EUROPE

This PDF document was made available from www.rand.org as a public service of the RAND Corporation.

Jump down to document ▼

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world.

Support RAND

Browse Books & Publications

Make a charitable contribution

For More Information

Visit RAND at www.rand.org

Explore RAND Europe

View document details

Designing a National Standard for Discovery Metadata

Improving Access to Digital Information in the Dutch Government

Jeff Rothenberg, Irma Graafland-Essers, Harry Kranenkamp, Abigail Lierens, Constantijn van Oranje, Rob van Schaik

RAND EUROPE

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

RAND® is a registered trademark.

Colophon

General

Client and policy accountability:	Ministry of the Interior and Kingdom Relations (Ministerie van Binnenlandse Zaken en Koninkrijksrelaties; http://www.minbzk.nl/)
Projectmanagement:	Advies Overheid.nl (http://advies.overheid.nl/)
Report written by:	RAND Europe (http://www.randeurope.org), assisted by Gyata Management Consulting (http://www.gyata.nl) and Doxsupport (http://www.doxsupport.nl)

Rights

The Hague, August 2004

Preface

This report, prepared for and funded by the Advies Overheid.nl and the Ministry of the Interior and Kingdom Relations (BZK), presents the results of a short project on the use of metadata (i.e., information that describes other information) to enhance the transparency of government in the Netherlands. The purpose of this project was to evaluate the risks and issues involved in introducing a nationwide standard for metadata to describe online government information in the Netherlands and to propose such a standard, based on the Dublin Core (DC) metadata standard, which has served as the basis for a number of similar initiatives in other countries. In addition to this report on the project itself, a Handbook was developed to motivate and explain the proposed standard to the government information providers for whom it is intended. The research project began 22 April 2004 and was to be completed 30 September 2004.

This report presents the results of our analysis, organized according to the three major phases in which the project was conducted. The first phase examined the needs of government information providers, to enable us to understand their concerns and what a metadata standard should offer them. The second phase examined the Dublin Core and a number of other national metadata standards, to enable us to understand their capabilities and limitations and to propose a discovery metadata standard for the Netherlands. The third phase explored implementation issues that had emerged in the first two phases. Our conclusions and recommendations are drawn from the results of all three of these phases. The report should be of interest to information providers, such as website and database managers in government organizations at all levels.

The research was conducted by a consortium under the leadership of RAND Europe (Leiden, the Netherlands) and included Gyata Management Consulting (Rotterdam, Netherlands), Doxsupport (Oosthuizen, Netherlands) and the RAND headquarters office in Santa Monica, California, United States.

RAND Europe is an independent not-for-profit policy research organization that serves the public interest by improving policymaking and informing public debate. Clients are European governments, institutions, and firms with a need for rigorous, impartial, multidisciplinary analysis of the hardest problems they face. This report has been peer-reviewed in accordance with RAND's quality assurance standards (see http://www.rand.org/about/standards/) and therefore may be represented as a RAND Europe product.

For more information about this report, please contact: metadata@rand.org.

For more information about this project, please contact:

For more information about RAND Europe, please contact:

Contents

Summary

There are currently more than 1200 separate government websites in the Netherlands, but the relationships among the documents and data at these sites are not always obvious, and most of the information is in unstructured, textual form, making it difficult to find desired information or to compare or compile information from distinct sites. Yet the action program 'Andere Overheid' ("Transfiguring government" of the Netherlands Cabinet) of December 2003, section 1.1 states that: "For www.overheid.nl and other government portals, a search engine will have to be developed that can quickly and efficiently show the citizen the government information, services and organizations he is looking for." All information whose publication is mandated by law must be made available via a government search engine by 2007.

This report presents the results of a short project conducted by RAND Europe and Gyata Management Consulting for Advies Overheid.nl, which maintains the government's portal website www.overheid.nl and was assigned by the Netherlands Ministry of the Interior and Kingdom Relations to develop the above mentioned search engine and to develop projects that aim to enhance the transparency of the Netherlands' government on the internet.

This project examined the issues and evaluated the risks involved in introducing a nationwide standard for metadata to describe online government information in the Netherlands and proposed such a standard, based on the Dublin Core metadata standard, which has served as the basis for a number of similar initiatives in other countries. The project also developed a Handbook to motivate and explain the proposed standard to the government information providers for whom it is intended.

The primary motivation for a nationwide metadata standard is to help online information providers at all levels of the Dutch government make their resources more easily accessible to citizens and commercial users (companies), where "accessible" means easy to discover, locate, view, and use. In addition, a standard should improve the interoperability of information and services at different government sites, making it easier to compare and combine information from different government sources in consistent and meaningful ways. Finally, a standard should help improve the delivery of eGovernment services.

The project consisted of three major research phases. In the first phase, we examined the needs of government information providers, to understand their concerns and what a metadata standard should offer them. In the second phase, we examined the Dublin Core and a number of other national metadata standards, to understand their capabilities and limitations and to learn from the experiences of other countries. In the third phase, we explored implementation issues that had emerged from the first two phases.

Phase I analyzed Dutch government data and metadata needs, relevant Dutch government initiatives, laws, regulations, and policies, and the EU context. No comprehensive study of the needs of government information users in the Netherlands was available, and conducting such a study was beyond the scope of the project. However, we conducted a small-scale study of the current practices and needs of government information providers, using an online survey, in-depth interviews, and online research focused on the lessons learned by similar national efforts. This revealed many issues and concerns, including the fact that most Dutch government website managers do not see a need for a national metadata standard.

Although agencies and organizations in different levels of government or those dealing with different aspects of government or different subjects may have specialized needs for metadata, we concluded that the standard itself should not embody variant subsets of metadata or specific recommendations for default values or for the use of encodings or controlled vocabularies (CVs) for particular metadata elements. Instead, groups of organizations or individuals should be encouraged to form Communities of Interest (COIs), each of which can tailor such aspects of the standard to its own needs. Furthermore, in order to improve discovery, search terms (such as keywords) should be standardized and organized into taxonomies, thesauri, and ontologies, which can also best be created and maintained by COIs. In order to coordinate the creation and activity of COIs (including the creation and use of CVs), a national clearinghouse for COIs should be established.

In light of the information obtained in Phase I, Phase II of the project examined the Dublin Core standard and some of its competitors, including a number of national standards that have been derived from Dublin Core. Standards that were available in English served as primary input, including the eGMS (e-Government Metadata Standard) from the U.K., the AGLS (originally the Australian Government Locator Service, though no longer restricted to the public sector) and the related VAGLS (from Victoria) and NZGLS (from New Zealand), the Irish Public Service Metadata Standard (IPSMS), and the Canadian Treasury Board Information Management Standard (TBITS), as well as a number of other related efforts, including the CEN Workshop's eGovernment Metadata Application Profile v. 1.0 (CWA 14860, November 2003), GILS (Global Information Locator Service), the EAD (Encoded Archival Description) standard, the Warwick Framework, ISO-IEC_11179 on Metadata Registries (MDR), ISO/TS 23081 on principles of metadata for records management, the OASIS work on search interoperability and Topic Maps, and the European Interoperability Framework (EIF).

 One crucial issue that emerged from this analysis was the need to integrate metadata in the "access domain" (that is, metadata intended to improve discovery and access of information) with records management metadata (used to manage informational assets). When a records management function exists in an organization, it should not only manage that organization's informational assets but should also generate and manage most of the metadata describing those assets. In such cases, most discovery metadata should simply be derived from records management metadata. However, since records management may not be implemented yet in many organizations, the access domain may need to implement an interim surrogate records management function in the interim, to manage its own information resources: this should not involve significantly more effort than that which is

required to maintain a website in the absence of a supporting records management function.

A second important issue that became apparent was the need to allow more structure in metadata than the "flat" (unstructured) set of metadata elements (i.e., fields) provided in the Dublin Core and its derivatives. The widespread use of even a flat metadata set would represent a major step forward in improving discovery of online information in the Netherlands; but there are compelling reasons to allow metadata to be structured, including subsetting, encapsulation, and describing complex resources or relationships. Subsetting allows the creation of groups or subsets of metadata elements, e.g., administrative, records management, rights-management, intellectual property control, or elements that help ensure the longevity of information resources. Encapsulation allows combining metadata elements that are not merely similar in function but are tightly related to each other, such as those describing events (each event in the lifecycle of a resource has a number of different attributes, such as an instigating cause, a date, a responsible party, contributors, status, and result, all of which should ideally be bundled together). Finally, structured metadata facilitates describing resources that are more complex than distinct, static objects like individual documents or records, e.g., those that consist of multiple components or are generated dynamically or correspond to services rather than simple information.

Advies Overheid.nl's initial assumption for this study was that Dublin Core would form the basis of the Dutch national discovery metadata standard, whose temporary, working name is NL-meta. This premise, combined with our identified need for structured metadata led us to design a novel "optionally-structured" approach, in which flat, Dublin Core metadata elements can be augmented by structured metadata when necessary, for example, by using relational, object-oriented, or semantic net approaches.

Our analysis of the Dublin Core and alternative standards led us to adopt a design strategy for the Dutch national metadata standard based on four principles:

1. Use the Dublin Core standard as a starting point

2. Focus on metadata for discovery and access, while addressing the need to integrate the access domain with the records management domain

3. Keep the standard as simple as possible without limiting its future ability to serve the needs of diverse communities to describe the full range of online information resources and services

4. Consider implementation issues including: engaging relevant communities, motivating adoption, managing metadata in metadata systems, and ensuring that metadata can be used by finding-tools and systems

Phase III of the project analyzed implementation issues that had emerged in the previous phases, including the need to engage a wide range of stakeholders to create joint ownership for the standard and the need to create a Metadata Management function within each organization, ideally headed by a Metadata Manager who should develop a Metadata Management Plan and work with internal website managers, records managers, and publication managers, as well as with Advies Overheid.nl and relevant COIs to develop, implement and manage an overall metadata strategy.

We also developed a general roadmap showing a sequence of steps that should enable the Dutch government to meet its goal of providing online access to all legally mandated published information by 2007:

1. Engage all relevant providers of government information.

2. Develop and refine the NL-meta nationwide standard for discovery metadata.

3. Develop and refine "metadata-aware" finding-tools (e.g., search engines) that can access and utilize the discovery metadata in the NL-meta standard to improve discovery.

4. Develop and refine process models for the implementation and use of NL-meta discovery metadata by government information providers.

5. Develop and deploy instrumentation to measure the effectiveness of the online discovery and access of government information.

6. Promote and support the adoption and use of the NL-meta standard, metadata-aware finding-tools and implementation process models by all relevant government information providers.

7. Evaluate the effectiveness of online access to legally mandated published information, and use the results of this evaluation to refine the NL-meta standard, finding-tools, and implementation process models.

Technical implementation issues were also examined. In particular, we considered the question of whether to embed metadata in resources (e.g., using HTML or XML tags) or to keep them separate (e.g., in a database or other metadata management system), and we concluded that whichever approach is taken, the following criteria should be ensured:

- Metadata must remain logically linked to the resources they describe

- Metadata must be accessible to finding-tools

- Modifying metadata should minimize the risk of corrupting resources

- Revising resources should allow appropriate metadata to be easily updated

So long as these criteria are met, it is immaterial whether an embedded or separate metadata scheme (or a hybrid of the two) is adopted.

The general recommendations of the project for information providing organizations are:

1. A Metadata Management function should be created within each information provider organization, ideally headed by a Metadata Manager, and a Metadata Management Plan should be developed and executed by this function.

2. An interim surrogate digital records management function should be created in the access domain, if none already exists in an organization.

3. Communities of Interest (COIs) should be used to address the specialized needs of different user groups, such as choosing default values for specific metadata elements, selecting or developing encoding schemes or CVs, and developing taxonomies, thesauri, and ontologies to organize keywords and other terminology.

4. An optionally-structured approach to metadata should be used to enable subsetting, encapsulating, and describing resources that are more complex than distinct, static objects.

5. Systems that create, manage and use metadata must be considered in addition to metadata per se.

6. Semantic interoperability (i.e., the ability to meaningfully interchange information among different sources and systems) is an important goal that should be addressed by using standard encodings and CVs and by relying on COIs to identify semantic inconsistencies in their domains of interest.

Phase III also identified and analyzed a number of key risk areas:

1. Lack of information about users' needs or capabilities

2. Lack of input from provincial governments, ZBOs, water boards, etc.

3. Lack of involvement of records management and archives

4. Higher cost than benefit

5. Longevity of data and metadata

Mitigation strategies for each of these risk areas are embodied in the project's recommendations to Advies Overheid.nl:

1. Engage a broad stakeholder community via an online working group, including, at least:

   a. Users (e.g., citizens and companies using government information)

   b. Government information providers at all levels

   c. Relevant commercial data producers (such as publishers)

   d. Non-governmental websites offering government information

   e. Records management, archives, libraries and other metadata creators

   f. Commercial vendors of search-engines, records management systems, metadata or database systems, ontology management technology, etc.

   g. Academic institutions researching web use, discovery, and eGovernment

2. Support the adoption and use of the metadata standard:

   a. Facilitate creation and coordination of COIs and CVs

   b. Act as (or establish) a clearinghouse for COIs and CVs

   c. Support tool evaluation, selection and use for example, develop tool evaluation checklists

   d. Provide conformance verification and metadata quality control offered as an online service

   e. Develop and provide guidelines and training for the creation, management, and use of discovery metadata

       f.   For the role of Metadata Manager in information provider organizations

3. Promote a records management strategy for the access domain

4. Conduct follow-on studies to:

       a.   Analyze user needs

       b.   Measure baseline discovery effectiveness (based on user needs) and put in place ongoing instrumentation to measure improvement

       c.   Analyze the relevance of specific encoding schemes for Dutch metadata

       d.   Analyze the possibility of developing a multi-level, modular ontology

       e.   Develop process models for creating, maintaining and improving discovery

       f.   Develop desiderata for discovery-systems and the implied interfaces and services that such systems require of metadata systems

       g.   Develop desiderata for systems to semi-automatically generate, maintain, derive and transform metadata

5. Perform pilot projects to try out and evaluate the approach

The current set of 20 metadata pilot projects that Advies Overheid.nl is conducting may form the start of this effort. However, they should ideally be performed only after a baseline study of the effectiveness of discovery has been conducted and instrumentation for measuring improvements in discovery has been put in place.

A logical sequence of pilot projects aimed at trying out the proposed standard might be:

- Implement the simple NL-meta.DC+ metadata set

- Try out one or more structured NL-meta.Extended approaches

- Explore the use of CVs and ontologies

- Explore the development of metadata to model processes, in order to aid the discovery of online transactions, workflows or e-Services

6. Plan to revise the Handbook

Establish a review and feedback process to revise and evolve both the standard and its associated Handbook to produce a new version within a year or so after its initial release.

Our recommendations are intended to help the Dutch government create and implement a national standard for discovery metadata and promote its adoption and use by government information providers, as a way of improving the transparency of government.

Introduction

An increasing amount of government information, much of it in the form of documents or database records, is becoming available on the Internet. There are currently more than 1200 separate government websites in the Netherlands, but the relationships among the data at these sites are not always obvious, and it is difficult to establish appropriate linkages or associations among related information at different sites. Furthermore, most of these websites provide information in unstructured, textual form, which can make it difficult to find desired information or to compare or compile information from distinct sites.

The action program "Transfiguring government" of the Netherlands Cabinet (December 2003) section 1.1 states that: "For www.overheid.nl and other government portals, a search engine will have to be developed that can quickly and efficiently show the citizen the government information, services and organizations he is looking for."

This report presents the results of a short project conducted by RAND Europe, Gyata Management Consulting and Doxsupport for Advies Overheid.nl on the use of metadata (information that describes other information) to enhance the transparency of government in the Netherlands. The purpose of this project was to examine the issues and evaluate the risks involved in introducing a nationwide standard for metadata to describe online government information in the Netherlands and to propose such a standard, based on the Dublin Core (DC) metadata standard, which has served as the basis for a number of similar initiatives in other countries. In addition to this report on the project itself, a Handbook was developed to motivate and explain the proposed standard to the government information providers for whom it is intended.

The primary motivation for introducing a nationwide metadata standard is to help online information providers at all levels of the Dutch government make their resources more easily accessible to citizens and commercial users (companies), where "accessible" means easy to discover, locate, view, and ultimately use. (Note that although the ultimate purpose of locating and accessing information is to use it in some way, this Report does not focus on use, since doing so would require an understanding of users' needs and purposes, which was unavailable to this study. Metadata can improve use by helping users interpret information correctly, but this may require more than just discovery metadata.)

Making government information more accessible is seen as a key aspect of making government more transparent to its citizens, and one of the widely accepted benefits of metadata is the potential to improve accessibility. A nationwide metadata standard should help information providers describe their resources in a more effective and consistent way, thereby making it easier for users to find and access government information wherever it

exists. For example, if all documents and data related to water were described by a Subject metadata element containing the keyword "water" then searching for this keyword would make it much easier for a user to find all of these documents and data. Given this primary motivation for the standard, its chief functions must be to aid discovery and access. Simply put, the standard should help users discover resources that can help satisfy their information requests, and it should help them access (read or otherwise utilize) those resources once they have discovered them.

In addition, the proposed standard should improve the interoperability of information and services at different government sites, making it easier for users to compare and combine information from different government sources in consistent and meaningful ways. Finally, the standard should help improve the delivery of eGovernment services. All other potential uses of metadata are of secondary concern for this standard.

The proposed standard follows the model of the DC and most national metadata standards in terms of what it standardizes. It specifies mandatory, recommended, and optional elements and refinements, as well as recommended encoding schemes and controlled vocabularies for some (though not all) of these metadata fields, and it follows DC practice by allowing individual organizations to choose their own encoding schemes and controlled vocabularies and to add refinements (sub-elements) when necessary. It is also (like DC) implementation and platform independent: that is, it is intended to be implementable in many different ways on many different computer systems. However, it leaves considerable room for organizations and users to tailor its use to their specific needs, by adding refinements, establishing local default values for various fields, and optionally creating structured metadata (which the DC and its derivatives do not permit) to satisfy a range of potential needs discussed in detail below.

Our "optionally-structured" approach factors the proposed standard into two branches. One of these branches provides unstructured ("flat") metadata analogous to that in the DC standard and its derivatives: this can be used as a baseline implementation of the standard by organizations for which a flat metadata approach provides sufficient capability. The other branch allows the use of structured metadata (discussed further in Section 5.3), for example, to create subsets of related metadata elements, to encapsulate metadata elements that logically belong together, or to describe complex online resources such as dynamic or composite objects or eServices.

Unlike the national metadata standards in certain other countries (such as the U.K.), adoption and use of the proposed standard in the Netherlands is expected to be voluntary. The central government has no intention of imposing the standard as an official mandate for all information. The advantages of adopting and using the standard must therefore be made obvious, or it is unlikely to be used. Furthermore, a metadata standard in and of itself is simply a set of element definitions, optional refinements, encoding schemes, and controlled vocabularies. If the standard is to be adopted and used, the scope of the Dutch national metadata initiative must be expanded to include guidelines for implementing the standard, recommendations for the choice and development of encoding schemes, controlled vocabularies, or other extensions, and support for the acquisition of systems that can create, manage, and use the metadata defined by the standard. In particular, if the standard is to improve the discovery and access of online information, the actual metadata

values that users create under it must be made accessible to finding-tools such as web search engines like Google. Although these concerns are beyond the scope of a metadata standard, per se, they must form a central part of the overall endeavor of improving discovery and access.

There are risks in attempting to introduce such a nationwide standard. The most significant of these risks have to do with implementation, meaning not just the technical aspects of implementing a metadata standard but the entire process of getting it adopted, managed, and used effectively. Foremost among these risks is the possibility that any nationwide standard that is promoted centrally (e.g., by Advies Overheid.nl) may be seen as an unwarranted burden on government agencies. In particular, this is likely to be the case if implementing such a standard requires more effort than an agency feels is worth the benefit that the standard produces for that agency. This in turn is likely to be the case if the standard is overly complex, difficult to implement, inapplicable to an agency's specific needs, or inadequate to significantly improve the accessibility of information. Although central government in the Netherlands does not generally impose official requirements on lower-level government agencies, it is crucial to avoid the perception that a nationwide metadata standard would constitute an unfunded mandate: if such a standard is not willingly adopted by government agencies, it will be a waste of effort. This report and its associated Handbook suggest numerous steps that can be taken to mitigate this and other risks.

Because the motivation for this effort is the enhancement of the transparency of government, the focus of the proposed nationwide metadata standard is on the discovery and access of information--which we refer to as the "access domain" interpreted in the broadest sense. That is, the access domain includes all aspects of public access to government information. In the online context, this includes websites that provide government information, as well as other eGovernment services.

Metadata can support numerous additional functions, including records management, archiving, intellectual property and rights management, quality control, and long-term preservation. Our focus on discovery and access minimizes the burden that the standard places on agencies that conform to it; however, this report and the Handbook discuss various ways in which metadata in the access domain should integrate with metadata in other domains--notably records management. Furthermore, we explore many aspects of the discovery and access of online information that are often ignored in discussions of metadata.

Definitions

This section provides definitions of a number of key terms that are used throughout this report and whose meanings are rather specialized in the context of discovery and access metadata.

A **Community of Interest (COI)** is any group of organizations, agencies, departments, or individuals that share some common concern or interest. For example, different levels of government (national, provincial, municipal) might constitute different COIs, as might non-governmental agencies. In addition, groups of agencies or departments concerned with particular "realms" of government (law and regulations, licenses, European regulations, official publications etc.) or with specific problems or subjects (environmental factors, ethnic concerns, gender issues, etc.) may form COIs that cross levels of government or organizational boundaries. COIs can and often do overlap with each other and cut across traditional hierarchical divisions.

For a definition of **controlled vocabulary** see the discussion of **encoding**.

Discovery in our context is the process of finding information that is of interest. Note that we use the term to mean finding information rather than finding individual resources, since users are often seeking information that spans multiple resources; however, in the context of metadata, discovery typically means finding individual objects, such as documents.

The **Dublin Core** metadata standard consists of a small number of metadata elements, each of which describes some aspect of some resource (such as a document or database). The DC elements are analogous to the data fields found on a card in a library card catalogue. They specify information such as the author, publisher, publication date, format, location, etc. of the resource being described.

Following the terminology of the Dublin Core Metadata Initiative (DCMI), an **encoding scheme** is either a "syntax encoding" or a "vocabulary" though we prefer the term "controlled vocabulary" for the latter. A **syntax encoding** is a specific format for representing a given type of information, such as a date or object identifier. For example, 12-jun-04, June 12, 2004, and 2453168.5 (Julian date) are three ways of encoding the same day, whereas http://overheid.nl/home/sitemap is a way of encoding the name of a web-page. A **controlled vocabulary** (or "CV"), on the other hand, is normally a closed list of terms or identifiers, such as abbreviations for the names of cities or countries, or the names of alternative cartographic systems. For example, the VIND catalogue consists of a

list of services offered by municipalities to citizens and companies; this list might form the basis of a CV describing such services.

A **finding-tool** is any mechanism or service (typically, though not exclusively, a computer program) that can help users find information. This includes search engines (such as Google), database access and query systems, structured browsing programs, or any other tool that can locate and access information. Note that this is broader than the term "finding aid" as it is used in the library and record-keeping communities, where it typically means an index, inventory, abstract or description of some kind rather than an executable program.

Information is a broader term than **resource**, since it may be derived from numerous distinct resources, such as multiple documents or database queries.

The term **Metadata** is typically defined as data about data or information about information. Metadata can serve many different uses, including the description of the contents of a resource, explanation of how a data item is to be interpreted (for example whether it is expressed in English or metric units), statistical measurement or subjective evaluation of the quality of information, or administrative functions, such as managing, maintaining or locating resources.

A simple example of descriptive metadata is an entry in a library card catalog, which provides information about the author, publisher, publication date, format, etc. of a book or other object. Librarians tend to use metadata for this kind of descriptive information, whereas records managers and archivists also use it to describe the activity or business process context in which records were produced, for evidentiary purposes.

Discovery metadata should help users find relevant information by providing useful descriptive or categorical information. Certain kinds of metadata may be irrelevant for this purpose; for example, administrative metadata, intended to help manage resources, may be of little use for discovery, whereas archival record-keeping metadata may be of use for discovery primarily for lawyers or others seeking to hold government accountable for its actions.

We define the term **Metadata set** to refer to a specific set of metadata elements For example, the Dublin Core standard specifies a particular metadata set,, consisting of elements Title, Creator, Publisher, Date, etc.

Navigation in the online context implies moving through some logical structure, such as a table of contents or a menu or list of links that is structured to represent a hierarchy or network of information. Navigation is typically distinguished from searching (defined below), though this distinction can be blurred. Navigation is sometimes also confused with BROWSING, which implies a less organized wandering through a collection of information.

For a definition of **ontology** see the discussion of **taxonomy**.

The term **record** as used in the records management context means something that documents or provides evidence of an official action or activity performed by some organization, for example, the granting of a permit by some government agency. However, the term is also used more generally in this context to mean any information

resource, such as a document or database. (Note that the meaning of this word in the database context is quite different: there it means a particular collection of data values, such as a row in a relational table. We do not use the term in that sense in this report.)

The term **resource** is used in various, sometimes conflicting ways by the library, archives, records management, and web communities, which have been the primary users of metadata in this context. Most of these communities consider a resource to be a distinct, static object (such as a book or report) rather than something more amorphous and dynamic, such as a database or an interactive web-page. However, we use the term in its most general sense in this report, to include services and facilities as well as individual objects and both online entities and offline objects such as printed reports or CD-ROMs.

Searching in the online environment means using a computer to search the contents of resources for combinations of words or phrases. Note that if keywords in a resource represent categories corresponding to some navigational structure, then searching may be equivalent to navigation.

A **semantic net** is a mechanism for representing knowledge as a collection of nodes connected by a network of relationships. A semantic net can represent entities such as objects, concepts, events, and processes, as well as facts or relationship among such entities, such as the fact that a steering wheel is part of a car or the relationship between a grandfather and his grandchildren. Because semantic nets are very general and powerful, they can be used to represent essentially any knowledge or information.

A **standard**, as used in this report, is an accepted prescription of some sort, which specifies such things as how to perform some process, how to format or organize some document or how to choose allowable values for certain kinds of information. For example, a standard for a given type of report might specify what sections such reports should have and what those sections should contain. Similarly, a standard for how to specify a date might specify the order in which the day, month and year of the date should appear and whether each would be encoded by letters or digits (e.g., "June 3, 2004" vs. "03-jun-04" vs. ("04-06-03"). A metadata standard lists a set of metadata "elements" (or fields) that should be specified and describes the possible values of each of these.

For a definition of **syntax encoding** see the discussion of **encoding**.

Various techniques are used to manage and structure large collections of related terms, such as keywords. A simple hierarchy, such as the grouping of plants or animals, is often referred to as a **taxonomy**. Each term has exactly one location in a taxonomy, which reduces confusion but does not reflect the fact that many terms have multiple meanings. A **thesaurus**, on the other hand, like the library reference book from which it is derived, allows synonyms and other related terms to refer to each other by means of a limited number of built-in relationships and allows a given term to have different meanings in different contexts. An **ontology** is the most general terminological concept: it can have any structure, can allow individual terms to be present in any number of different contexts or categories, and can represent arbitrary relationships among terms and their contexts. Various tools for building semantic nets can be used to build ontologies; for example, Topic Maps allow terms and concepts to be represented as "topics" that can be linked in arbitrary ways.

Research Approach

We began with the assumption that a suitable nationwide metadata standard can achieve the desired effect of improving discovery, access, and (thereby) the transparency of government. Although there is little definitive evidence to support this assumption, it is logical and widely accepted. Starting from this assumption, we conducted a cursory survey of current metadata practices and problems within a small number of government information provider organizations in the Netherlands.

The project's scope did not permit conducting a survey of user needs, nor were we able to identify any large-scale relevant studies of this kind that had been performed in the Netherlands. On the one hand, the lack of any such input presented a significant risk to the project, but even the best study of user needs would be a somewhat futile attempt to track a moving target. If potential users of cellular phones had been asked ten years ago how they envisioned using such devices, it is highly unlikely that they would have predicted their current use patterns with any accuracy. Similarly, if asked twenty years ago, the vast majority of current computer users would no doubt have predicted little benefit from using such machines. Any survey of user needs would perhaps be valid for six months to a year, making it a questionable basis for long-term planning.

In the absence of empirical input about user needs, we felt it was crucial to get some idea of the plans and current practices of providers of online government information in the Netherlands. Since we were unable to field a large-scale survey, we attempted to get relevant information from several sources. First, we created an online survey that was advertised in an online newsletter to over 1200 government information sites. In addition, 30 of the ZBOs (Independent Government Agencies) were approached by letter. Finally, an e-mail message soliciting participation was sent to an additional list of approximately 165 contacts, obtained from Advies Overheid.nl through the Overheid.nl Monitor of Advies Overheid.nl.

The intent of this survey (discussed in further detail in Appendix A), was to obtain a broad view of the kinds of information that Dutch government websites are providing and whether they are using metadata to enhance discovery and access of this information. Unfortunately, for various reasons, response was far less than we had hoped. Moreover, we believe that the results are skewed by a strong self-selection factor, since only those sites that understand and actively employ metadata responded. (Evidence for this effect came from the fact that we received a number of phone calls and e-mail messages from recipients of our survey saying that they had not filled it in because they had little or no experience with metadata.) Despite these problems, we obtained answers to 38 surveys.

In parallel with our online survey, we conducted 29 in-depth interviews (mostly in person, but some by telephone) with key personnel at provider organizations that we had reason to think were already using metadata in significant ways or could help us to identify such users. These interviews were structured around a questionnaire (shown in Appendix B) whose intent was to identify major problems and issues that these organizations had encountered or expected to encounter in trying to implement a metadata scheme to improve discovery and access. These interviews (which included meetings with key implementers of the U.K.'s eGMS metadata standard) were also skewed, since we chose organizations that we knew had experience with metadata, which gave disproportionate weight to central government agencies. Nevertheless, these interviews were most enlightening and revealed many important issues, as discussed below under Phase III and Conclusions.

Finally, we augmented our online survey and interviews with considerable online research, digesting whatever we could find about national metadata standards and efforts in other countries. In the process, we discovered a number of other handbooks, similar to the one we were producing, which provided useful input. We combined all of this input to produce recommendations for a national Dutch metadata standard and an implementation strategy to encourage its adoption and use.

This still left the question of what strategy to employ in designing such a standard. For example, we could have simply selected the most widely accepted standard or the simplest to use, the easiest to implement, the most complete, the most powerful, or the most flexible.

After analyzing many of the national standards that have been accepted or proposed by other countries (as discussed below under Phase II), we concluded that:

- Most of them combine or confuse the access and records management domains.

- Most of them are relatively inflexible and difficult to extend significantly.

- Few of them address issues of dynamic online information.

- Few of them address the use of metadata by search engines and other tools.

- Few of them address implementation issues.

In combination with the fact that Advies Overheid's initial assumption for this study was that Dublin Core would form the basis of the Dutch national discovery metadata standard, we therefore employed a design strategy based on the following principles:

- Use the Dublin Core standard as a starting point

- Focus on metadata for discovery and access, while taking into account the need to integrate the access domain with the records management domain.

- Keep the standard as simple as possible without limiting its future ability to serve the needs of diverse communities to describe the full range of online information resources and services.

- Consider implementation issues, including engaging relevant communities, motivating adoption, managing metadata in metadata systems, and ensuring that metadata can be used by finding-tools and systems.

We used these principles to develop the proposed NL-meta standard using as our model several other national metadata standards, notably the eGMS from the U.K. (as suggested by Advies Overheid.nl) and the work of the CEN (European Committee for Standardization) Workshop's eGovernment Metadata Application Profile v. 1.0 (CEN Workshop Agreement, 2003). The result retains most of the simplicity of the Dublin Core standard but embeds it in a simple but general "optionally-structured" model that provides far more flexibility and power than the DC or any of its derivatives.

The remainder of this report presents the results of our analysis, organized according to the three major phases in which the project was conducted. The first phase examined the needs of government information providers, to enable us to understand their concerns and what a metadata standard should offer them. The second phase examined the Dublin Core and a number of other national metadata standards, to enable us to understand their capabilities and limitations and to propose a discovery metadata standard for the Netherlands. The third phase explored implementation issues that had emerged in the first two phases. Our conclusions and recommendations are drawn from the results of all three of these phases.

The next section discusses Phase I, including our analysis of the needs of government data producers in the Netherlands and our proposed strategy for enabling a national standard to be flexible enough to adapt to the needs of different user groups.

Section 5 (Phase II) discusses the Dublin Core standard and some of its alternatives, focusing on the important relationship between the access domain and the digital records management domain and our model for how the former should ideally derive much of its metadata from the latter. This section also discusses the motivation for providing structured metadata, leading to our proposed "optionally-structured" metadata approach and the specific set of metadata elements and refinements that we propose for the Dutch national metadata standard.

Section 6 (Phase III) discusses the primary implementation issues that we feel are important in establishing a useful and successful national metadata standard.

Section 7 identifies our major findings. We first recommend general steps that should be taken to implement and utilize a national metadata standard. We then analyze the risks involved in this endeavor and suggest ways of mitigating them. Finally, we recommend some specific steps that Advies Overheid.nl can take to advance the adoption of the proposed standard.

SECTION 4 Phase I: Analysis of Dutch government data and metadata needs

In Phase I, we analyzed the needs of government data producers in the Netherlands, in the context of relevant Dutch laws, regulations, policies, programs, projects and initiatives, as well as relevant developments in the EU. This context is discussed in Section 4.1. We then examined a number of issues that arise within information producer organizations themselves, as discussed in Section 4.2 below. In particular, we studied the different needs that different branches, levels, or types of government departments or agencies might have for metadata, and we considered alternative ways of making the national standard flexible enough to adapt to these needs. For example, different groups of users may require different encoding schemes or controlled vocabularies for certain metadata elements (e.g., for temporal or spatial coverage, which specifies the time period or geographical area covered by a resource). Or they may find it useful to supply automatic default values for different elements, such as expiration dates for periodic publications. Similarly, different groups may rely on different sets of keywords or search terms, which may need to be represented as distinct terminologies. We concluded that the best way to allow for all such variations in the use of a national metadata standard is to encourage the formation of distinct Communities of Interest, each of which can develop its own guidelines for using the standard and can support its own specialized terminology.

One of the more surprising results of our survey was that most of the government information provider organizations we contacted are relatively satisfied with their current discovery capabilities and see no reason to change. This has significant implications for the promotion and implementation of a nationwide discovery metadata standard and implies that Advies Overheid.nl must undertake a serious marketing effort if such a standard is to be adopted. (Of course, it is ultimately user satisfaction that matters, but in the absence of a comprehensive empirical study of user satisfaction, information providers have only limited user feedback and their own intuition to help them evaluate the effectiveness of the discovery process.) Although we have no magical solution to this lack of perceived need among information providers, we suggest various strategies in this Report (e.g., in Sections 4.2.2 and 6.1.1) to attempt to address this impediment to the success of a national discovery metadata standard for the Netherlands.

4.1   The Dutch and EU contexts

This section discusses Dutch projects and initiatives that are relevant to metadata, the Dutch legislative and regulatory context, and the EU regulatory context.

4.1.1   Dutch projects related to metadata

Though nearly all municipalities in the Netherlands now have some sort of website, the quality, content and structure of these sites varies greatly. At the other end of the spectrum, the central government level also exhibits significant variance, and there is no central authority that can enforce standards and policies.

The ministry of internal affairs (BZK) is the responsible coordinating ministry for eGovernment activities of the Dutch government. It has developed the Overheid.nl central portal for all government web publications, services and organizations. This portal has grown since its launch and is the centerpiece of BZK's digitization campaign. The RVD's Activity Index initiative (discussed below) is of particular relevance to Advies Overheid.nl's nationwide metadata standard effort.

eGovernment efforts relevant to the metadata initiative include:

- Programma andere overheid (Transfiguring government program)

- Notitie 'Op weg naar de Elektronische Overheid' (Towards an electronic government)

- Nota 'Naar optimale beschikbaarheid overheidsinformatie' (2000)   (Towards optimal availability of governmental information)

- Overheid.nl

- Activiteiten Index (Activity Index)

- Bedrijvenloket (Business counter)

- Webwijzer award (Award for the best Dutch governmental website)

- Produkten catalogus (for municipalities) (Product catalogue)

- Metadata pilot projects (Advies Overheid.nl)

- Quality of web interfaces project (TNO/Cinnamon)

- Search engine project (ZENC)

- The InterLAB working group of ministries (on records management metadata)

The urgency of eGovernment policy has dropped in the last few years, but it has recently received new impetus through the appointment of Mr. De Graaf, as the Netherlands' new minister for governmental reform. Also, at the central government level, patterns of cooperation are developing between civil servants that allow for a bottom-up approach to

choosing standards and developing joint projects and policies. Although these activities are not formally coordinated, these new informal structures allow for a reasonably effective exchange of information and ensure that most of the officials involved have a broad understanding of the various projects that are currently running.

International benchmarks (EU/SIBIS (Statistical Indicators Benchmarking the Information Society), OECD, WEF (World Economic Forum), Accenture, eEnvoy/Booz Allen) have also increased pressure on Dutch policymakers, as the Netherlands is lagging behind other EU member states in creating a favorable eEnvironment. However, these precedents seem to have the greatest impact on the departments that are already convinced of the need to increase the eGovernment effort; they do not generate a similar response at lower government levels and departments that have been lagging behind.

Although the above projects and initiatives vary in their relevance to the introduction of a national discovery metadata standard in the Netherlands, the multiplicity of these efforts suggests that the time is ripe for a standard of this sort (as well as for standards in the records management domain). A nationwide discovery metadata standard could, if successful, greatly improve the coordination and interoperability of many of these efforts and of the information resources they produce. In addition, the current project aimed at developing an improved search engine for Overheid.nl should complement this metadata standard effort by helping to make discovery metadata available to finding-tools.

4.1.2 The Dutch regulatory context

A number of laws and regulations in the Netherlands are also potentially relevant to the introduction of a discovery metadata standard.

The Wet Kenbaarheid Publiekrechtelijke Beperkingen (Law on provision of information on public restrictions concerning real estate) draws up a registration model and methodology for public restrictions (65 restrictions are listed to be covered, e.g., permits, destination decrees, etc.) on the use of real estate. This information used to be scattered around different government bodies and was difficult for users to retrieve. The law obliges all public bodies involved in setting restrictions on the use of real estate (municipalities, water boards, etc.) to register the information in a uniform way and to guarantee that the user can access all the information through a single entry point. Government is held liable for not providing the right information or not providing it within four days of the decision to apply a restriction.

 To make this happen, all government organizations involved need to link their restriction registration systems with the central real estate register (kadaster). The expectation is that the system can be extended to other areas of public policy as well. The law explicitly states that its implementation should also lead to organizational changes and changes in the way governments interact among themselves and with citizens. The law should provide more transparency, better service and a legal base for citizens to hold government liable for failing to act. It is expected that the law's effect--once it is in force--will create a new momentum for municipal and provincial governments, and water boards to link their information systems in order to develop shared services. The law is therefore likely to be a

driver for these organizations to consider the use of metadata and to adopt a metadata standard for all such government information.

The Law on Public Electronic Traffic (Wet Elektronisch Bestuurlijk Verkeer, WEBV) has been drafted to determine the status of electronic documents, an issue which has long been a barrier to introducing more information online. This draft law is an extension of the framework law 'Algemene Wet Bestuursrecht' (AWB - General Public Law). WEBV provides a regulatory framework for all electronic interaction and transaction between government and citizens. It is expected to go into effect this year. In the meantime government and municipalities are invited to experiment with implementation of the provisions of the WEBV. The ministry of internal affairs has developed a handbook and guidelines for its implementation.

The Personal Data Protection Act (Wet Bescherming Persoonsgegevens, WBP) regulates all use of personal data. The use of personal data plays an important role in many areas, such as public administration, health care, social security, telecommunication, financial services and direct marketing. Personal data must be protected in order to safeguard the rights and freedom of citizens. The WBP applies to the fully or partly automated processing of personal data, and the non-automated processing of personal data entered in a file or intended to be entered therein. Not all kinds of personal data are covered by this act, and there are some specified exceptions. Only data used to identify persons is protected, and restrictions are set on processing personal data. WBP identities four risk-levels that determine what kind of protective measures are required. Personal data under level I to III are under protection of the act, whereas data under risk-level 0 (public level) are not under its protection. Generally accepted personal data, such as appears in phone books, public sites, etc., are classified under level 0.

The WBP sets conditions for the lawful processing of personal data. It has a high standing in the Netherlands and is strictly followed by the Dutch government and commercial companies. The execution of the WBP is controlled by the Dutch Data Protection Authority (DPA). This is an independent government agency (ZBO) in charge of checking the use of personal data and guaranteeing the protection of citizens' privacy now and in the future. The Dutch DPA advises the government, tests codes of conduct, studies technological developments, provides information, handles complaints, evaluates processing of personal data and, if necessary, takes enforcement action.

At the moment most government websites do not contain data under protection of the WBP, so the act has limited impact on Advies Overheid.nl's effort to establish a Dutch nationwide metadata standard. However, WBP may raise some concerns when the metadata element "author" (or "creator") is used to describe the civil servant who has actually written the content of an information resource. In the Dutch government, the ministers are responsible for what is written by civil servants. If the names of civil servants were listed in a metadata field, users would be able to trace documents to the civil servant who wrote them. This would also facilitate combining personal data, for example to infer the responsibilities or political leanings of specific authors, as represented in various publications to which they have contributed.

Though WBP has only limited relevance for current data on government web sites, there could be issues when linking databases containing personal data, e.g., to provide shared

services, as the WBP sets conditions for linking personal data with other data. Although this again goes beyond the strict context of establishing a metadata standard, it is a concern that needs to be noted and addressed when fully exploiting the benefits of metadata for improved discovery.

In this context it is worth noting that the government is in the process of overhauling the way it collects personal data. The program Streamlining Key Data (Stroomlijning Basisgegevens) was launched based on the principle that the government should not ask what is already known. Currently each government agency is, in principle, responsible for the independent collection of the data it needs to perform its duties. In the future the data required by a number of agencies would, in principle, be collected once--in one place and at one time--and then be made available to the relevant agencies. A large number of ministries, the Association of Netherlands Municipalities and the Dutch DPA are engaged in the joint implementation of the Streamlining Key Data Program. The Dutch DPA ensures compliance of the program with the Personal Data Protection Act. The Streamlining Key Data Program will be used as a back office system. The parties involved are required to protect the data under their control: This should automatically protect front office information.

As long as there is no link between personal data and other data, there is no protection under the WBP. It is important to note that private and public organizations have an independent responsibility for protecting personal data. In developing online transactional services they must design their systems in a way that ensures that users will only have access to the information for which they have been granted permission.

The overall impact of these laws and regulations on discovery metadata appears to be less than we expected, with the possible exception of the personal data privacy issues noted above. Nevertheless, they imply the increasing importance of online information in the Netherlands, which in turn implies the need for improved discovery and access to such information.

4.1.3 The EU context

The ambitions of the EU to promote eGovernment and eCommerce are ahead of most national developments, including those in the Netherlands. This lack of compliance with EU guidelines may be damaging to the national policy, but this has not led to significant political pressure at the Dutch national level. At the municipal level these European objectives are largely unknown.

eEurope is the central program to stimulate the eEnvironment in the EU. It is part and parcel of the Lisbon strategy and works through a series of instruments: policy measures, coordination of member states' policies, establishing best practice, benchmarking and setting policy targets. Targets include:

- Under the eEurope 2002 Action Plan, Member States agreed to provide all basic services online by the end of 2002.

- By the end of 2004, Member States should have ensured that basic public services are interactive, where relevant, accessible for all, and exploit both the potential of broadband networks and of multi-platform access.

- By the end of 2005, Member States should carry out a significant part of public procurement electronically.

- Under the eEurope 2005 Action Plan, by 2005 Europe should have:

  - modern online public services

  - e-government

  - e-learning services

  - e-health services

  - a dynamic e-business environment

  and, as an enabler for these:

  - widespread availability of broadband access at competitive prices

  - a secure information infrastructure

Governments pledge to provide interactive public services, accessible to all, and offered on multiple platforms. If this is to become a reality, it will require improved discovery metadata to enable citizens and others to find and access these services, both within the Netherlands and by remote access.

Under the eEurope action plan with support of the IDA program (Interchange of Data between Administrations), the EIF (European Interoperability Framework) was developed. The Commission and member states will draft a list of public services for which interactivity and interoperability are desirable. The EIF is concerned with three aspects of interoperability, i.e., organizational, semantic, and technical. EIF is not prescriptive; it attempts to provide a framework in which governments can develop largely compatible policies that should support a greater level of pan-European interoperability. The EIF principles are shared by most member states, and although the EIF itself does not assert a great deal of actual or political pressure on national governments to comply, any EU member state's metadata standard is expected to take the principles of the EIF into account.

The EU Directive on Services in the Internal Market (COM (2004) 2 def 2 / SEC (2004)21 could serve as a good "stick" to promote the urgency of adopting a metadata standard. The directive is a follow-up to the Lissabon Council conclusions (2000) and the Commission's Internal Market Strategies (December 2000 and May 2003). It is a framework directive, which aims at reducing barriers for cross-border service within the internal market. The Directive specifically requires member states to provide a one-stop shop for EU service providers, where they must be able to consult and obtain all relevant information and permits required for conducting business by 31 December 2008. This service and accompanying procedures should be provided in electronic form.

These EU activities provide additional motivation for the creation of improved discovery metadata. For example, they broaden the user base of those who may need to find and

access services to include other Europeans. This has significant multi-lingual implications for both online resources and the discovery metadata that describe them.

4.2 Intrinsic metadata issues

This section discusses metadata issues that emerged from our survey as affecting information provider organizations.

4.2.1 Tailoring metadata to different needs

One of the key questions that we set out to answer in this phase of the project was the extent to which different branches, levels, or types of government departments or agencies have different needs for metadata. Advies Overheid.nl was particularly interested in whether agencies dealing with different types or realms of government information (e.g., law and regulations, licenses, European regulations, official publications etc.) might have different needs. An agency dealing with financial information might have different needs from one dealing with laws and regulations or one dealing with scientific, historical, or environmental information. For example, an agency working in a given realm might require specialized metadata fields (elements) to describe its information in terms of geographical/cartological coordinates or to provide instrument calibration metadata for environmental or climatological measurement data. Similarly, certain metadata elements might be more important for some realms than for others; for example, cadastry data might require geographical coverage metadata, whereas such coverage information might be optional or irrelevant in certain other realms. In addition, certain metadata elements might have easily specified default values in certain realms (such as the name of a publishing organization, temporal coverage for tax information pertaining to a given tax year, or expiration dates for periodical reports), whereas others may not.

Despite the fact that different realms may have specialized metadata needs, the majority of the metadata elements that are relevant for discovery and access are common across nearly all realms, making it unattractive to add specialized metadata to a national standard. The need for consistency and interoperability across realms to facilitate discovery argues that a national metadata standard should not be burdened with a large number of optional fields or variant forms, most of which would be meaningful to a limited number of users. Indeed, all of the existing national metadata standards attempt to minimize variant or optional fields, though some are more successful at this than others.

Similarly, it appears to be inappropriate to define default element values in the standard. The choice of which elements should have default values will be different for different organizations dealing with different subject matter and having different policies and procedures. For example, a tax organization that is preparing information concerning the current tax year may allow the temporal coverage metadata for most of the resources it is creating to default to the current year, whereas an agency that publishes quarterly reports may default the temporal coverage for such reports to the relevant 3-month period. It seems best to encourage organizations to define relevant defaults for themselves on a case-

by-case basis and to encourage metadata system developers to provide features that enable users to automate the provision of default values wherever needed.

4.2.2 Communities of Interest, supported by a national clearinghouse

One approach that we identified for enabling agencies in different realms to develop and use different metadata is based on the approach that is currently being used in the RDF/XML (Resource Description Framework/eXtensible Markup Language) community to deal with the development of multiple ontologies and namespaces. This involves the notion of Communities of Interest (COIs), each of which takes responsibility for its own terminology and metadata. The concept of a COI is a very general one: a COI can correspond to a level of government (municipal, provincial, central) or a realm, but it can also cut across such hierarchical divisions to represent any collection of organizations or individuals that share any common concern or interest (for example, privacy or gender issues). Various COIs already exist among Dutch government organizations, notably those surrounding GFO efforts, such as GFO Zaken.

A further use of COIs is in the modular development of ontologies. It is far more realistic to allow each COI to develop an ontology for its area of interest than to develop a single, top-down ontology covering all such areas. If ontologies are developed using suitable approaches (such as Topic Maps or similar semantic net tools), it may be possible to treat the ontology developed by each COI as a "module" that can be merged with other such modular ontologies, thereby creating an overall ontology from the bottom up. This would require some form of coordination among COIs, as suggested below.

Among their other roles, COIs should keep track of, and possibly participate in, the work of appropriate professional groups in developing and evaluating standards, encoding schemes, and CVs that are relevant to the concerns of the COI.

Finally, the COI mechanism may help counteract the perceived resistance of many organizations to introduce new discovery metadata procedures, since it replaces the direct intervention of central government in such matters by an emphasis on cooperation among organizations that share common concerns and missions. It is possible that organizations that see no reason to change their online practices may be convinced of the need to do so if approached by other organizations with which they need to interact.

In order to reap the benefits of COIs, each organization will have to coordinate its metadata decisions with other organizations in the appropriate COIs. Furthermore, each COI will require some minimal administration and coordination with other COIs, and this work will in most cases have to be performed by the COI's member organizations.

If all of the COIs were logically disjoint and did not interact with each other,aj and if users never needed to combine or compare information from across multiple COIs, then each COI could develop its own metadata and terminology independently, without fear of

overlap or inconsistency. Unfortunately, this is far from the case, so adopting a COI approach requires some centralized coordination. The online stakeholders' working group proposed in Section 6.1.2 should be useful in promoting the idea of COIs, but in addition, we recommend the establishment of a national COI clearinghouse, which would provide a means for groups to determine whether a relevant COI already exists and, if not, to register a new one so that others become aware of its existence. The clearinghouse would perform a similar function for encoding schemes, CVs, taxonomies, thesauri, and ontologies so that, for example, an information provider that needs to choose or create a CV for some specific purpose (such as specifying geographic coverage) can determine whether a suitable CV is already in use by some relevant COI and, if not, which COI might help the organization choose or define such a CV. Moreover, if COIs develop their own, independent ontologies, the clearinghouse could coordinate the development and merging of these into an overall ontology. Finally, the clearinghouse would help mediate and resolve conflicts among COIs; however, if it fulfills its primary coordination function of minimizing the overlap in the scope of distinct COIs, then such conflicts should be minimized, and it should be possible for each COIs to make its own decisions.

Creating a national COI clearinghouse obviously requires resources, as well as care in placing it within an appropriate organization in central government. These and other implementation issues involved in establishing a national COI clearinghouse are discussed in Section 6.1.4 below. Since the role of the clearinghouse would be primarily one of coordination, it should not be necessary for it to be large or powerful. Nevertheless, the creation of such a clearinghouse is essential to ensure that the COI mechanism fulfills its potential rather than producing confusion.

The COI approach puts the burden for specialized extensions to the metadata standard in the right place: the organizations in each COI take responsibility for their own extensions, whereas the standard itself need not be complicated by numerous variants and additions. For the COI approach to succeed in the Netherlands, it must be supported by the proposed online stakeholders' working group as well as by a national COI clearinghouse.

4.2.3   Encoding schemes and controlled vocabularies

Advies Overheid.nl specifically requested that we explore the question of where encoding schemes and controlled vocabularies should be chosen or defined, and Appendix A of the Handbook lists relevant encoding schemes for most of the elements in the proposed NL-meta standard. In addition, the choice of encoding schemes appears to be an area where COIs may have their greatest impact.

Following the terminology of the Dublin Core Metadata Initiative (DCMI), an encoding scheme may be either a "syntax encoding" or a "vocabulary" though we prefer the term "controlled vocabulary" (CV) for the latter. These specify (respectively) the format and allowed values of particular fields, such as dates, object identifiers, abbreviations for the names of cities or countries, etc. In many cases, a metadata element is of limited use for discovery without an associated encoding scheme or CV that indicates the form or range of its allowed values.

A syntax encoding is a specific format for representing a given type of information. For example, 12-jun-04, June 12, 2004, and 2453168.5 (Julian date) are three ways of encoding the same day, whereas http://overheid.nl/home/sitemap is a way of encoding the name of a web-page. A controlled vocabulary, on the other hand, is normally a closed list of terms or identifiers, such as abbreviations for the names of cities or countries, or the names of alternative cartographic systems.

Many common metadata elements have well-accepted standard encoding schemes and CVs, and these are referred to in most standards as "best practice" for the given element. Unfortunately, in many cases, several competing standards exist, such as 12-jun-04, 12-06-04, 06-12-04, 04-06-12, June 12, 2004, etc. or the alternative, language-dependent sets of abbreviations for country names. Still, it is not difficult to choose appropriate standards for most common metadata elements.

For more esoteric or specialized metadata elements, CVs can be adopted or developed by relevant COIs. This appears to work best when a given CV is chosen or developed at the lowest level possible, i.e., by the smallest COI that includes all (or most) users of that CV. This requires some centralized coordination, as we recommend below, to enable a COI to determine what appropriate CVs may already exist for a given purpose. Similarly, central coordination is required to enable organizations to determine whether an appropriate COI for the creation of a new CV already exists--and if it does not, to create one.

It is important to retain the discovery perspective when choosing encoding schemes and CVs. The purpose of controlling the values of a discovery metadata element is to make it easier for users to find information by searching for a known set of values. Encoding schemes or CVs that are overly complex may defeat this purpose, as discussed below, unless finding-tools provide powerful interfaces to help users construct queries utilizing these schemes.

The development and use of CVs in most metadata elements is relatively straightforward, but two cases are more problematic. One of these is the choice of a geospatial encoding scheme for the Coverage element in NL-meta, and the other is the encoding of Subject keywords.

4.2.3.1 Geospatial encoding schemes

When feasible, best practice is to use names for geospatial areas rather than coordinates. Such names should be drawn from an appropriate CV, for example, a list of city or jurisdiction names. Another alternative is to use geocodes, such as postal codes, which can also be considered a CV. The advantage of using names or geocodes is that they are (at least ideally) easier to interpret than lists of coordinates describing geographical areas. On the other hand, if such a name or code is abbreviated, its meaning may not be obvious, and if the definition of the entity it names is allowed to change over time (for example, if the boundaries of a city or jurisdiction can be redefined), then the meaning of the name or code must be interpreted in the proper temporal context, whereas coordinates such as latitude and longitude generally have more permanent meanings.

In any case, no name may exist for a particular geospatial region of interest, which may leave no alternative but to describe the region in terms of coordinates. Several

comprehensive approaches have been developed to describe arbitrary geospatial regions of this kind, notably the U.S. Federal Geographic Data Committee standard FGDC-STD-001-1998 and the more recent ISO 19115. Each of these constitutes an entire metadata scheme in its own right, rather than an encoding scheme for use within a single metadata element, such as Coverage, but they are currently the most widely accepted methods of specifying complex geospatial information. Using either of these schemes to supply a value for the NL-meta.DC+:Coverage element could be done by setting the value of the Coverage element to a reference (for example a URI or DOI) denoting an FGDC or ISO metadata set describing the desired region. Alternatively, if the NL-meta.Extended structured option (as discussed below in Section 5.4) were employed, the FGDC/ISO metadata set could be made a sub-component of the NL-meta.Extended metadata set.

A simpler approach, supported by the Dublin Core Metadata Initiative (DCMI), is the DCMI-Point and related DCMI-Box schemes, which encode geographic coordinates (such as latitude, longitude and elevation) in a text string that can then be used as the value of a Coverage element.

These two extremes, the FGDC/ISO approach and the DCMI-Point/Box approach, currently appear to offer the best alternatives for complex geospatial encoding. The choice between them should be determined by each organization or COI on the basis of its specific needs for geospatial representation. However, using any complex geospatial encoding scheme of this sort may defeat the purpose of discovery metadata, since few users may have the sophistication to utilize such a scheme when searching for information. For example, it is unlikely that a user searching for information concerning some geographical area will be able to supply values that match the FGDC or ISO 19115 metadata (or even a simple set of DCMI-Point/Box coordinates) describing this area. In order to make such geospatial metadata useful for discovery, it may be necessary to construct powerful finding-tool interfaces that support geographic browsing and that match regions selected graphically by a user against regions described by geospatial metadata.

4.2.3.2 Encoding schemes for Subject keywords

The use of keywords in a Subject element is of particular significance. Searching for keywords is currently one of the most common discovery techniques, employed by World Wide Web search engines such as Google. However, these free-text search engines simply search for words in the content of documents or web pages, without knowing which of these words are intended to be keywords. In order for a search engine to find an information resource on the basis of keywords in a Subject metadata element associated with that resource, those keywords must be made visible to the search engine and must be identified as keywords. This is an aspect of the use of discovery metadata that is often ignored in metadata standards, most of which discuss only the metadata elements themselves. Search engines (or, more generally, finding-tools) may have to be modified to enable them to find and recognize metadata elements such as Subject; in particular, if metadata elements are not embedded in online resources but are instead maintained in separate databases, then finding-tools must be modified so that they can access these

databases. We discuss some aspects of this problem elsewhere in this report, but here we simply note that good keywords can greatly improve discovery, assuming that they are made accessible to finding-tools. Yet the generation and choice of appropriate keywords by metadata creators and the use of those keywords by users require considerable effort.

Various approaches can be used to manage and structure large collections of related terms, such as keywords. A simple hierarchy, such as the grouping of plants or animals, is often referred to as a taxonomy. A taxonomy places terms under higher-level terms that include them. A taxonomy requires each term to have exactly one location in the structure, which reduces confusion but does not reflect the fact that many terms have multiple meanings. A thesaurus, on the other hand, like the library reference book from which it is derived, allows synonyms and other related terms to reference each other by means of a limited number of built-in relationships and allows a given term to have different meanings in different contexts. A thesaurus subsumes a taxonomy, since a taxonomy can be represented by a simple thesaurus in which there are no synonyms or other references. Finally, an ontology is the most general terminological concept: it can have any structure, can allow individual terms to be present in any number of different contexts or categories, and can represent arbitrary relationships among terms and their contexts. An ontology subsumes a thesaurus, since a thesaurus can easily be implemented as an ontology. Various tools for building semantic nets can be used to build ontologies; for example, Topic Maps allow terms and concepts to be represented as "topics" that can be linked in arbitrary ways.

The development and use of taxonomies, thesauri, or ontologies to manage keywords can greatly improve discovery. However, the development of such terminological systems requires coordinated effort across government organizations by appropriate COIs, with the help of the national COI clearinghouse.

4.2.4 The Activity Index

The Activity Index effort can be seen as a start at developing a taxonomy of government functions, albeit in the limited context of Dutch government policies embodied in the budget and governing agreement, within the scope of central government departments, and pertaining only to the four year mandate of the acting government. The taxonomy that is being developed by this ongoing effort may be its most important contribution to the nationwide discovery metadata endeavor. Since it uses a minimal number of mandatory metadata elements, chosen from the Dublin Core set, it does not imply the need to add any additional elements to the proposed NL-meta standard.

The Activity Index allows each organization to decide for itself whether to embed metadata in the resources that are described or to keep it separate, as discussed under embedded vs. separate metadata, below.

4.2.5 Navigating, searching, and "finding-tools"

Discussions of discovery tend to focus on searching, but it is useful to broaden this to include the concept of "navigation" as well. The distinction between these two in the

online world is somewhat blurred, but website creators often provide buttons, menus, or structured trees or networks of links, analogous to tables of contents, that allow users to navigate from section to section (or page to page) of a website. This is seen as quite distinct from allowing users to search for text inside documents or other resources. However, if searching can involve keywords that describe sections in the navigational structure, then searching becomes equivalent to navigation. Similarly, if searching allows users to traverse a structured collection of terms (such as a taxonomy or ontology), this amounts to navigating through a set of search terms, which can be thought of equally as navigation or searching.

Browsers and search engines are currently the most common tools for finding information on the web, but there are many other existing and potential tools of this sort, such as database query systems, object repositories, and semantic nets (e.g., Topic Maps). In order to include all such tools in one term, we refer to them here as "finding-tools" (which is a generalization of the more restricted notion of a "finding aid"). As noted repeatedly in this report, if the use of discovery metadata is to fulfill its promise, metadata must be made accessible to finding-tools, to help users find what they seek.

4.2.6 Longevity

All digital artifacts--even simple text files--consist of streams of bits that must be interpreted by software (i.e., "rendered") in order to be made human-perceptible. All digital information is therefore vulnerable to becoming inaccessible or unusable as the programs that render it or the computers on which those programs run become obsolete.

In order to ensure that online information remains accessible, readable, usable, and meaningful to users in the future, it is necessary to create preservation metadata to describe the process necessary to render the information in a form that human users can see, understand, and utilize.

Most of the literature on preserving digital information assumes that it will "migrate" into new formats over time, thereby remaining readable in these future formats. Since migration involves conversion from one format to another, it inevitably loses some aspects of the original, and since it must be performed repeatedly, such conversion losses accumulate over time. Still, if migration is employed, preservation metadata must indicate what format (and version of that format) a resource is in, and each time a resource is converted, preservation metadata should document the conversion process, to enable future users to evaluate the quality of the migration effort.

A more formal approach to preservation involves the logical description of the form and content of a resource, so that it can be rendered from this description in the future. In this case, the preservation metadata must constitute the formal description (or a pointer to it).

A final approach to preservation involves the use of emulation of hardware platforms to enable the original saved rendering software for a resource to be run (under emulation) on future computers. In this case, the preservation metadata should consist of pointers to a suitable emulator and suitable versions of the original rendering software for the resource in question.

Assuming that digital information can be kept readable, it must also retain its meaning over time. This is more problematic, since it depends on proper semantic conversion or explanation, as the meanings of words and even concepts evolve over time. In the absence of a complete, formal semantic description facility (which is beyond the current state of the art), this must rely on human interpretation.

The longevity of metadata (as opposed to that of the resources described by the metadata) is a somewhat simpler issue. Since metadata values are not highly application-dependent, they are not as vulnerable to loss as more general online information. Metadata represented in a particular relational, object-oriented, or semantic net form may be application-dependent to the extent that proprietary or unique aspects of these representations are utilized. However, the need to use metadata on an ongoing, daily basis implies that they will naturally migrate into new metadata systems as needed over time, thereby hopefully avoiding loss.

A more troubling issue concerns the revision of metadata or of the metadata fields (elements) themselves, i.e., evolution of the metadata standard. If metadata are kept separate from the online resources they describe, this is not a major problem, although it may require modifying a metadatabase or other representation. However, if metadata are embedded in the resources they describe, then changing the metadata requires modifying these resources, which is an unappealing prospect, unless it can be fully (and safely) automated. This issue forms one of the motivations for our discussion of embedded vs. separate metadata, below.

Finally, the semantics of metadata must be preserved. For example, if the meaning of search terms evolves over time, it may no longer be appropriate for them to denote the same resources. This is as yet an unsolved problem, but the use of thesauri and ontologies may offer indirect linkages of this sort, which are more easily changed than direct linkages from metadata to specific resources. For example, rather than searching old documents for a new keyword, a thesaurus can be used to map the new keyword into an older one that denoted the intended meaning in older documents.

Although longevity tends to be an afterthought, it must be considered in a discovery metadata scheme if metadata are not to become obsolete or misleading in the future.

4.2.7 Semantic interoperability

Finally, although our limited investigation of user needs was unable to find much direct evidence for it, we believe that one important class of discovery involves combining or comparing information that comes from multiple government sources and sites. In particular, researchers are often faced with such tasks, for example, when gathering, aggregating, or comparing information from different geographical or political districts or looking for information on such subjects as privacy or gender, which cut across traditional government information realms. To help support such combining or comparing of information from different sites, discovery metadata must facilitate semantic interoperability, i.e., the consistent interpretation of information across different contexts. This can be aided by the development of ontologies that span these contexts, though in practice it has been difficult to develop such comprehensive ontologies across multiple

COIs. New tools for building semantic networks, such as Topic Maps, may help by enabling the merging of independent community-specific ontologies developed by different COIs.

In the end, although we could easily have spent many more months performing this phase, we believe we obtained a qualitatively valid understanding of the major issues involved in developing a national metadata standard. While future surprises are always possible, we are fairly confident that we have identified the most significant risks and implementation concerns that accompany this undertaking.

SECTION 5 **Phase II: Analysis of Dublin Core and alternatives**

Having explored contextual issues in Phase I, we proceeded to Phase II of the project, in which we analyzed the Dublin Core standard and some of its alternatives. Our focus on Dublin Core was one of the premises of the project, since Advies Overheid.nl's initial assumption was that it would form the basis of the Dutch national discovery metadata standard. Similarly, Advies Overheid.nl initially suggested that one of Dublin Core's derivatives, the eGMS from the U.K., might serve as a model for a Dutch national standard.

The results of our surveys as well as our analysis of Dublin Core, eGMS and other standards revealed the important relationship between the access domain and the records management domain, as noted above. This led us to develop a model for how the former should ideally derive much of its metadata from the latter.

In addition, the needs of information providers to describe dynamic, composite objects (such as databases and websites) as well as eServices, led us to focus on a major limitation of DC and those standards that are derived from it, namely, their lack of structured metadata, which makes it difficult for them to group different kinds of metadata together or to describe complex, structured resources. In order to circumvent this limitation without imposing a burden on users for whom a simple DC metadata scheme is sufficient, we developed what we call an "optionally-structured" approach. After discussing these issues, this section presents the specific set of metadata elements and refinements that we propose for the Dutch national metadata standard.

5.1 Dublin Core and alternative standards

The Dublin Core metadata standard consists of a small number of metadata elements each of which describes some aspect of a "resource" (such as a document or database). The DC elements are analogous to the data fields found on a card in a library card catalogue. They specify information such as the author, publisher, publication date, format, location, etc. of the resource being described. Each element can be restricted by an "encoding scheme" that specifies the format of its allowed values; for example, dates may be constrained to be in the form "YYYY-MM-DD" with 4 digits for the year, followed by 2 digits each for the month and the day. "Controlled vocabularies" can also be used to restrict the values of an element to those in a given set, such as standard abbreviations for country names (e.g., NL,

UK, US). Each element can have only a single value, but that value can be a string of text, such as a multi-word title or an abstract for a report. In addition, some elements, such as Creator, can be repeated, to represent multiple authors. The basic DC contains 15 elements, but users are free to consider only some subset of these mandatory (i.e., requiring that their values be filled in), while other elements can be made merely recommended or optional. In addition, each element can have various sub-elements (called "refinements"); for example, Date can have refinements such as Date.created and Date.issued. Users of the standard can add new refinements to each element if needed.

The DC standard has been adopted or adapted by many organizations and governments as the basis for their own standards. We examined quite a number of these national efforts, focusing on those in English, including the eGMS from the U.K., the AGLS (originally the Australian Government Locator Service, though no longer restricted to the public sector) and the related VAGLS (from Victoria) and NZGLS (from New Zealand), the Irish Public Service Metadata Standard (IPSMS), and the Canadian Treasury Board Information Management Standard (TBITS), as well as a number of other related efforts, including the CEN Workshop's eGovernment Metadata Application Profile v. 1.0 (CWA 14860, November 2003), GILS (Global Information Locator Service), the EAD (Encoded Archival Description) standard, the Warwick Framework, ISO-IEC_11179 on Metadata Registries (MDR), ISO/TS 23081 on principles of metadata for records management, the OASIS work on search interoperability and Topic Maps, and the European Interoperability Framework (EIF).

Several alternatives to DC (such as GILS and EAD) offer richer sets of metadata elements to describe resources. However, most of these alternatives are strongly oriented toward records management or archiving. The widespread acceptance of DC as the basis for national metadata standards suggests that its element set is sufficient for most discovery purposes and that the additional power of any such alternative must be carefully weighed against its complexity and the increased burden it places on implementers.

5.2   Discovery vs. records management metadata

As we studied these standards, it became apparent that many of them combine metadata for discovery with metadata for records management and/or archival purposes. Records management metadata include any information that is used to manage information assets or to maintain accountability for their origin, purpose, use, and stewardship. For those standards that arise out of records management or archival concerns, combining discovery and records management metadata is entirely appropriate; however, at least some of the standards appear to suffer from a failure to properly conceptualize or analyze the distinction between the access domain and the records management domain.

The Dublin Core arose from the library domain and is understandably focused on describing distinct, concrete resources, such as books and their digital equivalents. DC retains a distinct "cataloguing" flavor, which combines the access and management domains. It is simple and relatively easy to implement, and it is essentially "flat" (i.e., unstructured). That is, all of its elements (Title, Creator, Publisher, Date, etc.) are on the same level and cannot be grouped together into more complex collections of elements, sub-

elements, sub-sub-elements, etc. The element "refinements" (Date.Created, Date.Valid, etc.) in DC do provide a partial second level of structure, but these refinements do not provide a general-purpose, multi-level structuring facility.

Dublin Core and all of its derivatives are strongly oriented toward describing distinct, concrete resources like reports, whereas online information of the kind that citizens and companies often seek may take numerous, complex forms that may not be easily described in terms of individual resources. Nevertheless, the simple, flat Dublin Core approach provides a good foundation for discovery metadata.

When a true digital records management function is implemented in an organization, it serves as the owner and manager of metadata describing the organization's information resources. (This function overlaps that of library, documentation and publication functions, and metadata may also be generated by an organization's business functions, but for simplicity, we refer to all of these functions here as records management.) This implies that most if not all of the metadata describing these resources should be generated or collected, maintained, and managed by records management, though records management metadata are not generally created or collected for discovery purposes. In contrast, the access domain, which includes websites that make information available over the network, should ideally obtain the metadata describing information resources from records management, rather than generating such metadata itself. For example, such metadata values as the Title, Publisher, Publication-date, etc., for a resource should simply be obtained from the records management domain. This one-way path for metadata from records management to access is shown in Figure 1a (based on a concept of Hans Hofman of the Dutch National Archives).
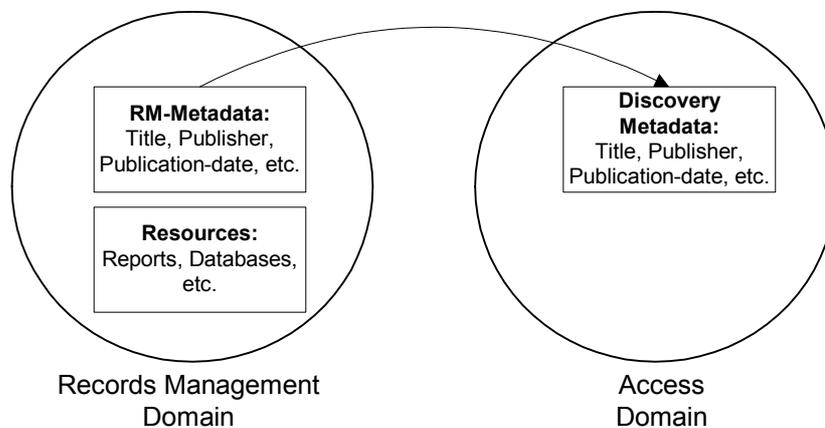
Figure 1a: Derive access metadata from records management

One important exception to the simple flow of metadata shown in Figure 1a lies in the generation of search terms, keywords, and ontologies. The access domain may need to produce a richer or at least different set of such discovery metadata than is required by records management, and it may need to translate search terms that are used internally by records management into terms that are more meaningful to users searching government websites. In some cases, these additional terms may be of interest to the records

management domain as well, or records management may offer to serve as the repository for them; but in most cases, they will be of interest only to the access domain, so it makes more sense for them to be stored, maintained, and managed there. This is illustrated in Figure 1b. In addition, the figure shows that although the access domain may often reference online resources that are owned by records management, it may also create its own copies of resources, for example in order to "hand craft" web-friendly representations of documents or databases. Any such copies that are created in the access domain must then be stored and managed within that domain (unless records management is willing to store and manage them).
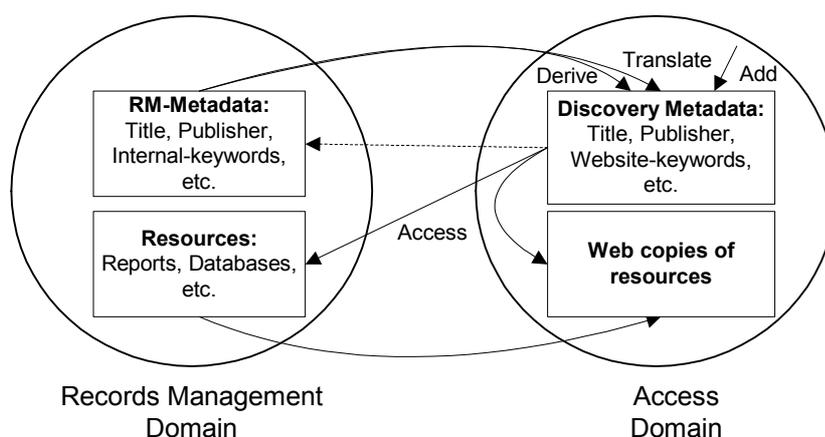
Figure 1b: Derive/modify/store access metadata from records management

As noted above, search terms should ideally be organized into CVs, represented by taxonomies, thesauri, or ontologies. The generation of ontologies of useful categories of terms requires considerable work and coordination among COIs, and the resulting CVs must be maintained by their own records management function, which may be required to span multiple organizations, corresponding to the members of a COI. It is an open question where this COI records management function should be performed.

Unfortunately, the current records management landscape in the Netherlands is very diverse. There is as yet no widespread agreement on exactly what metadata should be created or maintained in the records management domain, and few organizations have implemented true digital records management functions, let alone records management systems. Therefore, although the flow illustrated in Figure 1 may be the ultimate ideal, it may not be feasible until some time in the future. In those organizations whose access domain has a need for metadata and whose records management function is not yet in place, it will be necessary for the access domain to create and manage its own metadata prior to the implementation of a true digital records management function. In such cases, the access domain may need to create some metadata that it can use for administrative (i.e., records management) purposes, even though that may be irrelevant for discovery and access, which are the mission of the access domain. That is, the access domain may need to implement an interim surrogate digital records management function, simply to enable

it to manage (e.g., store, access, and maintain) its own information resources, such as webpages and the documents, records, databases and other assets to which they are linked. Implementing this interim surrogate digital records management function should not require a significant additional effort beyond the basic administrative tasks which any website manager must already perform to maintain a website. We merely point out that these tasks should be thought of as a form of records management.

Whether or not an interim surrogate digital records management function must be implemented in the access domain of an organization, it is important for the Metadata Manager to maintain close ties to records managers within the organization and to maintain an ongoing effort to integrate access domain metadata with records management metadata.

5.3   The need for structured metadata

One of the key questions concerning Dublin Core and its derivatives is whether their essentially flat structure is sufficient. As noted above, the elements in a flat metadata structure all exist on the same level. Although DC element refinements add a partial second level, a user of the standard cannot create new levels as needed or regroup elements, for example, to allow representing each event in the lifecycle of a resource as having an instigating cause, a date, a responsible party, contributors, status, and result, all of which are bundled together. There are a number of arguments for providing more structuring capability in a metadata standard than the limited mechanism provided by the element refinements in DC. We group these arguments into three categories, which we refer to here as subsetting, encapsulating, and describing complex resources or relationships.

Subsetting refers to the ability to group different subsets of metadata elements together into groups with related meanings or functions. For example, metadata elements might be grouped as administrative, records management, rights-management, intellectual property control, or elements that help ensure the longevity of information resources. Subsets of this sort can make larger sets of metadata elements easier to understand and use. Another use for subsets is to hide collections of elements that are of interest only in certain cases; for example, e-mail header fields (such as "addressee" or "cc") are of interest only when describing e-mail messages and should ideally be hidden in all other cases. The element refinements in DC provide a kind of subsetting, but the subsets they form are not always meaningful, as for example, various refinements of Date (such as Date.created, Date.revised, Date.redacted, Date.destroyed) may serve quite different functions, spanning discovery and records management.

Encapsulation is similar to subsetting, but it addresses the need to combine metadata elements that are not merely similar in function but are logically tied to each other. An example of this is the concept of an event in the lifecycle of a resource, as mentioned above. Although DC-like elements or refinements could be used to represent each of the attributes of an event (cause, date, contributors, result, etc.), there is no mechanism for bundling or encapsulating them into a single representation of an event: ideally, there should be a bundled set of these attributes for each event, e.g., Created(cause, date, contributors, result), Published(cause, date, contributors, result), Revised(cause, date,

contributors, result), etc. A related example is geographic coverage, which may require more complex description than a single or small number of pre-defined element refinements can provide, e.g., arbitrary sets of coordinates and place identifiers.

The final and perhaps most compelling need for structured metadata is to describe more complex resources or complex relationships among resources. In particular, Dublin Core and most other library or archival description schemes are still focused on distinct, static objects, such as books or individual records. In the online world, however, many information resources are far more complex than this, consisting of multiple components, each of which may require a different interpreter (application program or viewer) to render it as part of the overall result desired by a user. In many cases online information is generated dynamically, by means of access to a database or GIS (graphic information system) or by the execution of programs (Active Server Pages, Java Server Pages, etc.), which may not have obvious values for elements such as Title, Creator, or Publication Date. Furthermore, users are often looking for information rather than specific documents or resources: they may not know or care where the desired information resides. An information site that caters to such questions may need to describe sources of information that span multiple documents, databases, or even websites. Moreover, many sites provide services rather than simple information, for example interactive forms or transactions to accomplish tasks such as filing taxes, requesting permits or licenses, or making payments. Few existing metadata schemes address these issues, but structured metadata would provide a more powerful mechanism for describing composite objects and resources of these kinds. Finally, it is sometimes necessary to describe complex relationships among documents, records, agents and activities or events (mechanisms for representing such relationships in the records management domain, are discussed in ISO 23081-1).

5.4   An "optionally-structured" approach

There are a number of alternative ways of designing structured metadata, including the use of a relational database (RDB), object-oriented ("O-O") techniques (often implemented on top of an RDB), or semantic nets (such as Topic Maps). Each of these is subsumed by the next, in the sense that an O-O approach can do anything an RDB can do, whereas a semantic net can do anything that either of the other two approaches can do.

An RDB approach to metadata might consist of a set of relational tables, each of which contains information about one metadata element or refinement for each resource. Using links between the rows of these tables (e.g., references to "foreign keys"), metadatabase designers can create the effect of structured subsets or groupings of metadata as described above. Users can then access these groups of metadata elements by issuing queries that join tables (in the usual, relational database sense of the word "join").

An O-O approach would go a step further by creating named structures or groups of metadata elements, which could then be accessed by name, without having to construct queries to create these groups by joining multiple tables. In practice, O-O approaches are often implemented on top of relational databases, simply to take advantage of some of the features of most RDB systems, such as multiple-access, record-locking, backup and

recovery mechanisms, etc. But the O-O approach allows both metadata managers and metadata users to think directly in terms of objects without having to know which tables must be joined to create the functionality of objects. That is, the O-O approach provides "encapsulation" which is simply the ability to treat collections of items (in this case, metadata elements) as a single named unit. Most O-O approaches provide a class-subclass (or "IS-A") relationship and (often) a part-whole relationship, as well. Some of them allow arbitrary relationships to be defined, though few allow relationships themselves to be viewed as objects with their own attributes. (Note that the Dublin Core "dumb-down" rule--which says that a user of a refinement can ignore the refinement name and think of it as an instance of the higher-level element instead--is simply a restatement of the object-oriented "IS-A" relationship, which defines a subclass as a refinement of its parent class.) O-O approaches are well suited to representing hierarchies, but they are limited in representing more complex structures, such as nets of webpages or ontologies.

A semantic net approach is even more powerful, allowing arbitrary relationships to be defined and relationships themselves to have their own attributes. Semantic nets can do anything that an RDB or O-O approach can do, and they are well-suited to representing complex structures, such as webpages, ontologies, workflows, processes, or services.

Note that any of these approaches can easily represent a flat structure, such as the list of elements and refinements in Dublin Core or any of its derivatives.

Each of these structuring approaches has its advantages and limitations, and a structured metadata approach that utilized a relational database would be quite different from one based on objects or semantic nets. Yet we feel that the choice of which of these (or other) techniques to use to add a structuring capability to a metadata approach is a technical implementation issue that should not be constrained by the standard itself. We therefore provide, instead, an abstract framework that allows any such technique to be used to add structured metadata to the standard.

Since this framework also allows an implementer to ignore the standard's structuring capability and use a flat, DC-like list of elements, we refer to this as an "optionally-structured" approach. The widespread use of even the flat, unstructured part of the proposed standard would represent a major step forward in improving discovery and access, but the optional use of structure makes the standard more flexible and powerful.

5.5   NL-meta: a proposed nationwide discovery metadata standard

Figure 2 presents a class diagram for the metadata in the standard: the top-level NL-meta class can have one of two instantiations, represented on the left by the NL-meta.Extended subclass and on the right by the NL-meta.DC+ subclass. The NL-meta.Extended subclass can in turn consist of any structured metadata scheme, including RDB, object-oriented, or semantic net approaches. But the NL-meta.DC+ subclass is limited to the specific set of DC-derived elements and refinements listed in Section 5.5.1 below.

This framework allows a given "metadata set" to be either flat or structured. (Since the term "record" is overloaded in the context of this discussion, we do not use it in the database sense to denote a collection of metadata element values. Instead, we use the term

"metadata set" to refer to such a given collection of values. For example, we refer to the Dublin Core description of a single resource, consisting of values for Title, Creator, Publisher, Date, etc., as a single metadata set.)

In order to create a metadata set using the framework in Figure 2, the Metadata Manager of an organization that is implementing the standard first decides whether to use the structured (left) or flat (right) subclass of the top-level NL-meta class. If the flat NL-meta.DC+ subclass is chosen, then the simple list of elements described in Section 5.5.1 is used, and the result is similar to the Dublin Core model. On the other hand, if the structured NL-meta.Extended subclass is chosen, then the implementer is free to create a structured metadata set using a relational database, objects, semantic nets, or any other suitable technique. For example, in order to represent events in the lifecycles of resources (e.g., creation, publication, release, revision, invalidation), an object-oriented structure might be defined, in which each named event consists of a number of related sub-elements, such as cause, date, responsible party, contributors, status, result, etc.

Any given NL-meta metadata set should consist of either an NL-meta.Extended set or an NL-meta.DC+ set, but probably not both, to ensure that a given metadata set cannot contain conflicting metadata represented by both DC-derived elements and some other, structured representation.

(Note: these would re-implement NL-meta.DC+)
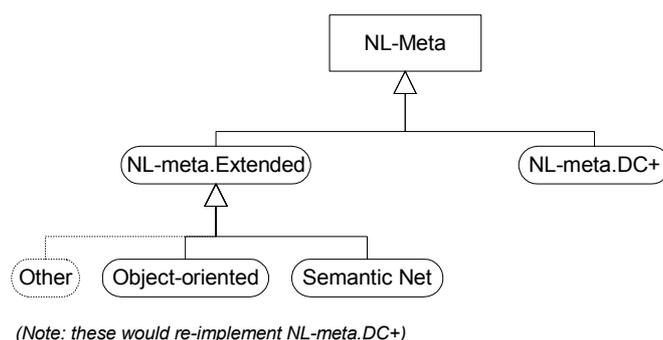
Figure 2: An "optionally-structured" metadata approach

Since the left (structured) option can embed a representation of a flat NL-meta.DC+ metadata set, a structured metadata set can contain both structured and unstructured metadata within it. For example, an RDB metadata representation could contain either one table for each NL-meta.DC+ element or refinement or a single table containing all NL-meta.DC+ elements and refinements as its columns. Similarly, an O-O representation could contain a single object whose attributes corresponded to the NL-meta.DC+ elements and refinements. That is, any structured representation can be used to recreate a flat metadata set, just as a hierarchical organization chart can represent a company with a "flat" management structure (i.e., one whose employees are all on the same level, with no one reporting to anyone else).

An implementer who is uncomfortable with a structured metadata approach or has no current need for it can create metadata on the right-hand branch of the model, using the simple NL-meta.DC+ approach. If and when the need for structure arises, this metadata set can be replaced by a structured approach (using the left-hand branch of the model) by simply representing each of the NL-meta.DC+ elements in the new, structured metadata set, as described above (at least the mandatory elements of the standard would have to be represented in this structured form). This allows most implementers to ignore the structuring capabilities of the standard and use it as if it were a DC derivative. Yet any implementer who needs more power than the DC approach provides can use the left-hand branch of the model to create structured metadata, either now or in the future when the need arises. This approach to extensibility is far more flexible and powerful than the simple refinement mechanism of DC (though refinements can still be added on the right-hand branch of the standard, to extend the NL-meta.DC+ elements, if desired).

5.5.1 The NL-meta.DC+ elements

The Obligation entries in Table 2 (NL-meta.DC+ elements) have the following meanings:

Table 1 Meaning of obligation entries

Obligation entry	Meaning
M	Mandatory
MiA	Mandatory if Applicable
R	Recommended
O	Optional
Admin	Administrative

The last of these (Administrative entries) are elements that must be implemented in order to satisfy basic resource management. Except for these, the standard includes only those elements that offer some advantage for discovery: records management elements are eliminated, since it is the responsibility of records management to create and maintain them. However, as mentioned above, if no records management function is in place in an organization when a discovery metadata scheme is first implemented, those elements marked as Admin must be created in addition to the other, discovery-oriented elements in the standard, to enable the access domain to perform its surrogate digital records management function. If a digital records management function exists, it should not be necessary for the access domain to create and maintain these Admin elements, since records management will manage metadata of its own to serve these administrative needs. A definition of the elements can be found in Appendix E.

For traceability, each element and refinement in Table 2 is preceded by an indicator of its origin. This indicator is either "DC" for Dublin Core, or "OVERHEID" if the element and/or refinement are not part of Dublin Core but items borrowed from extension of Dublin Core like "eGMS" (from the U.K. national standard), "CEN" (from the CEN Workshop), and "FI" (from the Finnish national standard), or added by Advies Overheid.nl.

Table 2 NL-meta.DC+ elements

Scheme	Element name	Refinement (Qualifier)	Encoding Scheme	Obligation
DC	date		Y	M
DC	date	created		
DC	date	valid		
DC	date	available		
DC	date	issued		
DC	date	modified		
OVERHEID	date	nextVersionDue		
OVERHEID	date	updatingFrequency	Y	
DC	language		Y	M
DC	subject		Y	M
OVERHEID	subject	category	Y	
OVERHEID	subject	keyword	Y	
OVERHEID	subject	person		
OVERHEID	subject	processIdentifier		
OVERHEID	subject	programme		
OVERHEID	subject	project		
DC	title			M
DC	title	alternative		
OVERHEID	title	abbreviation		
OVERHEID	audience		Y	MiA
OVERHEID	audience	educationLevel		
DC	coverage		Y	MiA
DC	coverage	spatial		
DC	coverage	temporal		
DC	creator		Y	MiA
OVERHEID	creator	personalName		
OVERHEID	creator	corporateName		
DC	identifier		Y	MiA
DC	identifier	bibliographicCitation		
DC	publisher		Y	MiA
DC	type		Y	MiA
OVERHEID	type	administrativeBody		
OVERHEID	type	aggregation	Y	
OVERHEID	type	organisation	Y	
DC	format		Y	R
DC	format	extent		
DC	format	medium		
OVERHEID	status		Y	R
OVERHEID	accessibility			O

Scheme	Element name	Refinement (Qualifier)	Encoding Scheme	Obligation
DC	contributor		Y	O
OVERHEID	contributor	personalName		
OVERHEID	contributor	corporateName		
DC	description			O
DC	description	tableOfContents		
DC	description	abstract		
OVERHEID	mandate			O
DC	relation		Y	O
DC	relation	isVersionOf		
DC	relation	hasVersion		
DC	relation	isReplacedBy		
DC	relation	replaces		
DC	relation	isRequiredBy		
DC	relation	requires		
DC	relation	isPartOf		
DC	relation	hasPart		
DC	relation	isReferencedBy		
DC	relation	references		
DC	relation	isFormatOf		
DC	relation	hasFormat		
DC	relation	conformsTo		
DC	source		Y	O
OVERHEID	preservation			Admin
DC	rights		Y	Admin
OVERHEID	rights	copyright		

5.5.2   Rationales for elements and refinements

The overall rationale for which elements and refinements to include in the NL-meta.DC+ set is: except for "Admin" elements, include only those that provide enough benefit for discovery to warrant the effort needed to create and manage them. If elements in the initial version of the standard do not in practice satisfy this criterion, they should be eliminated, whereas if additional elements are found to satisfy it, they should be added.

Detailed rationales for individual elements and refinements in the NL-meta.DC+ set are provided in Section 4.3 of the Handbook and are not repeated here.

SECTION 6 **Phase III: Analysis of implementation issues**

In phase III of the project we analyzed implementation issues that we had identified in earlier phases, such as the relationship between the access and records management domains. As part of this analysis, we realized the importance of creating a Metadata Manager role for each information provider organization and giving this role broad responsibility for managing and coordinating metadata, both within the organization and across various Communities of Interest. In addition, we explored the idea of establishing a national clearinghouse to coordinate the creation and activity of COIs and their creation and use of CVs (see Section 6.1.4). This phase also examined the key technical implementation issue of whether metadata should be "embedded" in resources (for example, by inserting HTML or XML tags, such as <meta date="07-05-2004"> in each resource) or should be kept in databases or other structures that are separate from the resources themselves. Finally, we analyzed some of the issues involved in creating metadata for older, pre-existing resources, as opposed to new ones.

In addition to drawing on the collective experience of our project members with previous implementation efforts, we gathered input on implementation issues from a number of sources, including our interviews with website managers and other government information providers, published reports on metadata implementation, and discussions with the eGMS and eGOV offices in the U.K.

6.1   Policy and bureaucratic implementation issues

This section discusses key implementation issues involving policy, bureaucracy and organization. Section 6.1.1 discusses the need to engage a wide range of stakeholders, whereas Section 6.1.2 discusses the use of an online working group to support stakeholder interaction. Section 6.1.3 discusses the need for a Metadata Management function and a Metadata Manager in each organization, and Section 6.1.4 discusses the need for a national COI clearinghouse. Section 6.1.5 offers a roadmap for achieving the 2007 goal of providing online access to published information, and Section 6.1.6 discusses some remaining issues.

6.1.1 **Engaging stakeholders**

Not surprisingly, one of the main issues that emerged from all of our sources was the importance of engaging all relevant stakeholders as active partners in the creation and implementation of a national metadata standard. Standards that are developed in a vacuum tend to be received with skepticism or hostility by those who must implement them. To paraphrase an American Revolution slogan: "No implementation without representation."

Engaging stakeholders has several, well-documented benefits. In addition to ensuring that all relevant concerns, needs and interests are at least considered during the design phase of a standard, engaging stakeholders greatly increases the likelihood of acceptance, since they become joint owners of the standard. In the Netherlands, where compliance with a metadata standard would be voluntary rather than centrally mandated, this joint ownership factor is especially important. Furthermore, since many information provider organizations see no reason to implement new discovery metadata, their engagement in an ongoing dialogue on the needs and benefits of such metadata may be a crucial factor in convincing them of the utility of this endeavor. Therefore, although this project has produced an initial proposed standard (and an initial Handbook for its use), the larger stakeholder community can and should be included in future discussions, evaluations, and revisions of the proposed standard and Handbook.

Among the groups that should be engaged are records managers and archivists. These groups bring an important perspective to metadata and should eventually create, maintain, and manage much of the metadata needed in the access domain (even though this will not be possible at present in organizations that have not yet implemented a true digital records management function).

In addition, it is important to engage records management stakeholders in order to properly conceptualize the distinction and relationship between access domain metadata and records management metadata before implementing the former. Metadata standards efforts that have failed to do this initially have had to do so later, at the cost of considerable rework and revision.

Most organizations that have a records management function do not seem to have integrated or coordinated it with their website management. For example, many of the ministries that we interviewed indicated that their records management and websites consist of "islands of information" that are isolated from each other, often using different definitions of metadata and different controlled vocabularies. In such cases, it is obviously crucial to bring these isolated groups together to coordinate their definitions and use of metadata.

If no records management function is in place when an access domain metadata scheme is implemented in an organization, some records management metadata and functions may have to be created in the access domain, simply to manage data (and metadata) resources. Records management input to this process will make it more effective and more likely to integrate with whatever records management approach is eventually implemented. One of

the keys to enabling future integration of this sort is to use stable, universal identifiers when describing entities, so that records managers can determine whether access domain metadata already exist for a given entity that they later incorporate in their own domain.

6.1.2 Online working group for stakeholders

Stakeholders can be engaged by a number of mechanisms. The current Advies Overheid.nl online working group provides one potentially effective way of involving relevant parties in decisionmaking without incurring the cost of meetings and travel. This working group can be expanded to include all government information producers as well as representatives of users, such as citizen organizations (e.g., burger@overheid.nl) and consumer groups (e.g., the Consumentenbond). If it is to be useful, however, the working group must not be allowed to grow to the point that it becomes unwieldy. In order to balance the competing goals of making it big enough to be inclusive and small enough to be responsive, it may be factored into sub-groups and make use of online committees and action groups that represent the larger membership in performing most of the actual decisionmaking.

The online working group should function as a knowledge and support center, utilizing active participation by Advies Overheid.nl staff. It should serve as a resource for organizations that want to improve discovery of their online offerings, and it should be used to motivate organizations and other stakeholders to engage in an ongoing decisionmaking process surrounding metadata and discovery. Finally, working group interactions should be captured to help update the Handbook with additional specifics, such as process models for creating and using metadata.

In summary, the online working group should be used proactively to generate support for the adoption of the NL-meta standard and to counteract any resistance that may arise in organizations that do not yet see the need for such a standard or for the use of metadata to improve discovery.

6.1.3 The need for a Metadata Management function and a Metadata Management Plan

The introduction of metadata into an organization that produces online information requires careful management. Most important is the management of expectations, both nationally and within each organization. This must include recognition of the fact that the creation, management, and use of metadata will add a new burden. A Metadata Management function should be created within each organization, ideally headed by a designated Metadata Manager. A Metadata Manager should be well-versed in the use of metadata and familiar with the organization's missions as well as its records management, resource management, and online publication and e-Service functions. (The hard-won experience of several other national metadata efforts suggests that the creation of a Metadata Manager role is a key element in the successful introduction of metadata in any organization. Therefore, although a Metadata Management function can be performed

without having a Metadata Manager, we assume throughout this Handbook that such a manager will exist in each organization.)

Since few government organizations recognize the need for the formal use of metadata to improve discovery of their information assets, it may be difficult to convince many organizations of the importance of establishing a Metadata Management function. Advies Overheid.nl's online working group should help disseminate arguments for creating this function, explaining the potential benefits of metadata when coupled with finding-tools that can utilize metadata to aid discovery. In smaller organizations, website managers may serve as default Metadata Managers, so long as the scope of their activities includes coordinating efforts across multiple internal websites and interacting with records management and other organizations in relevant COIs.

The first job of a Metadata Management function should be to develop a Metadata Management Plan for the organization. This plan should include at least:

- An overall metadata strategy

- A metadata creation and management strategy

- An internal and external metadata coordination strategy

- A metadata quality control strategy

- A strategy for making metadata accessible to finding-tools

- An action plan for implementing these strategies

The Metadata Manager should work with internal website managers, records managers, and publication managers, as well as with Advies Overheid.nl and relevant COIs to develop, implement and manage the overall metadata strategy. This strategy should specify the expected roles and interactions of records management, library, publications, website management, and metadata management functions.

The metadata creation and management strategy should specify where discovery metadata will be developed and managed and whether it will be embedded in online resources or maintained separately, in a database of metadata (i.e., a "metadatabase), as discussed in Section 6.2.1.

The internal and external metadata coordination strategy should provide procedures for coordinating metadata choices and ensuring consistency of metadata across multiple internal websites within the organization and with other relevant external organizations (for example, through the use of appropriate COIs).

The metadata quality control strategy should include procedures for managing changes to the metadata scheme used within the organization. Quality control should include ongoing verification of the existence and form of appropriate metadata and validation of their correct relationship to the resources they describe.

The metadata management plan should ensure that metadata will be accessible to finding-tools, in order to aid discovery. This may require choosing or developing tools that make metadata accessible to existing finding-tools or choosing or developing finding-tools that

can recognize embedded metadata within online resources or that can access metadata in separate metadatabases.

Finally, an action plan should be developed to assign roles and timeframes to the tasks required to implement the metadata management plan.

The Metadata Manager should oversee and coordinate efforts to create and use metadata consistently and interoperably across multiple internal websites as well as among related agencies, government portals, and relevant COIs. This should include coordinating the use and evolution of metadata refinements and extensions, encoding schemes and CVs, taxonomies, thesauri, and ontologies with appropriate COIs.

6.1.4   The need for a national clearinghouse for COIs

As pointed out above, in the discussion of Phase I (Section 4.2.2), the concerns and subject areas of COIs may sometimes overlap, and users may need to combine or compare information from across multiple COIs. This means that the metadata and terminology (including encoding schemes, CVs, ontologies, etc.) developed by each COI must be coordinated with those of other COIs. If COIs are allowed to form and act without such coordination, the result is likely to be a lack of semantic interoperability.

In order to mitigate this risk but still take advantage of the COI as a mechanism for subdividing the task of tailoring the use of the national metadata standard to the needs of specific groups, it is important to establish some kind of national clearinghouse that can help coordinate the efforts of COIs. This function may be performed by Advies Overheid.nl or some other central government organization. The clearinghouse should act as a knowledge and support center for COIs and organizations and COIs.

Although it may be possible to perform much of the desired function of a national COI clearinghouse by means of an online working group, some agency must still support this function. At a minimum, this would require maintaining a database that describes the functional scope and lists the member organizations and points of contact of each COI. In addition, the clearinghouse would have to develop mechanisms for mediating any conflicts that may arise among COIs. The design of these mechanisms lies beyond the scope of this project, but as pointed out in Section 4.2.2, if the clearinghouse fulfills its primary coordination function of minimizing the overlap in the scope of distinct COIs, conflicts among COIs should be minimized, and it should be possible for each COIs to make its own decisions.

We strongly recommend that Advies Overheid.nl explore the establishment of a national COI clearinghouse, either under its own auspices or within some other appropriate agency, in order to ensure that the COI mechanism fulfills its potential.

6.1.5   A general roadmap for providing online access to published information

The following roadmap suggests a sequence of steps that should enable the Dutch government to meet its goal of providing online access to all legally mandated published information by 2007:

1. Engage all relevant providers of government information.

Whether this is done via an online working group or in combination with other mechanisms, such as newsletters, symposia, or meetings, it is vital that information provider organizations at all levels of government be brought on board early in the endeavor, if it is to succeed. This process should be ramped up quickly so that virtually all relevant organizations become active participants by mid-2005.

2. Develop and refine the NL-meta nationwide standard for discovery metadata.

The initial draft of this standard developed by this project should be reviewed, tested, evaluated and revised by relevant stakeholders within the first year of its publication, in order to improve confidence in its applicability and usability.

3. Develop and refine "metadata-aware" finding-tools (e.g., search engines) that can access and utilize the discovery metadata in the NL-meta standard to improve discovery.

Concurrent work should proceed on the development of finding-tools that can access the discovery metadata specified in the NL-meta standard (whether it is embedded in resources or kept separately) and can use it to improve the discovery process. Initial versions of such tools should be made available to information provider organizations by mid-2005.

4. Develop and refine process models for the implementation and use of NL-meta discovery metadata by government information providers.

This should include process models for implementing the NL-meta standard, for using it to create and maintain discovery metadata for whatever online information an organization publishes, and for ensuring that the resulting metadata can be accessed by finding-tools that are available to online users. These process models should be refined, extended, tested, and evaluated by relevant stakeholders to ensure that they are easily implemented and effective. Initial, tested process models for implementing and using unstructured NL-meta.DC+ metadata should be made available to organizations by mid-2005.

5. Develop and deploy instrumentation to measure the effectiveness of the online discovery and access of government information.

It is necessary to develop and utilize instrumentation of this kind in order to evaluate the degree to which the goal of making government information available online has been realized. Instrumentation should be developed, and baseline measurement using this instrumentation should be completed, by the end of 2005.

6. Promote and support the adoption and use of the NL-meta standard, metadata-aware finding-tools and implementation process models by all relevant government information providers.

In order to meet the 2007 goal, the government should actively promote the use of discovery metadata and its associated tools by all organizations that publish online government information. Since promotion by itself may be insufficient, substantive support should be provided, in the form of published guidelines, process models, freely-available finding-tools, online consultation and working group participation, the national COI clearinghouse, and feedback from repeated measurements of the

effectiveness of discovery and access, to help organizations improve the transparency and availability of their information. This should be an ongoing, evolving process, but basic support functions should be put in place by mid-2005.

7. Evaluate the effectiveness of online access to legally mandated published information, and use the results of this evaluation to refine the NL-meta standard, finding-tools, and implementation process models.

Measurement and evaluation of this sort is necessary in order to determine whether the goals of improved transparency and access have been met. Without such measurement, the endeavor will be run "open-loop" without a firm basis for deciding whether it has succeeded. Initial measurement and evaluation of the improved effectiveness of pilot projects and "early-adopter" organizations should begin by early 2006, to enable other organizations to utilize feedback from these evaluations by the goal date of 2007.

6.1.6   Additional bureaucratic implementation issues

It is important to recognize that any metadata scheme is merely a means to an end. Discovery and access metadata will be useful only if their creation and maintenance are supported and only to the extent that they are made available and accessible for discovery and access purposes. That is, metadata descriptions of resources are useful only if search engines or other finding-tools can access and use them. This implies that defining a metadata standard must be seen as merely one of several necessary steps. Further steps should be taken to explore ways of making discovery metadata accessible to finding-tools. In addition, systems and process models must be put in place to help create, maintain, manage, and use discovery metadata.

Finally, a number of steps should be taken to support the adoption and continued evolution of the metadata standard and to coordinate and facilitate the formation of COIs and their development of CVs and ontologies. Our recommendations (Section 7) include a number of specific tasks that Advies Overheid.nl might undertake to pursue these goals.

6.2   Technical implementation issues

This section discusses key implementation issues involving policy, bureaucracy and organization.

6.2.1   Embedded vs. separate metadata

When implementing metadata, a common question is whether they should be embedded in the resources they describe or kept separate and linked to those resources. There are advantages to both of these approaches, and in the interest of avoiding "early binding" (i.e., choosing an implementation too soon), we discuss these two alternatives here without

choosing among them. We also point out that it is possible to combine these approaches to some extent.

Embedding metadata in online resources amounts to inserting metadata element names and values (attribute-value pairs) into online documents, databases, etc., using mechanisms such as HTML or XML tags or RDF descriptions. Actually, as we have pointed out above, much of the online information and service that users seek is dynamic and may not consist of distinct objects, so it may be difficult to find an appropriate place to insert metadata in such cases; but we ignore that problem for the moment. The advantage of inserting metadata into resources in this way is that search engines, such as Google, can thereby (in principle) find the metadata and allow users to search for these attribute-value pairs. For example, a tagged value such as <meta title="Encoded Archival Description"> allows searching for this embedded title in a resource. In order for metadata to be useful for discovery purposes, it is vital that search engines and other finding-tools be able to access them, and embedding provides a simple mechanism for achieving this. On the other hand, the perception of at least some of our surveyed information provider sites is that current search engines do not do a very good job of using embedded metadata tags of this sort, though in the absence of a baseline measurement of discovery effectiveness, this perception is difficult to verify.

Unfortunately, there are several disadvantages to embedding metadata in resources. For one thing, it requires editing or modifying the resources themselves, which may not be feasible in all cases. If a resource is generated by a government website for online access purposes, or is produced as a copy of some other resource, then it may be possible to insert such information in it. But if the site merely provides access to a digital resource that was produced elsewhere (for example, a PDF report or a TIFF image), it may be impractical or impossible to insert anything into the resource (this is especially true for offline resources that are described by online metadata, for example CD-ROM publications). Furthermore, whenever a resource is revised, its metadata must be reinserted into it, and whenever the metadata themselves are modified (for example, to insert new keywords, category descriptors. or search terms) the resource must be edited. Finally, as noted above, for dynamic resources such as database views, interactive forms or active web-pages, it may not be meaningful to insert metadata into the resource itself.

Keeping metadata separate from the resources they describe (i.e., in a database or some other kind of metadata management tool) avoids the problem of having to modify the resources themselves to insert or update metadata. Furthermore, it allows metadata to be kept in a more flexible form, such as in a relational or object-oriented database or in a semantic net (such as a Topic Map). This facilitates exploiting the optionally-structured approach we propose. Finally, it enables metadata to describe resources that are not editable and that are dynamic or ephemeral, such as database views, transaction facilities or other e-Services.

The risk of this approach is that metadata values may become logically disconnected from the resources they describe. However, the use of one of the accepted universal schemes for identifying objects, such as ISBN (International Standard Book Number) or DOI (Digital Object Identifier) should greatly reduce this risk.

Note that the use of stable, reliable identifiers when describing resources is crucial for another reason as well. If an organization creates discovery metadata for a resource in the access domain before implementing a records management function, the use of stable identifiers should enable future records managers to determine that such access domain metadata already exist for the given resource when they later incorporate the resource into their own domain; this should help prevent the creation of inconsistent metadata for the resource.

Finally, separate metadata must be made accessible to finding-tools. This may be done by "exporting" metadata tags to be embedded in resources (a hybrid approach), or it may require providing query-language interfaces to databases or application program interfaces (APIs) to other systems that store the separate metadata.

As the hybrid solution suggests, the embedded and separate metadata approaches are not mutually exclusive. For example, metadata may be managed in a database or represented in a semantic net (such as a Topic Map) and used to generate and export tags that are then embedded in resources. This is an advantage only if the export process can be automated, but if so, it combines some of the best (and worst) of both approaches. That is, it makes it easier for finding-tools to access the metadata, but it requires updating the resources themselves whenever the metadata are changed.

Whichever implementation approach is chosen, it is necessary to enable finding-tools (such as search engines) to access discovery metadata, if they are to be of any use. The embedded approach has the advantage that embedded metadata tags can be found as text by free text search engines, though it helps if those engines understand metadata tags and interpret them as such. If metadata are kept separate from the resources they describe, finding-tools must be given access to the separate metadata to enable them to find the desired resources. This may ultimately be an advantage, if finding-tools are designed to be able to browse or search ontologies or other terminological aids.

In order to leave future implementers free to choose whichever approach best suits them, we do not take a stand on the question of embedded versus separate metadata. We merely note that whichever approach is taken, the following criteria should be ensured:

- Metadata must remain logically linked to the resources they describe

- Metadata must be accessible to finding-tools

- Modifying metadata should minimize the risk of corrupting resources

- Revising resources should allow appropriate metadata to be easily updated

So long as these criteria are met, it is immaterial whether an embedded or separate metadata scheme (or a hybrid of the two) is adopted.

6.2.2 Creating metadata for old versus new resources

The AGLS Implementation Manual (AGLS, 2003) lists a minimum set of resources that require AGLS metadata:

- Home pages

- Topics/services in high demand

- Information required by agency clientele to understand their entitlements to government assistance or obligations

- Pages that provide actual online services to the public

- Pages describing obligations by the organization

- Entry points to specific online services and indexes

- Major formal publications

- Media releases

- Major entry points to indexes and menus to a range of closely related topics, programs or policies

- Information about agency powers affecting the public

- Substantial descriptive or marketing information

The AGLS manual also describes the use of a Harvest Control List (HCL), which is an index that makes separate discovery metadata visible to search engines.

The creation of discovery metadata to describe new online resources can and should take advantage of technical mechanisms for automating at least part of the metadata creation and capture process. For example, a document management system or even a word processor may provide features for capturing the creation date, author, and type of a new resource when it is created. Unfortunately, such automatic facilities are often inadequate for true metadata creation. For example, they may insert the login name of a personal author, rather than the preferred, impersonal organization name, or they may update a "revision date" every time a document is edited, rather than when a significant change is made to it. Nevertheless, it should be possible to obtain or modify applications or to develop tools to capture at least some such metadata automatically for newly created resources. In particular, organization-specific and function-specific default values for certain metadata fields, such as publisher and organizational author, language, coverage, type, format, creation date, etc., may be automatically supplied at this time, though it must be possible to override such defaults at the time of creation or later.

Creating metadata for older resources that already exist may be less amenable to automation. If embedded metadata are to be added to old resources, it may be necessary to modify the resources themselves, which may be difficult or impossible in some cases, as discussed above. Moreover, the values for some metadata fields (such as coverage) may be harder to obtain after the fact, requiring careful analysis of the resources themselves or discussion with original authors or publishing organizations, assuming they are still available. Finally, for offline resources that are described by online metadata, some metadata elements and refinements may be inapplicable. These problems are similar to the well-known problems of dealing with legacy data, and although none of them are insurmountable, they imply the need to develop detailed, organization-specific strategies for generating metadata for older resources.

6.3   **The Handbook**

As requested by Advies Overheid.nl, a draft Handbook was developed in parallel with this report. The Handbook provides the rationale and motivation for using the proposed standard and offers advice to information provider organizations on how to implement and use it.

Conclusions and recommendations

This section summarizes the most important conclusions that are discussed above under the phases of the project and formulates our recommendations based on these conclusions. Section 7.1 gives general recommendations for actions that should be carried out by organizations that adopt and use the proposed metadata standard. Section 7.2 analyzes key risks involved in promoting the standard, along with suggested ways of mitigating each risk. Section 7.3 recommends specific actions that Advies Overheid.nl should take to advance the national metadata standard endeavor.

7.1 General recommendations

This section recommends actions that should be taken by the community of information provider organizations in order to adopt and use the proposed national metadata standard.

7.1.1 Create a Metadata Manager role and a Metadata Management Plan

It is important to designate a Metadata Manager within each organization, whose job it is to develop and implement a Metadata Management Plan for the organization and to coordinate efforts among internal websites and with relevant COIs.

7.1.2 Create an interim surrogate digital records management function if needed

As discussed above, if the access domain in a given organization creates metadata before a records management function has been put in place, it may have to create some administrative metadata as well, to satisfy its own records management needs. For example, if it is desired to ensure the continued accessibility and usability of online resources, preservation metadata for those resources must be generated and maintained; if this cannot be done by records management or archiving functions, it must be done in the access domain. Similarly, metadata may have to be created and maintained simply to store, manage, and use the resources created within the access domain. The Metadata Manager may therefore have to create an interim surrogate digital records management function to manage the access domain's own resources until a true digital records management function is implemented for the organization. The effort required to implement this interim surrogate digital records management function should not be significantly greater than that which is needed to maintain a website in the absence of a

supporting records management function, but this effort should be conceptualized as a records management function. In addition, the Metadata Manager should coordinate this effort with records managers within the organization, eventually transferring most of this effort to a true records management function, as one emerges.

7.1.3   Rely on Communities of Interest

The concept of a Community of Interest (COI) is useful for dealing with the many, disparate ways of grouping users of a national metadata standard. Rather than forcing a single, hierarchical decomposition to fit all such groupings (such as levels of government, functional areas, subject-matter concerns, cross-cutting issues, etc.), COIs provide a flexible, dynamic mechanism for allowing such groups to form as needed. These COIs can then decide for themselves such issues as which subsets of optional metadata elements to implement, which fields can have default values supplied for them, and what encoding schemes and CVs should be adopted or developed for specific metadata elements. Furthermore, if it is decided to create a comprehensive ontology to describe government information (and/or function), COIs can be tasked with creating modular, community-specific ontologies that can be merged together. In order to avoid inconsistency in all of these activities across COIs, some degree of centralized coordination is required. This should include coordination of encoding schemes, CVs and ontologies, to help ensure that these are made consistent among overlapping COIs.

7.1.4   Allow optionally-structured metadata

As argued above, the lack of structure in the Dublin Core approach has a number of shortcomings. In order to retain the simplicity of the DC approach while allowing for much more powerful extensions, we have proposed what we call an "optionally-structured" approach, which gives a metadata implementer the option of using relational, object-oriented, semantic net, or other structures to create far more powerful and flexible metadata representations.   Any such structured implementation should include a structured representation of the unstructured NL-meta.DC+ metadata elements to allow older, unstructured metadata to be retained in the new, structured approach. In order to avoid inconsistency between structured and unstructured representations of the NL-meta.DC+ elements, it may be safest to require that any given implementation use only one or the other of these options.

7.1.5   Expand scope to include systems that create, manage and use metadata

Because any metadata scheme is merely a means to an end, it is vital to expand the scope of Advies Overheid.nl's efforts beyond the definition of a standard to consider the acquisition or development of metadata creation and maintenance systems and the exploration of interface mechanisms for making discovery metadata accessible to search engines and other finding-tools. In addition, this expanded scope should include the question of whether to attempt to develop community-specific or comprehensive ontologies to help users navigate through and search for information and services of interest. There are (at least) two ways of

using such ontologies to help discovery: either they can be used in the "back end" of a metadata system to generate search terms (e.g., keywords) which are then exported to online material so as to be visible to search engines, or they can be made directly accessible by users, who can then browse through them to find relevant terms and topics.

7.1.6 Address semantic interoperability

Semantic interoperability is a worthy goal that motivates much of the W3C's Semantic Web effort and is called for in the EIF (along with organizational and technical interoperability). Simply stated, semantic interoperability is the ability to meaningfully interchange information among different sources and systems. Any difference in interpretation can lead to a lack of semantic interoperability. A costly ($125 million) example of a failure of semantic interoperability was NASA's 1999 Mars Climate Orbiter, which was lost when a force parameter expressed in English units (foot-pounds) was entered into a computer program that was expecting metric units (newtons).

Users searching for online information are perhaps most vulnerable to this problem when they access related information at different sites or derived from different sources (which may be unobvious, especially if the results are collected on a single site). This may occur when attempting to combine or compare information, for example, comparing or aggregating statistics for different geographic areas or different governmental departments whose responsibilities overlap in some way. Although this kind of comparison and combination may not be all that common and may be most frequently conducted by researchers, it may also be performed by the average user, who may be less likely to notice semantic inconsistencies that lead to meaningless or misleading results.

Basic metadata practices can help minimize semantic inconsistency, for example, using standard encodings and CVs (assuming they are used in consistent ways and with the same meanings). But the best guard against such problems may be COIs that are concerned with cross-cutting issues, which involve information that spans multiple sites, since members of a COI may be more likely to notice semantic inconsistencies in their areas of interest.

7.2 Risk Assessment and mitigation

This report has identified a number of risks in Advies Overheid.nl's current approach toward attempting to establish a nationwide standard for discovery metadata. This section organizes and assesses those risks and summarizes what can be done to mitigate them. Section 7.3 proposes a number of specific recommendations for things we believe Advies Overheid.nl can do to further this endeavor.

7.2.1 Lack of information about users' needs or capabilities

As we have noted, any survey of current or projected user needs may be misleading, since the technology and ubiquity of online government information is evolving so rapidly. Nevertheless, the absence of information about user needs poses a significant risk, since it

makes it difficult to predict whether any metadata approach that is adopted will satisfy these unknown needs. Users should at least be engaged as stakeholders in the ongoing process of designing and implementing a metadata scheme, so that their interests are represented. Ideally, surveys should be designed that can be used to track user needs as they evolve over time.

Similarly, the lack of any baseline measurement of the current effectiveness of discovery mechanisms makes it difficult to know whether any metadata scheme that is implemented has a positive effect. A baseline study should be designed and conducted to attempt to fill this void, and instrumentation should be developed to evaluate future improvements.

7.2.2 Lack of input from municipal or provincial governments, ZBOs, water boards, etc.

The full range of government information providers who will be affected by a nationwide metadata standard should be engaged as stakeholders, in accordance with the principle of "No implementation without representation." If this is not done effectively, and even if compliance is strictly voluntary, the proposed standard may be perceived as a central government edict (or unfunded mandate), which may lead to unnecessary resistance. Even if this perception does not arise or does not cause resistance, failure to engage the broad community of potential implementers as joint owners of the standard may marginalize it and reduce its acceptance.

7.2.3 Lack of involvement of records management and archives

Advies Overheid.nl has already engaged key records management and archives stakeholders in its efforts, which has been very helpful and was a key factor in our identifying this risk. However, the distinction between the records management domain and the access domain should be clarified, and a strategy (such as that proposed in Section 7.3.3 below) should be developed to ensure that metadata can be made to flow appropriately between these two domains. If the access domain implements metadata before records management in any given organization, this strategy should provide a migration path to the eventual mechanism shown in Figure 1. Premature reliance on the mechanism in Figure 1 incurs the risk of being unimplementable if a records management function is not in place by the time the access domain needs to generate and manage metadata.

7.2.4 Higher cost than benefit

The fact that many of the information providers we contacted are relatively happy with their current discovery facilities implies that they may be a "hard sell" for a nationwide discovery metadata standard. The key to marketing such a standard is to show that the benefit of adopting it outweighs the cost.

Any metadata standard intended to improve discovery and access runs the dual risks of either being too complex and burdensome to be used or providing insufficient improvement in discovery and access to be worth the effort of implementing it. In fact,

these two risks are not mutually exclusive: a standard can be both overly burdensome and functionally insufficient at the same time.

While the simplest versions of Dublin Core (employing the minimum number of metadata elements) do not seem to place an undue burden on information providers, neither do they provide great leverage for discovery of online resources, many of which, as discussed above, may be inadequately described by simple attributes such as title, author, and publication date. Most of the more elaborate standards derived from Dublin Core extend it in the records management domain, adding elements used for such functions as rights management, disposal, or access control, which are of marginal relevance for discovery.

The most useful additions for discovery are likely to be richer sets of search terms used in a subject field, as produced by taxonomies, thesauri, or ontologies. However, considerable effort may be required to create such terminologies, and this must be balanced against the expected improvement in discovery. The best way to achieve this balance appears to be to engage COIs in the modular development of such terminologies, so that the effort required is divided among many participants, while the benefit is multiplied across many potential users.

7.2.5   Longevity of data and metadata

One final risk is that online information--or the discovery metadata that describes it--may become unusable over time, thereby undermining the investment made in collecting it and making it available. As discussed in Section 4.2.6 above, all digital information, whether it is considered data or metadata, is susceptible to becoming inaccessible or unusable as the programs that render it or the computers on which those programs run become obsolete. Various solutions to this digital longevity problem have been proposed, but implementing them is probably beyond the scope of the Advies Overheid.nl metadata effort. Nevertheless, we have included preservation metadata (i.e., metadata designed to facilitate preservation) in the proposed standard, to support whatever long-term preservation strategy is employed for online information.

The longevity of the metadata values poses a separate risk. In order to implement the metadata standard, one or more metadata management systems must ultimately be adopted or developed, and these systems must employ some appropriate preservation strategy to ensure the longevity of the metadata they manage. Since most metadata values are relatively easy to render without running complex application software, a simple migration scheme is likely to be an effective way of preserving them, though formal description is also a possibility.

The revision of metadata or of the metadata fields (elements) themselves, should not pose a major risk if they are kept separate from the online resources they describe. On the other hand, if metadata are embedded in the resources themselves, then changing the metadata requires modifying these resources, which does present a risk, unless the update process can be fully (and safely) automated.

Finally, the semantics of metadata must be preserved as the meanings of terms change over time. As discussed in Section 4.2.6, this may require using mechanisms such as thesauri or ontologies to map old terms into new ones with equivalent meanings.

7.3   Recommendations to Advies Overheid.nl

This section formulates a number of specific recommendations for how Advies Overheid.nl should proceed in order to promote the use of a nationwide discovery metadata standard.

7.3.1   Engage a broad stakeholder community via an online working group

In order to increase the involvement of all of the groups that may be affected by or interested in a national discovery metadata standard, Advies Overheid.nl should make a concerted effort to include such groups in an ongoing discussion focused on the design and implementation of the standard. This may be done by means of Advies Overheid.nl's existing online working group, assuming the scope of this group can be expanded appropriately. This working group should include, at least:

- Users (e.g., citizens and companies using government information)

- Government information providers at all levels

- Relevant commercial data producers (such as publishers)

- Non-governmental websites offering government information

- Records management, archives, libraries and other metadata creators

- Commercial vendors of search-engines, records management systems, metadata or database systems, ontology management technology, etc.

- Academic institutions researching web use, discovery, and eGovernment

Engating stakeholders in the design and implementation process would serve several purposes at once. It would help avoid the perception (and reality) of imposing a de facto unfunded mandate, help ensure that all relevant needs and interests are taken into account, and create a broad set of joint owners who will be more likely to use the results of their joint efforts. This is seen as a crucial part of the marketing effort that Advies Overheid.nl must undertake to promote the use of the NL-meta standard.

7.3.2   Support the adoption and use of the metadata standard

If it is to be widely adopted and utilized, the standard must be supported by Advies Overheid.nl in a number of ways. The exact form and extent of this support may vary depending on Advies Overheid.nl's resources and priorities, but it should ideally involve the following activities:

- Facilitate creation and coordination of COIs, encoding schemes, and CVs; act as (or establish) a clearinghouse for COIs, encoding schemes, and CVs

- Support tool evaluation, selection and use; for example, develop tool evaluation checklists

- Provide conformance verification and metadata quality control; offered as an online service

- Develop and provide guidelines and training
    - For the creation, management, and use of discovery metadata
    - For the role of Metadata Manager in information providerorganizations

If Advies Overheid.nl cannot provide all of these functions, it should at least develop a strategy for how and where they can be provided.

7.3.3 Promote a records management strategy for the access domain

It may be useful for Advies Overheid.nl to develop and recommend a concrete strategy for how an information provider organization can integrate its access and records management domains. Such a strategy should be refined with input from records management experts, but a first approximation might consist of something like the following:

a) Designate and empower a Metadata Manager in each organization

b) Define needed access metadata, CVs, etc.

c) If these already exist in the records management domain, create one-way pipeline from records management to access, performing any needed translation, repackaging or restructuring

d) If not, discuss the shortfall with records management, and decide where to supply the missing metadata

e) If necessary, create any remaining missing access-specific metadata that may not be of interest to records management

f) Use CVs wherever possible, to enhance access:

- Find existing CVs from appropriate COIs via the COI clearinghouse

- Coordinate the use of CVs with records management

- Use international standards where applicable

g) Where no appropriate CV exists, consider forming a partnership between records management and an appropriate COI to create a new CV

These steps should help an organization establish appropriate ties between records management and access, while allowing for the fact that a true digital records management function may not yet be implemented in many organizations.

7.3.4 Conduct follow-on studies

A number of additional studies should ideally be performed to reduce the risk of the discovery metadata effort. These would refine and validate the insights and recommendations in this report, based on further research and analysis. The most important such studies appear to be:

- Analyze user needs for discovery and access

> Identify current and projected needs of citizens and others for finding and accessing government information in the Netherlands.

- Measure baseline discovery effectiveness (based on user needs) and put in place ongoing instrumentation to measure improvement

  > Measure not just the effectiveness of search queries at finding relevant "hits" but the overall effectiveness of the process of discovering appropriate government information resources and services.

  > As part of the baseline study, create instruments that can make additional measurements in the future, to see if interventions have improved discovery.

- Analyze the relevance of specific encoding schemes and CVs for Dutch metadata. Many different encoding schemes and CVs are used throughout the world and within the Netherlands

  - Look at existing national IT standards and their relevance for the Dutch government

  - Investigate encoding schemes and CVs that are used within the Dutch government

- Analyze the possibility of developing a multi-level, modular ontology

  > The topmost level of such an ontology can be developed by central government, while lower level modules should be developed by appropriate COIs, especially if tools (such as Topic Maps) can be found to automate the merging of these different resulting ontologies into one.

  > The alternatives of using such ontologies to generate and export discovery metadata (search terms) or allowing users to browse the ontology directly to find what they are seeking should be investigated.

  > The maturity of ontology creation tools (such as Ontopia) should be evaluated.

- Develop process models for creating, maintaining and improving discovery metadata

  > Develop concrete processes that government information provider organizations can use to help create, maintain and improve discovery metadata over time

- Develop desiderata for discovery-systems and the implied interfaces and services that such systems require of metadata systems

  > Create guidelines for designers and implementers of such finding-tools and access systems, in order to improve the use of discovery metadata by such systems.

- Develop desiderata for systems to semi-automatically generate, maintain, derive and transform metadata

  > Create guidelines for designers and implementers of content management, web-publication, and related systems, to enable these systems to create and maintain discovery metadata and derive and transform metadata among systems and across domains.

Studies such as these would provide valuable input to organizations that create government information, as well as to designers and vendors of metadata management systems and finding-tools. They would also help define appropriate pilot projects (such as those recommended in the next section) that would further increase the likelihood of success of implementing a nationwide discovery metadata standard.

7.3.5 Perform pilot projects to try out and evaluate the approach

In addition to engaging stakeholders via an online forum and conducting studies such as the ones recommended above, we suggest conducting small pilot projects to investigate the full range of issues involved in implementing, managing, and using discovery metadata. The current set of 20 metadata pilot projects that Advies Overheid.nl is conducting may form the start of this effort. However, in order to learn as much as possible from such efforts, they should ideally be performed only after a baseline study of the effectiveness of discovery has been conducted and instrumentation for measuring improvements in discovery has been put in place.

A logical sequence of pilot projects aimed at trying out the proposed standard might be:

- Implement the simple NL-meta.DC+ metadata set

- Try out one or more structured NL-meta.Extended approaches

- Explore the use of CVs and ontologies

- Explore the development of metadata to model processes, in order to aid the discovery of online transactions, workflows or e-Services

Among other things, these pilots should focus on the use of discovery metadata by finding-tools such as search engines or ontology browsers.

7.3.6 Treat the Handbook as evolutionary

Advies Overheid.nl should establish a review and feedback process to revise and evolve both the standard and its associated Handbook to produce a new version within a year or so after its initial release.

The recommendations in this report are intended to help the Dutch government create and implement a national standard for discovery metadata and promote its adoption and use by government information providers. Developing such a standard is a key step toward the admirable goal of improving the transparency of government information and action for citizens and others in the new digital age. However, the standard itself is just a first step toward this goal; if it is to be effective, it must be embedded in a larger context that includes the many implementation issues discussed in this report.

LITERATURE USED

Literature used

AGLS, National Archives of Australia, AGLS Metadata Element Set Part 1: Reference Description, version 1.3, 2002.

Alkema, Ynze and Kruizinga, Jelle, Definitiestudie Datamodel Toegankelijke Overheidsinformatie voor het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, versie 1.0, September 2003.

Barhma, S., E-government Unit, State Services Commission, New Zealand Government Implementation of a DC-based Standard – Lessons Learned, Future Issues, March 18, 2003.

Beckett, David, IAFA Templates in use as Internet Metadata, http://www.w3.org/Conferences/WWW4/Papers/52/, accessed 9 June 2004

Belastingdienst, A closed-loop search service with Topic Maps and XML, XML Europe 2004 Amsterdam

Booth, Keitha, Mediating for metadatastandards: competing demands of e-government, archivists and librarians for resource description in New Zealand, 2001.

Bridges, Environmental Information Access in Minnesota, Foundations Project Usability Testing: Dublin Core Metadata and Controlled Vocabulary Study.

Bridges, Environmental Information Access in Minnesota, Foundations Project Usability Testing: Bridges Search Interface Study.

Brodie, N., Government of Canada Metadata Framework, Presentation to Canadian Metadata Forum, September 20, 2003

Bussel, G.J. van, Horsman, P.J., and Waalwijk, J., Remano 2004, March 30, 2004

Cabinet Office, Office of the e-Envoy, e-Government Metadata Standard, Version 3.0, April 29, 2004.

Cammarata, S., D. H. Shane, and P. Ram, IID: An Intelligent Information Dictionary for Managing Semantic Metadata, The RAND Corporation, R-3856-DARPA, 1991.

CEN Workshop Agreement, CWA 14855, Dublin Core Application Profiles, November 2003

CEN Workshop Agreement, CWA 13988, Guidance information for the use of Dublin Core in Europe, April 2003

CEN Workshop Agreement, CWA 14859, Guidance on the use of metadata in eGovernment, November 2003

CEN Workshop Agreement, CWA 14857, Mapping between Dublin Core and ISO 19115, "Geographic Information – Metadata", November 2003

CEN Workshop Agreement, CWA 14860, Dublin Core eGovernment Application Profiles, November 2003

CEN Workshop Agreement, CWA 14856, Guidance material for mapping between Dublin Core and ISO in the Geographic Information Domain, November 2003

Commissie van de Europese Gemeenschappen, Mededeling van de Commissie, eEurope: Een informatiemaatschapij voor iedereen, COM(2002) 263 definitief, May 28, 2002, http://europa.eu.int/comm/lisbon_strategy/pdf/263_NL.pdf, Accessed June 2004

Cumming, Maewyn; Tomatoes are not the only fruit; a rough guide to taxonomies, thesauri, ontologies and the like, December 2002;

http://www.govtalk.gov.uk/schemasstandards/gcl_document.asp?docnum=681

DCMI Usage Board; DCMI Metadata Terms; http://dublincore.org/documents/dcmi-terms, November 2003

DARE, Stichting Surf, Use of Dublin Core, version 1.0, October 2003.

Dekkers, M., The Dublin Core Metadata Initiative, Current Status and Future Developments, 2001

Dekkers, M., AMI consult S.àr.l, The European Approach to Metadata, final report, October 31 2001.

EAD (Encoded Archival Description), http://www.loc.gov/ead/ead2002.html, Accessed July 8, 2004

e-Grant Publishing, Digitale overheid gaat volgende fase in: Provincieloket Overijssel is online, persbericht, 11 may 2004, http://presscenter.nl/Message/default.asp?NewsID=1715&CatID=110, Accessed June 8, 2004-07-01

e-Grant, Handleiding overheidsloket, 2004

e-Provincies, Plan van Aanpak Informatiearchitectuur 2004, versie 2.0 March 11, 2004.

European Commission, Richtlijn van het Europese Parlement en de Raad betreffende diensten op de interne markt, COM(2004)2, February 25, 2004, http://europa.eu.int/prelex/detail_dossier_real.cfm?CL=nl&DosId=188810, Accessed June 2004.

European Commission, Decision of the European Parliament and of the Council of 12 July 1999 on a series of guidelines, including the identification of projects of common interest, for trans-European networks for the electronic interchange of data between administrations (IDA) 1719/1999/ECOfficial Journal L 203 , P. 0001 – 0008, August, 3, 1999.

European Commission, IDA, European Interoperability framework, for pan-European eGovernment services. IDA working document. v4.2, January 2004

European Commission, MIReG activity (Managing Information Resources in eGovernment), website EU, http://europa.eu.int/ISPO/ida/jsps/index.jsp?fuseAction=home, Accessed June 2004

Garshol, Lars Marius; Metadata? Thesauri? Taxonomies? Topic Maps! *Making sense of it all* Ontopia; www.ontopia.net Accessed 2 June 2004

Garshol, Lars Marius; Living with Topic Maps and RDF Ontopia; www.ontopia.net Accessed 2 June 2004

Geodan, GeoKey: de sleutel tot uw (geo-)informatie website, http://www.geodan.nl/nl/product/geokey/gk_alg.htm, Accessed June 2004

Geodan, National Clearing House Geo Informatie (NCGI), website http://www.geodan.nl/nl/project/ncgi/index.htm, Accessed June 2004

GILS, www.gils.net

Government of Canada, Ad Hoc Committee of Federal Metadata Experts Metadata Action Team, Council of Federal Libraries, Metadata Implementation Guide for Web Resources, 2nd edition, May 2003

Heery, Rachel; Patel, Manjula,; Application profiles: mixing and matching metadata schemas, 2000, Ariadne Issue 25

Heijst, G. van, Voorstel voor een geïntegreerde taxonomie voor Regering.nl, Postbus51.nl en Overheid.nl , Oryon Rapport, 2004.

Hillmann, Diane; Using Dublin Core; http://dublincore.org/documents/usageguide/; August 2003

ICTU, Handleiding Consolidatie (draft), June 23, 2004

IOG-werkgroep "Conceptueel Model", Netland Schetsboek Provinciaal Referentiemodel, November 2002.

IPSMS, The Irish Public Service Standard – User Guide, version 1.1, June 2002

ISO, ISO 23081/TC, Technical Specification, first edition 2004-05-01

ISO, ISO 15836-DC

Itafit, De kunst van het vinden, April 2002

Kennisnet, De internetorganisatie voor het onderwijs, website http://www.kennisnet.nl/, Accessed June 2004.

Kennisnet, Inrichting, Eigendom, onderhoud en beheer voor metadata v1.0, maart 2004

Koninklijke Bibliotheek, Dublin Core Simple Format, http://www.kb.nl/donor/rapporten/Dcsimpleformat.html, October 1998, Accessed May 6, 2004

LogicaCMG, Local Government Association of Tasmania (LGAT), Metadata Tools and Analysis, December 2003

Machine-Readable Bibliographic Information Committee (MARBI), The MARC 21 Formats: Background and Principles, Revised November 1996, http://www.loc.gov/marc/96principl.html, accessed 10 June 2004

Miller, Eric, An introduction to the Resource Description Framework, D-Lib Magazine, May 1998, ISSN 1082-9873

Ministerie van Binnenlandse Zaken 'ICT en de Overheid', website http://www.minbzk.nl/ict_en_de_overheid, Accessed June 2004

Ministerie van Binnenlandse Zaken, Notitie 'Op weg naar de Elektronische Overheid', July 14, 2002, http://www.minbzk.nl/contents/pages/9116/elektronische_overheid.pdf, Accessed June 2004,

Ministerie van Binnenlandse Zaken, Nota 'Naar optimale beschikbaarheid van Overheidsinformatie', April 2002, http://www.minbzk.nl/grondwet_en/openbaarheid/publicaties/nota_naar_een, Accessed June 2004

Ministerie van Onderwijs, Cultuur en Wetenschap, De bron ontsloten, Haalbaarheidsonderzoek auteurs-gegenereerde metadata, August 1, 2003.

Minnesota Department of Administration, Office of Technology, Standard for Web Metadata, Standards: Information Resource Management Series, Version #: 21 no. 1, June 24, 2002

Minnesota Planning, Foundations Project, Information Needs Assessment: Summary of Survey Results, April 1998

Mlecoch, C. and Quam, E. State of Minnesota, Best Practice Guidelines for Web Metadata, April 2002

NZGLS Maintenance Agency, NZGLS Usage Guide Version 2.1, April 2004

Office of the e-Envoy, Minutes of Metadata Working Group meeting, March 11, 2004, http://www.govtalk.gov.uk/documents/MWG%20minutes%202004-03-11.pdf, Accessed June 2004,

Office of the e-Envoy, eGIF, e-Government Interoperability Framework, version 6.0, April 30, 2004

Office of the e-Envoy, Minutes of the Interoperability Working Group meeting, http://www.govtalk.gov.uk/documents/interopmins2004-03-17.pdf , Accessed June 2004, March 17, 2004

Office of the e-Envoy, e-GMS for websites, version 2.0, May 2003

Office of the e-Envoy, Policy Framework for a mixed Economy in the Supply of e-Government Services, Implementation Guidelines, V.1, December 2003

Pepper, Steve; The TAO of Topic Maps, Ontopia, April 2002, www.ontopia.net/topicmaps/materials/tao.html Accessed May 2004

Provincie Flevoland and Ordina, Provinciaal Referentiemodel, Presentation e-Provincies, April 2004

Quam, E., Minnesota Department of Natural Resources, Minnesota Metadata Guidelines for Dublin Core Metadata, Training Manual, September 2003.

Ricci, Christian, Developing and Creatively leveraging Hierarchical Metadata and Taxonomy, 2002-2004, http://www.boxesandarrows.com/archives/developing_and_leveraging hierarchical_metadata_and_taxonomy.php, Accessed June 14, 2004

Rijksvoorlichtingsdienst, Ministerie van Algemene Zaken, Afdeling Communicatiebeleid, Notitie werkgroep metadata project Activiteitenindex, March 14, 2004

Rijksvoorlichtingsdienst, RVD Taxonomie, 2004.

Rothenberg, J., "Ensuring the Longevity of Digital Documents",Scientific American, January 1995 (Vol. 272, Number 1), pp. 42-7;(International Edition pp. 24-29).

Rothenberg, J., "Metadata to Support Data Quality and Longevity", First IEEE Metadata Conference, April 16-18, 1996, NOAA Auditorium,NOAA Complex; Silver Spring, MD (published only on World Wide Web: http://computer.org/conferen/meta96/rothenberg_paper/ieee.data-quality.html)

Rothenberg, J., Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation, A Report to the Council on Library & Information Resources (CLIR), January 1999, ISBN 1-887334-63-7.

Rothenberg, Jeff, and Tora Bikson, Digital Preservation: Carrying Authentic, Understandable and Usable Digital Records Through Time, Report to the DutchNational Archives and Ministry of the Interior by RAND-Europe, The Hague, 1999.

Rusch-Feja, Diann, Metadata: Standards for retrieving WWW documents (and other digitized and non-digitized resources), http://www.stsci.edu/stsci/meetings/lisa3/ruschfejad.html, ASP conference series, Vol 153, 1998, Accessed April 30, 2004

Smith, M. APLAWS Guidelines for Metadata Standards Implementation for Local Government Websites, 2002

Swol, R.W. van and Wisse, E., NEONET - Netherlands Earth Observation Network, September 12, 2000

Taylor, Chris; An introduction to metadata; Cybrary, University of Queensland, July 2003; www.library.uq.edu.au/iad/oai; accessed 2 June 2004

The Victorian Government, Multimedia Victoria AGLS Victoria: Metadata Implementation Manual Version 1.2.1, A guide to the implementation and management of AGLS metadata in Victorian Government departments and agencies, November 2002.

Treasury Board of Canada, TBITS 39.1: Treasury Board Information Management Standard, Part 1: Government On-Line Metadata Standard, November 8, 2001

Treasury Board of Canada, TBITS 39.2: Treasury Board Information Management Standard Part 2: Controlled Vocabulary Standard, November 8, 2001

Tweede Kamer der Staten Generaal, Wet elektronisch bestuurlijk verkeer, kamerstuk 28483, July 22, 2002

Tweede Kamer der Staten Generaal, Wet Kenbaarheid Publiekrechtelijke Beperkingen Onroerende Zaken (PUBER, 28.218), 2002, http://www.referendumwet.nl/sites/referendumwet/contents/i000683/KST59452.pdf, Accessed June 2004

Tweede Kamer der Staten Generaal, Memorie van Toelichting bij de 'Wet Kenbaarheid Publiekrechtelijke Beperkingen Onroerende Zaken (PUBER)', 2002.

Tweede Kamer der Staten Generaal , 'Aktieprogramma Elektronische overheid', Tweede Kamer vergaderjaar 1998-99, 26387, nr. 1, http://www.recht.nl/doc/kst26387-1.pdf?RNLSESSION=0a33ebd58f35a12e7be34d4885948361, Accessed June 2004

Vertegenwoordigers BVE-sector, Metadatastandaard voor educatieve content: het toepassingsprofiel in de praktijk v1.0, maart 2004

Vertegenwoordigers BVE-sector, Metadatastandaard voor educatieve content: het toepassingsprofiel in detail v1.0, maart 2004

Vertegenwoordigers BVE-sector, Metadatastandaard voor educatieve content: Een praktisch toepassingsprofiel v1.0, maart 2004

Vlietinck, Erik, Hou het beheersbaar, Metadata: risicovol instrument voor betere efficiency, Computable Technologie nr.18, May 2, 2003

Wilson, Andrew and Aagaard, Palle, Survey of government implementation of DC, March 2003.

APPENDICES

Appendix A Survey and questionnaire results

In developing a metadata standard for the *access domain* in the Netherlands, we wanted to get information from as many government organizations as possible. Because of the number of organizations and the timeframe of the project it was clear that it would be impossible to speak to all the municipalities, waterboards, provinces and ZBOs (Independent Governmental Organizations). A distinction between three types of organizations was therefore made:

1. Organizations that actively use metadata
2. Organizations familiar with metadata
3. Organizations unfamiliar with metadata

We decided to conduct in-depth interviews with the first group of these organizations, which turned out to be mainly in central government. The second and third groups were targeted through an online survey. Face-to-face interviews with central government departments were conducted with the help of a questionnaire, whereas in order to obtain input from municipalities, waterboards, provinces and ZBOs, a survey was sent out in the hope of obtaining a representative sample. Some of the survey questions were similar to those in the interview questionnaire, in order to allow comparisons between the different types of government organizations. Based on the replies to the survey, some of the responding organizations were contacted by telephone or in person for follow-up interviews.

This appendix first presents our conclusions based on the combined results of these efforts. It then describes the methods that were employed and the detailed results of the survey and interviews. The questions in the survey are shown in Appendix B, whereas the interview questionnaire is shown in Appendix C.

Conclusions

Meaning and importance of metadata
There is uncertainty about the meaning of the term metadata, especially at the municipal level but also within departments at the ministry level. In those organizations in which the concept of metadata is understood, it is seen as providing added value, but most organizations do not use metadata (or at least are not aware of using it). The departments

responsible for websites are, in general, not familiar with the Dublin Core metadata standard.

Controlled vocabularies

Controlled vocabularies are hardly ever used on websites, and certainly not in the assignment of metadata, although many organizations do use product catalogues (for example, municipalities use VIND[1]). There seems to be a need for central assistance in the area of controlled vocabularies.

Information retrieval

Although some organizations are looking for a good solution to their search problems and are interested in using the results of our project, many organizations are satisfied with their search capabilities and see no need for change. For the ministries, offering the capability for users to retrieve information by navigating through structured menus is considered to be almost as important as offering the capability of retrieving information by searching for it. The importance of archiving content of websites is hardly acknowledged anywhere.

Co-operation between departments responsible for websites and record management

The cooperation between the departments responsible for websites and for digital records management is in general inadequate, or not present. Only one good example of cooperation between such departments was found. At most ministries there are no agreements about the definitions of metadata used for these two purposes. None of the ministries manage the coordination of metadata between their records management and their websites, and only one ministry has a dedicated metadata manager. At other ministries, a cooperative body of departments performs the management of metadata.

External co-operation

There is little or no cooperation between different organizations in the area of development and maintenance of their websites. The Advies Overheid.nl facilitates external cooperation for local government organizations, but implementation of the resulting ideas is substantially different within different organizations.

Methods

Survey method

The survey consisted of 20 questions that sought information about three main topics: Information offered, search capability, and interoperability. Questions on the *information offered* concerned the form in which information is presented and the information and services that are offered at websites. The questions on *search capability* focused on the creation, use, and maintenance of metadata and the existence and maintenance of controlled vocabularies. In addition, questions were asked about any limitations that organizations place on the publication of information on their websites or on the ability of users to search for such information. The questions on *interoperability* concentrated on the

---

[1] Vraaggerichte INteractieve Dienstencatalogus (Demand-oriented interactive services catalogue)

interactions between multiple websites within organizations and with websites of other organizations. A copy of the survey can be found in Appendix B.

Survey reponse

The Internet based survey was announced in a newsletter of Advies Overheid.nl (Year 5, no. 4. 26 May 2004), which included a link to the survey. 30 of the ZBOs were also approached by letter. When the response turned out to be limited, an email message was sent to an additional list of approximately 165 contacts, which was obtained from Advies Overheid.nl through the Overheidsmonitor of Overheid.nl.

In total, about 1200 newsletters were sent out, as well as the 30 letters to ZBOs and 165 messages to municipalities and waterboards. 44 surveys were returned. Of these, one was excluded because it represented a commercial organization. Two indicated that they did not know what metadata meant and that they had therefore not filled in the survey accurately. One filled in the survey incorrectly, and two were replies by ministries, which were also omitted from the analysis, as to avoid double counting (they had also filled in a questionnaire). This left a total of 38 surveys included in the analysis.

The 38 usable surveys were in response to 1400 requests, which is a response rate of 3%. The total number of surveys received is not enough to represent a useful sample, so caution is required when interpreting our results. Furthermore, the returned surveys probably do not constitute a representative sample, because only organizations familiar with the concept of metadata filled in the survey correctly. This was confirmed when some organizations indicated that they were not able to fill in the survey, since they did not understand its terminology.

Interviews (questionnaire) method

A questionnaire formed the basis for interviews with representatives of the ministries. In addition, some interviews were held with other government organizations (e.g., municipalities). A full list of organizations that were interviewed is:

Name of organization	Type of organization
Ministerie van Landbouw, Natuur en Voedselkwaliteit	Ministry
Ministerie van Volksgezondheid, Welzijn en Sport	Ministry
Ministerie van Algemene Zaken, Rijksvoorlichtingsdienst	Ministry
Ministerie van Onderwijs, Cultuur en Wetenschap	Ministry
Ministerie van Buitenlandse Zaken	Ministry
Ministerie van Sociale Zaken en Werkgelegenheid	Ministry
Ministerie van Economische Zaken	Ministry
Ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieu	Ministry
Ministerie van Verkeer en Waterstaat	Ministry
Belastingdienst (Tax office)	Other
National archives	Other
Provincie Noord-Brabant	Province
Gemeente Zwijndrecht	Municipality
Archives Den Haag	Municipality
Gemeente Den Haag	Municipality
E-Grant	Other

The questionnaire consists of 5 sections (Information offered, Structure of information, Permanency of information, Policy and implementations, and Interoperability). The section on *information offered* consists of questions regarding the form in which

information is offered as well as what information and services are offered on websites. Questions are posed regarding possible distinctions among users as well as the number of users (per type) of the website. Furthermore, inquiries are made into the accessibility and search capability of the website. Further questions are: are there any reasons that the possible expansion of the website is obstructed, and do laws such as WBP[2], WKPB[3] and WOB[4] prevent the publication of information? The questions on *structure of information* are partly the same as the questions posed in the section *search capability* of the survey. They focus on the creation, use, and maintenance of metadata, the use of any metadata standards, and the existence and maintenance of controlled vocabularies. Questions on limitations in the ability to publish or search for information on websites were also posed. Additional questions related to the existence of default values and formats, as well as the consistency and integrity of metadata. Finally, some questions regarding harmonization between metadata for a document management system and the website were posed. In the section *Permanency of information,* questions regarding the validity of information on the website and strategies to check this validity were posed. The section *Policy and implementation* aims at getting information on the policies that exist within organizations regarding the publishing of information on their websites, the responsibilities of different departments, the coupling of metadata to data, ownership of information on websites, and agreements with other organizations on publishing information on websites. As in the surveys, the questions in the section *Interoperability* focus on the existence and type of interaction between websites within organizations and with other organizations. A copy of the questionnaire can be found in Appendix C.

Results of the survey and questionnaire

The results are discussed in three parts. First, a comparison is made, where possible, between ministries and "other" government organizations. The category "other" is very broad and contains municipalities, ZBOs, provinces, and waterboards. In order to make this comparison, the questionnaire results for the category "other" were added to the results of the survey; data from similar questions in the survey and the interview questionnaire are combined. The results section ends with a description of the results of the interviews (based on the questionnaires and other comments). This description is presented for both the ministries and the "other" category.

Comparison Ministries and other governmental organizations
The comparison is based on 38 surveys, 9 questionnaires from ministries and 7 questionnaires from "other" government organizations. The graphs below therefore represent responses from 54 organizations.

---

[2] Wet Bescherming Persoonsgegevens (Personal Data Protection Act)

[3] Law on provision of information on public restrictions

concerning real estate

[4] Wet Openbaar Bestuur (Transparant Government Act)

The following figure shows the results of the question: which types of information are used on your website.
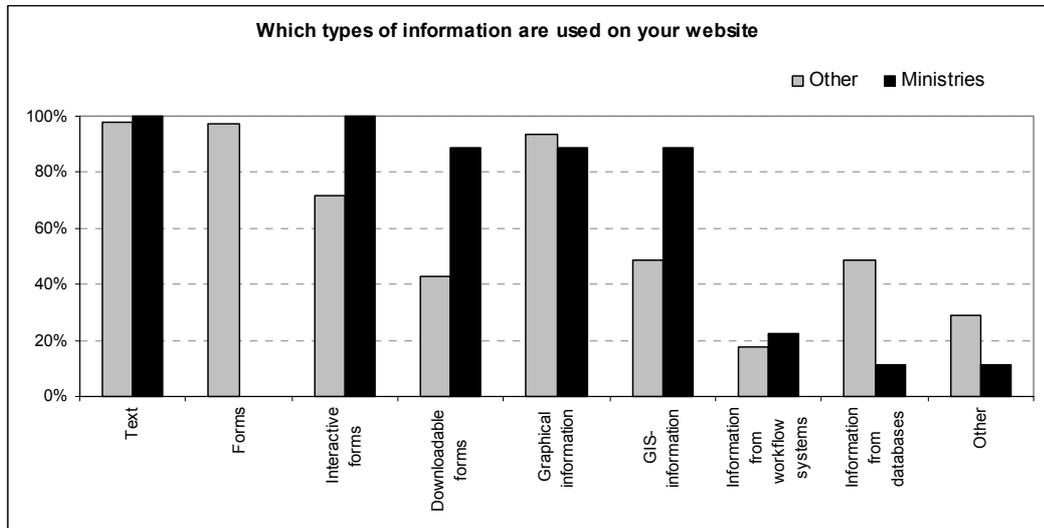
In the survey, there was only one category 'Forms', while in the questionnaire a distinction was made between 'Interactive Forms' and 'Downloadable Forms'. In the above figure, the grey bar under Forms is therefore the results of the survey, whereas the grey bars in the columns Interactive Forms and Downloadable Forms are the results of the questionnaires of the "other" government organizations.

The above figure shows that all websites have text and forms available. Also graphical information is widely available. GIS information can be found more often on ministry websites. Information from workflow systems and databases are not often available.

The following figure shows the type of information offered at government websites. The category "Raads-, Staten-, and bestuursinformatie" (Council, Provincial Council, and administration information) is only defined at the local and regional level and is therefore included in the "other" government organizations category.

What information does your website offer

☐ Other ■ Ministries
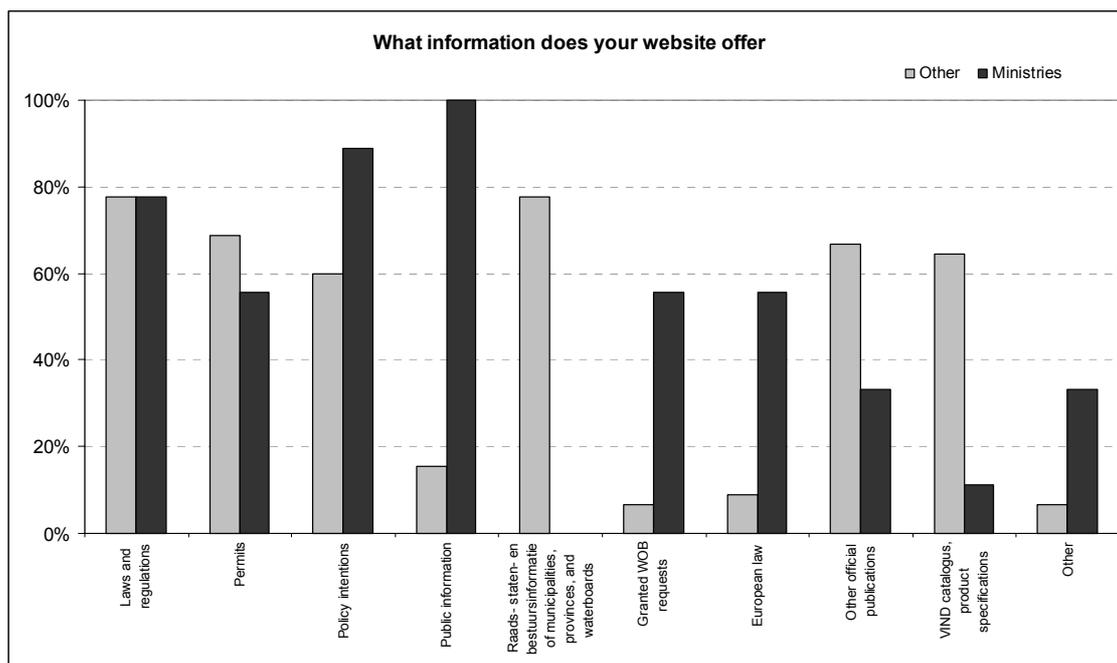
This figure shows that there are substantial differences between ministry websites and "other" websites. Nearly 80% of both ministry and other websites offer information on laws and regulations. Policy intentions and public information are more often found on ministry websites that on other websites. The VIND[5] catalogue and product information is typical of the information found on other websites.

Ministries use metadata more often than other government organizations. Only 11% of ministries do not use metadata, compared to 40% of other government organizations. The results for other government organizations are skewed, since it is probable that only those organizations familiar with the concept of metadata filled in the survey; it is therefore likely that more than 40% of "other" government organizations do not use metadata.

The existence of controlled vocabularies also seems to be more common in ministries than in other government organizations. Only 22% of ministries do not have controlled vocabularies, compared to 49% of other government organizations.

33% of other government organizations do not know if their website is fully searchable. 49% of websites of other government organizations are fully searchable, while 78% of ministry websites are fully searchable. One of the primary reasons that a website may not be fully searchable is if it uses databases, which search engines cannot search.

The following figure shows the categories on which one can search websites. Theme and keyword are the most common options, for both the ministries and other government organization websites. Language and target groups seem to be more important search possibilities for ministries than for other government organizations.

---

[5] Vraaggerichte INteractieve Dienstencatalogus (Demand-oriented interactive services catalogue)

On what categories can be searched
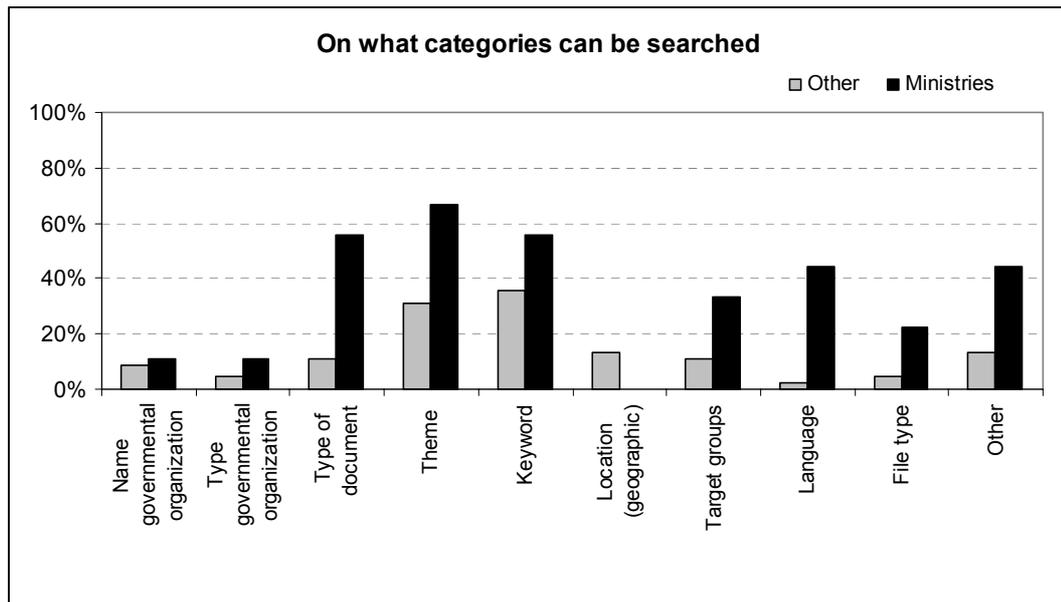
□ Other   ■ Ministries

A majority of ministries have more than one website (78%), while only 42% of other government organizations have multiple websites. Ministries use centralized management (33%) or the use of common standards (33%) as a means to achieve internal coherence (non-response 34%). Most other government organizations do not have a mechanism to assure internal coherence. Non-response in this category was much larger (more than 50%).

A majority of both ministries and other government organizations mention that there has not been consultation between organizations or departments to discuss the use of a metadata standard to improve the possibilities of combining and aggregating information from multiple websites.

A qualitative look at accessibility of government information online

In what follows, we present the most interesting results of our interviews with the 9 ministries and 7 "other" government organizations.

Ministries
The services provided are restricted to applying for permits on a limited number of websites. Few websites distinguish different user groups (e.g., citizens or companies). Expansion of the content of websites is limited mainly by organizational restrictions, although financial restrictions are also mentioned as a limiting factor.

Eight out of nine ministry websites present only current information; in general the only historical information is press information (available for up to several years). The information departments look after the content on the websites, where in practice both centralized and decentralized editors take care of this content. In general if the editors are centralized, the use and correctness of the metadata are perceived to be improved.

Six out of nine ministries own all the information presented on their website; three ministries provide information from organizations other than themselves. No comprehensive protocols or agreements on the use of information from other organizations were identified during the interviews.

Content Management Systems of a variety of suppliers manage the ministry websites. The use of websites is not measured in great detail. The central website of each ministry is linked to other ministry websites and to a lesser extent also to websites of independent organizations. If the management (and the layout) of the related websites is centralized, the user is able to change between different websites without notice. If the management of all websites is decentralized, many interoperability problems seem to occur e.g., with layout, metadata, and links.

All information on the main websites of the ministries is freely accessible. Restricted access to websites is found only on websites for special interest groups. Privacy sensitive information is not given on the websites, while the information designated by WOB is, or will be, published in the near feature. One ministry paid special attention to avoiding having private sensitive information on the website, by blocking the names of public servants within departments.

Information on websites can be retrieved through a menu structure (navigation) and through the use of a search engine. Free-text searching is possible on all websites, while searching for a list of subjects or for text combined with a metadata category is possible on 7 out of 9 websites.

A great variation is found at the websites of the ministries in the use of metadata. 8 out of 9 are using metadata fields, defined within their own organizations. International standards are not followed; each organization has implemented its own metadata set. Within almost all organizations the metadata for websites are developed for the website only, without coordinating these with departments responsible for document or records management. This practice results in "islands of information" within the ministries: the metadata for websites and records management are isolated from each other, and the definitions of the metadata and the controlled vocabularies (if present) have different meanings.

For ministries there is an important initiative, the Activity Index, in which a coordinated effort is being made to define a minimum set of metadata for websites. This initiative is limited to ministries and is only concerned with information relevant to a defined set of policies for the duration of the mandate of the coalition government.

At 4 out of 9 ministries the metadata are managed in a database, while at 1 ministry the metadata are embedded in information resources using HTML tags, and at 3 ministries the metadata are in the information but also in a database (a "hybrid" approach). One ministry is not using metadata for web content at all.

The implementation of document management systems and/or records management systems within the ministries is very limited and is only in the start-up phase. The ministries have formed a working group, the InterLAB, to coordinate the use of records management systems. The InterLAB is still working on developing metadata for records management. 5 out of 9 ministries have not chosen a standard for the preservation of

textual information, while the other 4 ministries are using both XML and Adobe PDF formats.

Other organizations

The interviewees in this group of organizations mentioned few services available on their websites. One website allows users to apply for some permits. Others have a web shop and cater to online payments. Most websites target several user groups; the most important distinction (made by 6 out of 7 organizations) is between citizens and companies. Only one interviewee does not distinguish between user groups. Most interviewees cannot give an estimate of the number of users of their websites. For most websites it is not possible to measure the distinction between the use of the website to obtain information and to provide services.

Most organizations interviewed have a content management tool in use. Two of them use Tridion, one Discoveryserver, and two "Smartsite".

Two of the organizations interviewed mention that not all the information on their website is accessible by all visitors due to the use of databases. The laws WBP, WKPB, and WOB do not prevent the publication of information. One organization mentions the existence of a separate log-in for privacy sensitive information and an active policy of publicizing all relevant documents and information.

Two organizations mention the use of an international standard, Dublin Core, in their search engine. Both acknowledge that the investment in time and/or money is small in choosing and using that standard. One organization examined Dublin Core but did not find it useful. It states "Dublin Core is a starting point but it is overrated." They also commented "the quality of outside search engines is too poor to find terms [expressed in] Dublin Core".

The use of controlled vocabularies is diverse among interviewees. Some do not use controlled vocabularies. One organization uses controlled vocabularies to generate Topic Maps. Others have a very limited set of controlled vocabularies that in general is determined within the organization. Only one interviewee mentions cooperation with other organizations in determining controlled vocabularies. Only one organization has a specific thesaurus manager, all other organizations allocated the maintenance of the controlled vocabularies to an existing department.

One interviewee indicates that the metadata fields **who**, **what**, **where**, and **when** are the crucial fields. One organization has a metadata manager who is responsible for consistency and integrity.

All organizations allow external users to suggest improvements in the website. If there are good suggestions they will be implemented. Four out of seven interviewees do not use a document management system. One organization uses two sets of metadata, one for each system. One has an internal harmonization process in place, while another will in the near future link both systems.

Three of the seven interviewees mention that they do not have a strategy for keeping information readable at all. Two use XML, but one uses it for archiving purposes only. 2 organizations have invalid information on their website, while 3 others check regularly for

the validity of the information and only show valid information on their website. There is a lot of diversity in policies regarding the information on the website among interviewees. There is a difference in who is responsible for the policy regarding web publications. At four organizations the publishing of information on the website is managed centrally, while two others use decentralized management, and one organization does both.

Nearly half of the organizations have their metadata embedded in resources, while the rest store metadata in databases with pointers to the resources. Only two interviewees mention that they do not have ownership of all information on their website; these have agreements with authors about rights.

Most websites are actually composed of a number of different websites. Only two interviewees indicate that they have only one site. One mentions that there is no tuning between the sites, one uses centralized management, and two interviewees use a standard. All interviewees mention links to external websites. Three of them mention no internal management; two have a (minimal form of) internal management and one mentions the use of a discovery server for internal management. External cooperation, in terms of tuning metadata with other organizations, is mentioned by only two organizations.

Appendix B: Internet Survey

Vragenlijst rond de websites van overheidsorganisaties.

De conclusies van de enquête worden beschreven zonder vermelding van de namen van de overheidsorganisatie. Om echter een betere indruk te krijgen van de antwoorden en tevens een controlemiddel te hebben op de ontvangst van de vragenformulieren verzoek ik u hieronder de naam van uw organisatie in te vullen. Bij aanvullende vragen kunnen wij dan nog contact met u opnemen.

1. **Vraag: Wilt u hieronder de naam van uw organisatie invullen en de gegevens van een contactpersoon over het onderwerp.**

   Naam organisatie: …………………………………….

   Naam contactpersoon: …………………………………….

   Bereikbaar onder telefoon: …………………………………….

   e-mail: …………………………………….

<u>Hoofdonderwerp: Inhoud van de informatie op uw website.</u>

De inhoud van de website wordt in het algemeen mede bepaald door de soort overheidsorganisatie.

2. **Vraag: Tot welke van de onderstaande categorieën kan uw organisatie worden gerekend?**

   ☐ Ministeries

   ☐ ZBOs

   ☐ Provincies

   ☐ Gemeentes

   ☐ Waterschappen

   ☐ Anders, te weten: …………………………………………...

3. Vraag: Met welke programmatuur beheert u uw website?

    ………………………………………………………………………

    ………………………………………………………………………

Gebruik van Data

Naast een groot deel aan tekstbestanden wordt er op diverse websites steeds vaker gebruik gemaakt van nieuwe mogelijkheden tot het presenteren van informatie.

4. Vraag: Van welke vormen van informatie wordt er momenteel of binnen afzienbare tijd op uw website gebruik gemaakt? (meerdere antwoorden mogelijk)

    ☐ Tekstbestanden

    ☐ Formulieren

    ☐ Grafische informatie (foto's, tekeningen, etc.)

    ☐ GIS-informatie

    ☐ Gegevens uit workflow-systemen

    ☐ Gegevens uit databases

    ☐ Anders, te weten: ………………………………………...

        …………………………………………………………………

5. Vraag: Welke informatie biedt uw website op dit moment?

    ☐ Wet- en regelgeving, inclusief decentrale wetgeving

    ☐ Vergunningen

    ☐ Beleidsvoornemens

    ☐ Raads-, staten- en bestuursinformatie van gemeenten, provincies en waterschappen

    ☐ Gehonoreerde WOB-verzoeken

    ☐ Europese regelgeving

    ☐ Overige officiële publicaties waarvan bekendmaking wettelijk is vastgelegd

    ☐ VIND catalogus / product specificaties

6. Vraag: Welke diensten biedt uw website op dit moment, met betrekking tot:

    ☐ Wet- en regelgeving, inclusief decentrale wetgeving, namelijk…

□ Vergunningen, namelijk ………………………………………………

□ Europese regelgeving, namelijk ……………………………………….

□ Overige officiële publicaties waarvan bekendmaking wettelijk is vastgelegd, namelijk………………………………………………………………….

□ VIND catalogus / product specificaties, namelijk …………………….

□ Anders, namelijk…………………………………………..

Doorzoekbaarheid

7. **Vraag: Heeft u de doorzoekbaarheid van uw website vergroot, door middel van het toekennen van metadata?**

    □ Neen (ga verder naar vraag 11)

    □ Ja

8. **Vraag: Gebruikt u hiervoor een standaard?**

    □ Neen

    □ Onbekend

    □ Ja, welke? ………………………………………………..

9. **Vraag: Als u metadata heeft toegepast, welke velden heeft u opgenomen?**

    ………………………………………………………………

    ……………………………………………………………

10. **Vraag: Wordt de automatisch aangemaakt of wordt de metadata door de medewerkers ingevuld?**

    □ Automatisch aangemaakt

    □ De metadata wordt door de medewerkers ingevuld

    □ De metadata wordt deels automatisch aangemaakt en deels door de medewerkers ingevuld

11. **Vraag: Maakt u gebruik van metadata om de duurzaamheid van informatie te regelen of te controleren?**

    □ Neen

    □ Ja, door gebruikmaking van de volgende velden:

    ………………………………………………………………

    ………………………………………………………………

12. **Vraag: Is uw website volledig doorzoekbaar door middel van externe zoekmachines?**

    ☐ Ja

    ☐ Onbekend

    ☐ Neen

13. **Vraag: Zijn er belemmeringen in het publiceren van alle mogelijke informatie?**

    ☐ Juridische overwegingen

    ☐ Intellectuele eigendomsrechten

    ☐ Bescherming persoonsgegevens

    ☐ Geregistreerde toegang/lidmaatschap

    ☐ Anders ....……………….................

Bij het invullen van metadata-velden dient de inhoud te voldoen aan afspraken rond manier van opmaak en inhoud. De inhoud van tekst-velden dient over het algemeen te voldoen aan een voorgeschreven waarde uit een tabel met mogelijkheden, de "controlled vocabulary", ofwel de gecontroleerde trefwoordenlijst.

14. **Vraag: Kan er met een gecontroleerde trefwoordenlijst gezocht worden?**

    ☐ Neen (ga verder naar vraag 18)

    ☐ Ja

15. **Vraag: Welke categorieën?**

    ☐ Naam overheidsorganisatie

    ☐ Type overheidsorganisatie

    ☐ Type document

    ☐ Thema

    ☐ Trefwoord

    ☐ Locatie (geografische namen)

    ☐ Doelgroepen

    ☐ Taal

    ☐ Bestand formaat

    ☐ Ander...

16. **Vraag: Kunt u voor elk van deze trefwoordenlijsten aangeven hoe deze zijn bepaald?**

	☐ Bestaande standaarden

	☐ Zelf gegenereerd

	☐ Anders………………………………………………………..

	……………………………………………………………….

	……………………………………………………………….

	……………………………………………………………….

	……………………………………………………………….

	……………………………………………………………….

17. **Vraag: Kunt u voor elk van deze trefwoordenlijsten aangeven door wie deze wordt beheerd?** ....

	……………………………………………………………

	……………………………………………………………

	……………………………………………………………

Interoperabiliteit

Interne operabiliteit: Een website van een overheidsorganisatie is in de praktijk vaak opgebouwd uit meerdere afzonderlijke websites, die voor de gebruiker één geheel vormen. Hierbij kan gekozen worden om alle afzonderlijke websites eenvormig te maken, zodat de afzonderlijke websites als zodanig niet herkenbaar worden

18. **Vraag: Is uw website opgebouwd uit meerdere afzonderlijke websites**

	☐ Neen (ga verder naar vraag 20)

	☐ Ja

19. **Vraag: Hoe wordt dan de onderlinge samenhang geregeld voor wat betreft de lay-out en de doorzoekbaarheid?**

	☐ Zonder onderlinge afstemming tussen de afzonderlijke sites

	☐ Met een gecentraliseerd beheer

	☐ Door middel van gemeenschappelijke standaarden, welke:
	……………………………………………………………

	☐ Anders: ………………………………………………………

Externe operabiliteit: Bij het gebruik van een portal voor het benaderen van de websites van een groot aantal verschillende overheidsorganisaties is het gebruik van identieke begrippen en definities een voorwaarde om de verkregen informatie te kunnen combineren.

20. **Vraag: Heeft er afstemming plaats gevonden met andere overheidsorganisaties om metadata of zoekterm een identieke betekenis te geven, zodat de gevonden informatie eenvoudiger te combineren of te aggregeren is met gegevens van andere websites?**

    ☐ Neen

    ☐ Ja, namelijk met…………………………………………………………………

Appendix C: Questionnaire

Doel van de bijgevoegde enquête.

De overheid stelt zich ten doel om transparanter te worden. Het beschikbaar stellen van meer informatie en diensten via internet wordt gezien als een belangrijke manier om daaraan invulling te geven.

Momenteel is al deze informatie verspreid over ongeveer 1300 overheidswebsites en -databases en is doorgaans zwak of niet gestructureerd. Met zoekmachinetechnologie is het in theorie mogelijk om deze informatie via één punt te ontsluiten. Maar zelfs als dat optimaal slaagt, blijft er sprake van een collectie die weinig samenhang vertoont. Zo is bijvoorbeeld niet betrouwbaar te bepalen of een document behoort tot 'wet- en regelgeving', of de informatie in het document (nog) geldig is, wie de eigenaar is, wat het onderwerp is etc. Als het al zou blijken uit de context waarin het document wordt gepubliceerd, dan is een groot deel daarvan verloren gegaan bij het indexeren ervan.

Het eenvoudigweg beschikbaar stellen van meer informatie en diensten via internet leidt derhalve niet vanzelfsprekend tot meer transparantie. Daarvoor is het informatieaanbod te versnipperd en te onsamenhangend, en zijn doel, context, bereik en/of geldigheid van individuele documenten onvoldoende eenduidig vastgelegd. Een mogelijke oplossing biedt het gebruik van metadata en om onderlinge afstemming te vergroten zou het wenselijk kunnen zijn om hiervoor één gezamenlijke standaard te kiezen.

Via het ICTU heeft Advies Overheid.nl aan RAND Europe, geassisteerd door Gyata Management Consulting bv en DOXsupport, opdracht gegeven tot het opstellen van een advies voor een Nederlandse standaard voor het vaststellen van metadata aan informatie op websites van overheidsorganisaties.

Bij het opstellen van het advies zal, naast gesprekken met een aantal overkoepelende organisaties, tevens rekening worden gehouden met de huidige situatie van, en toekomstige ontwikkelingen rond websites van de verschillende betrokken organisaties.

Bijgaande enquête heeft o.a. tot doel inzicht te verkrijgen in de ervaringen van overheidsorganisaties met het gebruik van metadata. Daarnaast zullen inzichten over diverse standaarden; voorgenomen of te nemen maatregelen rond de duurzaamheid van de informatie; informatie gebruik, etc. aan bod komen. De resultaten van bijgaande enquête vormen een belangrijk bron voor het advies rond het gebruik van metadata in Nederland.

Deze enquête zal worden gebruikt als leidraad bij gesprekken met organisaties en mensen, die kennis en ervaring hebben opgedaan met het gebruik van metadata. Tijdens de interviews wordt de enquête toegelicht en wordt beide partijen de gelegenheid geboden aanvullende informatie rond het onderwerp van de vragen uit te wisselen. Gedurende het interview zal de enquête worden ingevuld.

Vragenlijst rond de websites van overheidsorganisaties.

De conclusies van de enquête worden beschreven zonder vermelding van de namen van de overheidsorganisatie. Om echter een betere indruk te krijgen van de antwoorden en tevens een controlemiddel te hebben op de ontvangst van de vragenformulieren verzoek ik u hieronder de naam van uw organisatie in te vullen. Bij aanvullende vragen kunnen wij dan nog contact met u opnemen.

Wilt u hieronder de naam van uw organisatie invullen en de gegevens van een contactpersoon over het onderwerp .

Naam organisatie:	…………………………………….
Naam contactpersoon:	…………………………………….
Bereikbaar onder telefoon:	…………………………………….
E-mail:	…………………………………….

Hoofdonderwerp: Inhoud van de informatie op uw website.

De inhoud van de website wordt in het algemeen mede bepaald door de soort overheidsorganisatie.

21. **Vraag: Tot welke van de onderstaande categorieën kan uw organisatie worden gerekend?**

☐   Ministeries

☐   ZBOs

☐   Provincies

☐   Gemeentes

☐   Waterschappen

☐   Ander: …………………………………………………..

Naast een groot deel aan tekstbestanden wordt er op diverse websites steeds vaker gebruik gemaakt van nieuwe mogelijkheden tot het presenteren van informatie.

22. **Vraag: In welke vorm wordt informatie momenteel (of binnen afzienbare tijd) op uw website aangeboden?** *(meerdere antwoorden mogelijk)*

- ☐   Tekstbestanden
- ☐   Interactieve formulieren
- ☐   Te downloaden formulieren
- ☐   Grafische informatie (foto's, tekeningen, etc.)
- ☐   GIS-informatie (GIS: Geografische Informatie Systemen)
- ☐   Gegevens uit workflow-systemen
- ☐   Gegevens uit databases
- ☐   Anders, te weten: …………………………………………...

23. **Vraag: Welke <u>informatie</u> biedt uw website op dit moment?**

- ☐   Wet- en regelgeving, inclusief decentrale wetgeving
- ☐   Vergunningen
- ☐   Beleidsvoornemens
- ☐   Voorlichting
- ☐   Raads-, staten- en bestuursinformatie van gemeenten, provincies en waterschappen
- ☐   Gehonoreerde WOB-verzoeken
- ☐   Europese regelgeving
- ☐   Overige officiële publicaties waarvan bekendmaking wettelijk is vastgelegd
- ☐   VIND catalogus / product specificaties
- ☐ Anders, namelijk……………………………………………..

24. **Vraag: Welke <u>diensten</u> biedt uw website op dit moment, met betrekking tot:**

- ☐   Wet- en regelgeving, inclusief decentrale wetgeving, namelijk:
- ☐   Vergunningen ……………………………………………
- ☐   Europese regelgeving…………………………………….
- ☐   Overige officiële publicaties waarvan bekendmaking wettelijk is vastgelegd
- ☐   VIND catalogus / product specificaties: ……………………
- ☐ Anders, namelijk……………………………………………

25. **Vraag: Wordt er op uw website onderscheid in informatie gemaakt naar verschillende soorten gebruikers, en zo ja, welke groepen worden op uw website onderscheiden?**

    ☐ Neen

    ☐ Ja, de volgende groepen :

    ☐ Burgers

    ☐ Bedrijven / ondernemingen

    ☐ Politici en / of ambtenaren

    ☐ Anders, namelijk……………………………………….

26. **Vraag: Kunt U gegevens of een schatting geven van de aantallen gebruikers per categorie, zoals genoemd onder vraag 5?**

    ☐ Neen

    ☐ Ja, namelijk…………………………………………………...

    …………………………………………………………………….

27. **Vraag: Wordt het gebruik van de website op dit moment gemeten per onderdeel van de totale site, waarbij onderscheid gemaakt kan worden tussen informatie (Zie vraag 3) en diensten (zie vraag 4)?**

    ☐ Neen

    ☐ Ja, wat zijn hiervan de resultaten?………………………………

    ……………………………………………………………….......

    ……………………………………………………………….......

Het beheer van de grote hoeveelheid informatie op een website wordt vaak automatisch uitgevoerd met behulp van een 'content management systeem'

28. **Vraag: Wordt er voor het beheer van uw website gebruik gemaakt van programmatuur?**

    ☐ Neen

    ☐ Ja, namelijk ……………………………………............

Een aantal websites is <u>niet</u> doorzoekbaar door zoeksystemen van buitenaf.

29. **Vraag: Is alle op uw website aangeboden informatie voor alle bezoekers bereikbaar en doorzoekbaar door een externe zoekmachine?**

    ☐   Ja (ga verder naar vraag 11)

    ☐   Neen, afscherming van informatie door ....

    ☐   Neen, wegens ………………………………… ………………

30. **Vraag: Indien niet alle aangeboden informatie bereikbaar is voor alle bezoekers en doorzoekbaar door een externe zoekmachine, hoe wordt de informatie dan afgeschermd?**

    ……………………………………………………………….

    ……………………………………………………………….

31. **Vraag: Waardoor wordt eventuele verder uitbreiding van de inhoud van uw website tegengehouden?**

    ☐   Interne, organisatorische beperkingen

    ☐   Interne, financiële beperkingen

    ☐   Wet- en regelgeving

    ☐   Interne, technische beperkingen

    Toelichting: …………………………………………………

    ……………………………………………………………….

32. **Vraag: In hoeverre ondervindt u invloed van de WPB, PUBER en WOB in het ontsluiten van overheidsinformatie?**

    ☐   WPB (Wet Bescherming van Persoonsgegevens)

    ……………………………………………………………….

    ☐   PUBER (Wet Kenbaarheid Publiekrechtelijke Beperkingen)

    …………………………………………………………………

    ☐   WOB (Wet Openbaarheid Bestuur)……………………………

    ……………………………………………………………….

Hoofdonderwerp: Structuur van de informatie

De hoeveelheid van informatie op de website van een overheidsorganisatie vereist een bepaalde structuur, teneinde het kunnen vinden van de gewenste informatie te vereenvoudigen. Naast een menustructuur wordt vrijwel op elke website de mogelijkheid geboden gebruik te maken van een zoekmachine.

33. **Vraag: Hoe kan op uw website worden gezocht? (meerdere antwoorden mogelijk)**

  ☐ Op teksten of delen van tekst (vrije tekst)

  ☐ Op tekst, gecombineerd met vaste onderwerpen (bijv. diensten)

  ☐ Op tekst, gecombineerd met een categorie, vastgelegd in metadata.

  ☐ Door het gebruik van een query, samengesteld door een gebruiker

  ☐ Anders, namelijk ……………………………………………

Indien op uw website al gebruik wordt gemaakt van metadata, die is gekoppeld aan informatie op uw website, dan is het in het kader van het advies voor het opstellen van een standaard binnen Nederland interessant te bepalen, welke standaarden voor metadata in de praktijk al worden gebruikt.

34. **Vraag: Hoe zijn de gebruikte velden voor de metadata tot stand gekomen?**

  ☐ Er wordt geen gebruik gemaakt van metadata-velden.

  ☐ De metadata wordt automatisch gegenereerd door een applicatie

  ☐ Binnen de eigen organisatie zijn metadata-velden bepaald.

  ☐ Binnen een overkoepelende organisatie zijn metadata-velden bepaald.

    ▪ De overkoepelende organisatie is :

    ………………………………………………...........

  ☐ Er wordt gebruik gemaakt van een internationale standaard.

    ▪ De internationale standaard, die wordt gevolgd is:

    …..……………………………………………………

  ☐ Anders, namelijk ……………………………………………

    …………………………………………………………………

35. **Vraag: Indien bij vraag 14 is aangegeven dat een standaard wordt gevolgd, welke tijd c.q. investering is gemoeid geweest met het kiezen en volgen van de standaard?**

    ………………………………………………………………

Voor het gebruik van metadata zijn internationaal een aantal standaarden vastgelegd. Een bekende standaard hierbij is de Dublin Core standaard, die in een groot aantal landen de basis is voor een nationale standaard op het gebied van toekennen van metadata aan informatie.

36. **Vraag: Heeft U ervaring met de Dublin Core metadatastandaard?**

    ☐ Neen

    ☐ Ja, namelijk……………………………………………………

37. **Vraag: Zo ja, wat ziet U als de voor- en nadelen?**

    ☐ Voordelen…………………………………………………..

    ☐ Nadelen…………………………………………………..

Bij het invullen van metadata-velden dient de inhoud te voldoen aan afspraken rond manier van opmaak en inhoud. De inhoud van tekst-velden dient over het algemeen te voldoen aan een voorgeschreven waarde uit een tabel met mogelijkheden, de "controlled vocabulary". Indien de "controlled vocabulary" tevens wordt toegepast bij het zoeken naar documenten, dan wordt de toegankelijkheid vergroot.

38. **Vraag: Wordt bij het toekennen van metadata gebruik gemaakt van "controlled vocabularies"?**

    ☐ Neen (ga verder naar vraag 21)

    ☐ Ja, voor de volgende velden:

    …………………………………………………………..

39. **Vraag: Kan er via "controlled vocabularies" worden gezocht?**

    ☐ Neen (ga verder naar vraag 21)

    ☐ Ja, voor de volgende velden:

    …………………………………………………………..

    …………………………………………………………..

40. **Vraag: Hoe zijn de "controlled vocabularies" vastgesteld?**

    ☐ Door eigen organisatie

    ☐ Anders, namelijk ………………………………………….

41. **Vraag: Op welke categorieën kan er op uw website worden gezocht?**

    ☐ Naam overheidsorganisatie

□ Type overheidsorganisatie

□ Type document

□ Thema

□ Trefwoord

□ Locatie (geografische namen)

□ Doelgroepen

□ Taal

□ Bestandformaat

□ Ander...

42. Vraag: Welke organisatie, interne afdeling of externe organisatie, beheert de "controlled vocabularies", die binnen uw organisatie worden gebruikt ?

     ………………………………………………………………..

     ………………………………………………………………..

43. Vraag: Welke metadata-velden zijn bij het zoeken op uw website cruciaal?

     ………………………………………………………………..

     ………………………………………………………………..

44. Vraag: Welke metadata-velden hebben binnen uw organisatie altijd een identieke inhoud, default waarde, en welke inhoud wordt dan ingevuld als defaultwaarde?

     ………………………………………………………………..

     ………………………………………………………………..

45. Vraag: Welke notatiewijze (bijvoorbeeld voor tijd, datum, telefoon, getal) wordt binnen uw organisatie gevolgd ?

     ………………………………………………………………..

     ………………………………………………………………..

Door nieuwe technische mogelijkheden, of andere behoeften bij het zoeken naar informatie op uw website, kan het gebruik van de metadata-velden zijn veranderd. .

46. Vraag: Zijn aanpassingen aan metadata-velden aangebracht?

□ Neen, na eerste gebruik niet meer aangepast.

☐   Ja, welke……...………………………………………..

    Waarom: …………………………………………………

De gebruikers van uw website ervaren de zoekmogelijkheden soms anders dan de ontwerpers van de site. Ook omtrent de aangeboden informatie kan een gebruiker een ander oordeel vellen dan de beheerder van de website. Op een aantal websites wordt voor de gebruikers de mogelijkheid geboden om voorstellen voor wijzigingen of verbeteringen door te geven.

47. **Vraag: Staat u externe gebruikers toe om voorstellen te doen voor aanpassingen en verbeteringen?**

    ☐   Neen

    ☐   Ja en deze worden op de volgende wijze verwerkt:

        …………………………………………………………..

        …………………………………………………………..

Bij het bepalen van metadatavelden en de inhoud daarvan dient er voor te worden gezorgd, dat diverse metadatavelden niet met elkaar in tegenspraak zijn: consistent zijn.

Daarnaast dient de metadata volledig te zijn: de integriteit.

48. **Vraag: Hoe garandeert U de consistentie en integriteit van uw metadata?**

    Consistentie: ..…………………………………………..

    Integriteit: ..…………………………………………..

49. **Vraag: Indien uw organisatie gebruik maakt van een document management systeem, is er dan een interne afstemming tussen de metadata van dit systeem en de metadata van de informatie op uw website?**

    ☐   Neen, er wordt geen document management systeem gebruikt.

    ☐   Neen, beide systemen gebruiken hun eigen set van metadata.

    ☐   Ja, een interne afstemming

    ☐   Ja, een automatische afstemming door koppeling van de beide applicaties.

50. **Vraag: Indien uw organisatie gebruik maakt van een document management systeem, wordt er dan metadata, specifiek gericht op record management toegekend aan informatie?**

    ☐   Neen, er wordt geen document management systeem gebruikt.

    ☐   Ja, hierbij zijn binnen de eigen organisatie metadata-velden bepaald, namelijk:

        …………………………………………………………..

☐ Ja, hierbij wordt de volgende standaard gevolgd, namelijk:

…………………………………………………………..

Hoofdonderwerp: Duurzaamheid van de informatie.

51. **Vraag: Heeft u een strategie om de informatie leesbaar te houden in de komende jaren?**

☐ Neen

☐ Nog niet definitief

☐ Door migratie van gegevens, indien nodig

☐ Door omzetting naar een geaccepteerde standaard, bijvoorbeeld

▪ Adobe pdf-formaat

▪ XML

▪ Anders, namelijk………………………….

☐ Anders, namelijk…………………………….

Over het algemeen wordt op een website alleen actuele informatie opgenomen. Op sommige websites is ook informatie, die niet meer geldig is, nog steeds benaderbaar.

52. **Vraag: Wordt er op uw website alleen actuele informatie opgenomen en wordt informatie, die niet meer van toepassing is van de website verwijderd?**

☐ Ja

☐ Neen, verlopen informatie blijft opgenomen op de website.

53. **Vraag: Welke strategie wordt gevolgd om de actualiteit van de gepresenteerde gegevens te waarborgen?**

…………………………………………………………..

…………………………………………………………..

Hoofdonderwerp: Beleid en implementatie

54. **Vraag: Wat is het beleid op het gebied van webpublicaties?**

…………………………………………………………..

…………………………………………………………..

55. **Vraag: Wie is verantwoordelijk voor het beleid op het gebied van webpublicaties?**

....

56. **Vraag: Hoe is het publiceren van informatie op de website georganiseerd en met welke ervaringen?**

☐ Centraal

☐ Decentraal

Ervaringen hiermee:

....

57. **Vraag: Hoe is de metadata gekoppeld aan de informatie?**

☐ De metadata is opgenomen in de informatie

☐ De metadata is opgenomen in een database, waarin koppelingen met de informatie zelf zijn opgenomen.

☐ Anders, namelijk..

58. **Vraag: Is alle informatie op uw site uw eigendom, waardoor toegangsrechten, geldigheidsduur etc zelf kunnen worden bepaald?**

☐ Ja

☐ Neen.

59. **Vraag: Indien niet alle informatie op uw site uw eigendom is, hoe worden dan bijvoorbeeld toegangsrechten, geldigheidsduur etc afgestemd met de leverancier van de informatie?**

....

60. **Vraag: Welke maatregelen heeft u genomen om publicatie van informatie van externe leveranciers mogelijk te maken? Hierbij wordt gedacht aan protocollen of overeenkomsten met externe leveranciers van informatie.**

....

61. **Vraag: Welke maatregelen heeft u genomen om publicatie van privacy gevoelige informatie te voorkomen? Hierbij dient er tevens op worden gelet, dat ook door combinatie van gegevens van verschillende websites privacy-gevoelige informatie kan worden gevonden.**

....

Hoofdonderwerp: Interoperabiliteit

Een website van een overheidsorganisatie is in de praktijk vaak opgebouwd uit meerdere afzonderlijke websites, die voor de gebruiker één geheel vormen. Hierbij kan gekozen worden om alle afzonderlijke websites eenvormig te maken, zodat de afzonderlijke websites als zodanig niet herkenbaar worden

62.   **Vraag: Is uw website opgebouwd uit meerdere afzonderlijke websites**

☐   Neen (ga verder naar vraag 44)

☐   Ja

63.   **Vraag: Hoe wordt dan de onderlinge samenhang geregeld voor wat betreft de lay-out en de doorzoekbaarheid?**

☐   Zonder onderlinge afstemming tussen de afzonderlijke sites

☐   Met een gecentraliseerd beheer

☐   Door middel van gemeenschappelijke standaarden, namelijk:

………………………………………………………………….

☐   Anders: …………………………………………………………

64.   **Vraag: Zijn er op uw website links aanwezig naar andere, externe websites en hoe is het beheer van deze links geregeld.?**

☐ Neen

☐ Ja, zondere beheer

☐ Ja, beheer door …………………………………………………..

65. **Vraag: Is de metadata afgestemd met andere organisaties, waardoor bij het doorzoeken van meerdere van websites van meerdere organisaties uitgegaan wordt van een gelijke betekenis van de betreffende metadata?**

☐   De metadata is niet afgestemd met andere organisaties

☐   De metadata is afgestemd met andere organisaties, namelijk

…………………………………………………………………..

Appendix D List of organizations contacted

Interviews	Telephone and e-mail contacts
Ministries[6]:	**Municipalities**
Ministerie van Landbouw, Natuur en Voedselkwaliteit	e-Gemeenten
Ministerie van Volksgezondheid, Welzijn en Sport	Gemeente Zeist
Ministerie van Algemene Zaken, Rijksvoorlichtingsdienst	Gemeente Breda
Ministerie van Onderwijs, Cultuur en Wetenschap	Gemeente Rotterdam
Ministerie van Buitenlandse Zaken	Gemeente Archief Amsterdam
Ministerie van Sociale Zaken en Werkgelegenheid	Gemeente Enschede
Ministerie van Economische Zaken	**Provinces**
Ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieu	Provincie Noord Holland
Ministerie van Verkeer en Waterstaat	Provincie Utrecht
Co-ordinating Organisations and Programmes	Provincie Zuid Holland
e-Provincies	Waterboards
Interlab	Unie van Waterschappen
Vereniging van Nederlandse Gemeenten	Waterschap de Brabantse Delta
ELO Kennisloket	Waterschap Hunze en Aa's
Programma Stroomlijning Basisgegevens	**Other**
Provinces and Municipalities	Centraal Bureau voor de Statistiek
Gemeente van Den Haag	Commissariaat voor de Media

[6] Interview with Ministerie van Binnenlandse Zaken has been cancelled by the Ministry

Interviews	Telephone and e-mail contacts
Gemeente Archief van Den Haag	Zoek.nl
Gemeente Zwijndrecht	International
Gemeente Apeldoorn	Dublin Core
Provincie Noord Brabant	AMI Consult
Other Governmental Organisations	AGLS / Recordkeeping Innovation PTY Ltd. (Australia
Nationaal Archief	IT-og Telestyrelsen (Denmark)
Belastingdienst	
Koninklijke Bibliotheek	
Other Metadata Initiatives	
Surfnet	
Kennisnet	
Activity Index	
e-Grant	
Doxis	
International	
e-GMS UK	
National Archives UK	

Appendix E Description of NL-meta.DC+ elements

Mandatory elements	Origin	Description
date	DC	A date associated with an event in the life cycle of the resource. Typically, Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 [Date and Time Formats, W3C Note, http://www.w3.org/TR/NOTE-datetime] and follows the YYYY-MM-DD format.
language	DC	A language of the intellectual content of the resource. Recommended best practice for the values of the Language element is defined by RFC 3066 [RFC 3066, http://www.ietf.org/rfc/rfc3066.txt] which, in conjunction with ISO 639 [ISO 639, http://www.oasis-open.org/cover/iso639a.html]), defines two- and three-letter primary language tags with optional subtags. Examples include "en" or "eng" for English, "akk" for Akkadian, and "en-GB" for English used in the United Kingdom.
subject	DC	The topic of the content of the resource. Typically, a Subject will be expressed as keywords or key phrases or classification codes that describe the topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.
title	DC	The name given to the resource. Typically, a Title will be a name by which the resource is formally known.
audience (if applicable)	EGMS	A class of entity for whom the resource is intended or useful. A class of entity may be determined by the creator or the publisher or by a third party.
coverage (if applicable)	DC	The extent or scope of the content of the resource. Coverage will typically include spatial location (a place name or geographic co-ordinates), temporal period (a period label, date, or date range) or jurisdiction (such as

		a named administrative entity). Recommended best practice is to select a value from a controlled vocabulary (for example, the Thesaurus of Geographic Names [Getty Thesaurus of Geographic Names, http://www.getty.edu/research/tools/vocabulary/tgn/]). Where appropriate, named places or time periods should be used in preference to numeric identifiers such as sets of co-ordinates or date ranges.
creator (if applicable)	DC	An entity primarily responsible for making the content of the resource. Examples of a Creator include a person, an organization, or a service. Typically the name of the Creator should be used to indicate the entity.
identifier (if applicable)	DC	An unambiguous reference to the resource within a given context. Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Examples of formal identification systems include the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN).
publisher (if applicable)	DC	The entity responsible for making the resource available. Examples of a Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity.
type (if applicable)	DC	Type includes terms describing general categories, functions, genres, or aggregation levels for content. Recommended best practice is to select a value from a controlled vocabulary. To describe the physical or digital manifestation of the resource, use the Format element.
Recommended elements		
format	DC	The physical or digital manifestation of the resource. Typically, Format may include the media-type or dimensions of the resource. Examples of dimensions include size and duration. Format may be used to determine the software, hardware or other equipment needed to display or operate the resource. Recommended best practice is to select a value from a controlled vocabulary (for example, the list of Internet Media Types [http://www.isi.edu/in-notes/iana/assignments/media-types/media-types] defining computer media formats).
status	eGMS	The position or state of the resource.

Optional elements		
accessibility	eGMS	Indicates the resource's availability and usability to specific groups
contributor	DC	An entity responsible for making contributions to the content of the resource. Examples of a Contributor include a person, an organization or a service. Typically, the name of a Contributor should be used to indicate the entity.
description	DC	An account of the content of the resource. Description may include but is not limited to: an abstract, table of contents, reference to a graphical representation of content or a free-text account of the content.
mandate	eGMS	Legislative or other mandate under which the resource was produced.
relation	DC	A reference to a related resource. Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system.
source	DC	A Reference to a resource from which the present resource is derived. The present resource may be derived from the Source resource in whole or part. Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system.
Admin elements		
preservation	eGMS	Information to support the long-term preservation of a resource.
rights	DC	Information about rights held in and over the resource. Typically a Rights element will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. If the rights element is absent, no assumptions can be made about the status of these and other rights with respect to the resource.