



EUROPE

CHILD POLICY
CIVIL JUSTICE
EDUCATION
ENERGY AND ENVIRONMENT
HEALTH AND HEALTH CARE
INTERNATIONAL AFFAIRS
NATIONAL SECURITY
POPULATION AND AGING
PUBLIC SAFETY
SCIENCE AND TECHNOLOGY
SUBSTANCE ABUSE
TERRORISM AND
HOMELAND SECURITY
TRANSPORTATION AND
INFRASTRUCTURE

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](#)

Limited Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law as indicated in a notice appearing later in this work. This electronic representation of RAND intellectual property is provided for non-commercial use only. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use.

This product is part of the RAND Corporation technical report series. Reports may include research findings on a specific topic that is limited in scope; present discussions of the methodology employed in research; provide literature reviews, survey instruments, modeling exercises, guidelines for practitioners and research professionals, and supporting documentation; or deliver preliminary findings. All RAND reports undergo rigorous peer review to ensure that they meet high standards for research quality and objectivity.

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

Prepared for Advies Overheid.nl and funded by the Ministry of the Interior and
Kingdom Relations

The research described in this report was prepared for Advies Overheid.nl and funded by the Ministry of the Interior and Kingdom Relations.

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.

© Copyright 2005 Stichting ICTU

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the copyright holder.

Published 2005 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
201 North Craig Street, Suite 202, Pittsburgh, PA 15213-1516
RAND URL: <http://www.rand.org/>
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

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.