forms with validation processes and controlled vocabularies have been developed to ensure adherence to the necessary standards. Content and user behaviours are monitored and the hierarchies and search metadata modified as user needs become apparent. *ServiceTasmania Online* is still a developing service but will always be in a state of refinement and improvement. This however, is a strength not a weakness.

## **The development for a Whole of Government GRDS**

The development and provision of a GRDS based on the type of Information Architecture described in this paper requires a specific operational unit within government and will have a definite cost. The apportionment or redirection of funds to such a service will be a difficult task when government resources are limited. The cost of a GRDS may also be seen as a stand-alone and additional cost, lacking a persuasive business case to justify that expenditure. A GRDS business case will need to gain support from the highest executive levels within government due to the financial and cross-agency issues that are raised.

Governments employ more than simple monetary measures to assess its services as these often derive from social, political, and legislative obligations. The type of business case to support a GRDS will through necessity be complex and rely on social benefits and Whole of Government returns. Although such a business case is too complex to develop in detail here, it is relevant and necessary to at least ask whether a GRDS can generate a realistic business case.

### **Possible business case measures**

The basis for a GRDS business case is the argument that the expenditure within government of producing online content that is not found and used is lost expenditure. A GRDS is considered a fundamental and necessary tool to locate government content, and the difficulties for the user to locate content across a variety of government Agencies and jurisdictions, each with their own navigation structure and content style, can be demonstrated. The cost of the GRDS is therefore a cost necessary to government to realise overall government expenditure.

To support this, the cost of the GRDS can be compared against the actual cost of government online content production and delivery. Although the quantification of an absolute dollar amount for government online production is impossible here, it is possible to estimate the proportional costs of providing a GRDS against the cost of content production.

From this standpoint, experience in Tasmania suggests that any given Web page with significant content on a government server will be the result of at least a combined total of one hour's work - by the creator, html editor/translator, policy editor, etc. This average unit production time of an hour is likely to be a very conservative estimate.

Experience at *ServiceTasmania Online* also suggests the total indexing needs for government by a GRDS is likely to be between 5-10% of the total government resources produced. At *ServiceTasmania Online* an experienced metadata cataloguer can produce a metadata record in 1/4 of the unit production time for a resource. This produces a proportional cost for GRDS surrogate production of ca 2% of the cost of total content production. When the Whole of Government costs for Web computer and communications hardware acquisition, management and maintenance, and telecommunication costs are factored in, a conservative estimate would place the cost of GRDS indexing at less than 0.5% of government online expenditure.

As a proportional cost for web publishing across government it is difficult to argue that such a cost is unreasonable, especially when such a cost helps to maximise the return on the other investments made by government. In the private sector marketing is a cost seen as necessary to maximise production outcomes. Marketing is not a cost that requires justification as an independent figure, but is measured and justified according to overall returns and the full cost of production. In fact marketing is considered a cost of production. Resource discovery in an online government environment should also simply be considered an inherent cost of production.

It is difficult to quantify in economic terms the value to a member of the community of being able to find a desired item of information in an online GRDS environment. The convenience of accessing services or obtaining information in the home is apparent but not easily quantified. Although a small proportion of the community at present use online services, those services are available on a 7 day - 24 hour basis, and provide services in remote and isolated areas. The cost of online services is clearly cheaper than matching the same breadth of temporal and spatial coverage with physical services. And as the community and economy generally move into the online environment, government must position itself to ensure continued service relevance to its community.

## Conclusion

This paper has attempted to create a conceptual framework that defines the essential characteristics of a GRDS and has, where possible, tested that framework against general research in the area, empirical research from the Tasmanian Government Web site, and experience in operational service provision.

The necessary features of a GRDS have been identified as:

- 1) A defined and purpose built Information Architecture that encompasses and integrates user needs, behaviour, and the type of content available
- 2) Multiple searching methodologies that include:
  - government appropriate free text search engine
  - government appropriate subject directories and taxonomies
  - a variety of indices and listings to cater for a variety of users
- 3) A surrogate based system to limit and accurately describe government content
- 4) The use of enhanced surrogates to provide searching, site maintenance, and site navigation.

Such a service will involve a distinct cost to government but this expenditure can be justified if the desired outcomes of online service delivery across government are to be realised.

However, this is such a new field that the absence of conceptual models, detailed research, and shared experience is by far the most dominant problem for GRDS implementation. As government is the major stakeholder in such research, it is most appropriate that the conceptual and empirical development come from government. This research should include:

- An analysis of user behaviours, search strategies, skills
- The design characteristics of effective GRDS navigation structures and tools
- The development of effective taxonomies for government use
- The development of effective selection policies and processes to allow selective but efficient access to the full range of government resources
- Accurate costing models for online content production, online service provision, and the development of measures to evaluate indexing costs and relative efficiencies.

The experience at *ServiceTasmania Online* has served as the starting point for identifying the components of an effective GRDS, but alternate approaches and the experience of other jurisdictions also need consideration. However, the user-centric GRDS model proposed in this paper, linking needs, content, and behaviour, is fundamental to any GRDS.

## References and Bibliography

American Society for Information Science (2000) **ASIS Summit 2000: Defining Information Architecture**. [Internet]. Boston. Available from: < [http://www.asis.org/Conferences/Summit2000/Information\\_Architecture/index.html](http://www.asis.org/Conferences/Summit2000/Information_Architecture/index.html) > [Accessed 24 July, 2000]

Australian Defence Force Academy (2000) **Metachem: Welcome to a catalogue of chemistry resources**. [Internet]. Available from: < <http://metachem.ch.adfa.edu.au> > [Accessed 26 July, 2000]

Campbell, D. (2000) **Australian subject gateways - metadata as an agent of change**, IN: VALA2000 Conference: Books and Bytes: Technologies for the Hybrid Library. [Internet, Pdf file]. Melbourne, Victorian Association for Library Automation. Available from: < <http://www.vala.org.au/vala2000/2000pprs/fri2000.htm> > [Accessed 24 July, 2000]

Cheng, Y. & Shaw, D. (1999) Information Seeking and Finding. **Bulletin of the American Society for Information Science**. [Internet] Vol 25, No. 3, February/March 1999. Available from: < <http://www.asis.org/Bulletin/Feb-99/introduction.html> > [Accessed 24 July, 2000]

Dempsey, L. (2000) Scientific, Industrial, and Cultural Heritage: a shared approach, a research framework for digital libraries, museum, and archives. **Ariadne**. [Internet]. Issue 22. Available from: < <http://www.ariadne.ac.uk/issue22/dempsey/> > [Accessed 22 July, 2000]

Dublin Core Metadata Initiative. (2000) **Dublin Core Metadata Initiative**. [Internet]. Available from: < <http://purl.org/DC/index.htm> > [Accessed 26 July, 2000]

Education Network Australia. (2000) **EdNA**. [Internet] Available from: < <http://www.edna.edu.au> > [Accessed 26 July 2000]

Erdelez, S. (1999) Information Encountering: Its More Than Just Bumping into Information. **Bulletin of the American Society for Information Science**. [Internet]. Vol 25, No. 3, February/March 1999. Available from: < <http://www.asis.org/Bulletin/Feb-99/erdelez.html> > [Accessed 24 July, 2000]

Gordon, S. (2000) Information Architecture. **Bulletin of the American Society for Information Science**. [Internet]. Vol 26, No. 5, June/July 2000. Available from: < <http://www.asis.org/Conferences/Summit2000/gordon/index.htm> > [Accessed 24 July, 2000]

Head, A. J. (2000) At the Crossroads: Information Architecture and the "Purpose Crisis".. **Bulletin of the American Society for Information Science**. [Internet]. Vol 26, No. 5, June/July 2000. Available from: < <http://www.asis.org/Conferences/Summit2000/ajhead/sld003.htm> > [Accessed 24 July 2000]

Lagoze, C. (1997) From Static to Dynamic Surrogates: Resource Discovery in the Digital Age. **D-Lib Magazine: The Magazine of Digital Library Research**. [Internet]. June 1997. Available from: < <http://www.dlib.org/dlib/june97/06lagoze.html> > [Accessed 24 July, 2000]

Morville, P. (2000) Defining Information Architecture: "Strange Connections". **Bulletin of the American Society for Information Science**. [Internet]. Vol 26, No. 5, June/July 2000 Available from: < <http://www.asis.org/Conferences/Summit2000/morville/index.htm> > [Accessed 26 July 2000]

Morville, P. & Rosenfeld, L. (1998) Designing Navigation Systems **Webreview.com: Cross-training for web teams**. [Internet] 20 February 1998, Available from:  
< <http://www.webreview.com/pub/98/02/20/arch/index.html> > [Accessed 20 July 2000]

National Archives of Australia. (2000) **The Australian Government Locator Service (AGLS)**. [Internet] Available from: < [http://www.naa.gov.au/recordkeeping/gov\\_online/agls/summary.html](http://www.naa.gov.au/recordkeeping/gov_online/agls/summary.html) > [Accessed 26 July, 2000]

Nielsen, J. (2000) Search and You May Find, **Useit.com Alertbox**, [Internet] Available from:  
< <http://www.useit.com/alertbox/9707b.html> > [Accessed 24 July, 2000]

Nielsen, J. (2000) **Useit.com: usable information Technology** [Internet] Available from:  
< <http://www.useit.com> > [Accessed 24 July, 2000]

Office for Government Online. (2000). **Governnet**. [Internet] Canberra, OGO. Available from  
<: <http://www.ogo.gov.au/projects/services&innovation/governnet.htm> > [Accessed 24 July, 2000]

Office for Government Online. (2000). **GovernmentOnline**. [Internet]. Canberra, OGO. Available from:  
< <http://www.ogo.gov.au> > [Accessed 24 July, 2000]

Office for Government Online. (2000) **Online Information Service Obligations**. [Internet]. Available from: < <http://www.ogo.gov.au/projects/standards/oiso.htm> > [Accessed 24 July, 2000]

Payette, S. & Lagoze, C. (2000) Value-Added Surrogates for Distributed Content. **D-Lib Magazine**. [Internet]. June 2000. Vol 6 No 6. Available from:  
< <http://www.dlib.org/dlib/june00/payette/06payette.html> > [Accessed 24 July, 2000]

Pollock, A. & Hockly, A. (1997) What's Wrong with Internet Searching?. **D-Lib Magazine: The Magazine of Digital Library Research**. [Internet]. March 1997. Available from:  
< <http://www.dlib.org/dlib/march97/bt/03pollock.html> > [Accessed 24 July, 2000]

Rosenfeld, L. (2000) Making the Case for Information Architecture. **Bulletin of the American Society for Information Science**. [Internet]. Vol 26, No. 5, June/July 2000. Available from:  
< <http://www.asis.org/Conferences/Summit2000/rosenfeld/index.htm> > [Accessed 24 July 2000]

Rosenfeld, L. & Morville, P. (1998) **Information Architecture for the World Wide Web: Designing Large-scale Web Sites**. O'Reilly.

**SearchEngineWatch**. [Internet] Available from: < <http://searchenginewatch.com/facts/spotlight.html> > [Accessed 24 July 2000]

**Searchterms.com: A ranking of what people are searching for on the web**. [Internet] Available from : < <http://www.searchterms.com> > [Accessed 24 July, 2000]

Sokvitne, L. (2000) **An evaluation of the Effectiveness of Current Dublin Core Metadata for Retrieval**. IN: VALA2000 Conference: Books and Bytes: Technologies for the Hybrid Library. [Internet, Pdf file]. Melbourne, Victorian Association for Library Automation. Available from: < <http://www.vala.org.au/vala2000/2000pprs/fri2000.htm> > [Accessed 24 July, 2000]

Tang, X. (2000) Social Informatics and Information Retrieval Systems. **Bulletin of the American Society for Information Science**. [Internet] Vol 26, No. 3, February/March 2000. Available from: < <http://www.asis.org/Bulletin/Mar-00/tang.html> > [Accessed 24 July 2000]

Tasmania. Dept of Premier and Cabinet. (2000) **GIPS: Government Internet Publishing Standards, V 0.1**. [Internet]. Hobart, DPAC. Available from: < <http://www.service.tas.gov.au/govstds/homepage.htm> > [Accessed 26 July, 2000]

State Library of Tasmania. ed. (2000). **ServiceTasmania Online**. [Internet]. Available from:  
< <http://www.service.tas.gov.au> > [Accessed 30 July 2000]

State Library of Tasmania. ed. (2000). **Tasmania Online**. [Internet]. Available from:  
< <http://www.tas.gov.au> > [Accessed 30 July 2000]

Tennant, R. (2000) A Librarian's Perspective on Information Architecture. **Bulletin of the American Society for Information Science**. [Internet], Vol 26, No. 5, June/July 2000. Available from:  
< <http://sunsite.berkeley.edu/~manager/Presentations/ASIS/Boston/> > [Accessed 24 July, 2000]

**The Usable Web: 793 links about web usability**. [Internet]. 2000. Available from:  
< <http://www.usable.web.com> > [Accessed 24 July, 2000]

Veen, J. (2000) The Matrix Integration Architecture for the Masses. **Bulletin of the American Society for Information Science** [Internet]. Vol 26, No. 5, June/July 2000. Available from:  
< <http://www.asis.org/Conferences/Summit2000/veen/index.htm> > [Accessed 24 July, 2000]

**Webreview.com**. [Internet]. Available from: < <http://www.webreview.com> > [Accessed 24 July, 2000]

What people search for. **SearchEngineWatch**. [Internet]. Available from:  
< <http://searchenginewatch.com/facts/searches.html> > [Accessed 24 July, 2000]

Zwies, R. (2000) Observations on the American Society for Information Science Summit 200 Meeting: Defining Information Architecture. **Bulletin of the American Society for Information Science**. [Internet]. Vol 26, No. 5, June/July 2000. Available from: < <http://www.asis.org/Bulletin/June-00/zwies.html> > [Accessed 24 July, 2000]