

## Making Your Way Through Grey: Metadata, MARC and User Tools

Meagan Cooke, Senior Production Editor  
Sean S. Costigan, Executive Editor<sup>1</sup>

### Abstract

*Columbia International Affairs Online (CIAO) is the largest online collection of grey literature on international relations. Online since 1997, CIAO has grown to encompass both grey literature and published materials from over 200 contributing institutions. Currently the database contains over 10,000 papers and, as with anything of this magnitude, the challenge is not only continued growth in terms of content aggregation but also achieving successful user experiences and wider integration with other services. Our metadata is based on a small, controlled vocabulary that was developed in-house over the lifetime of the service. Pursuing an overhaul of CIAO's metadata for improved consistency has allowed for finer granularity in search results while also creating opportunities for the deployment of citation tools and rich cross-linking. In addition, the development of MARC records will allow for further integration with library resources and OPACs. This paper will discuss the impetus for the development of user and contextualization tools and our experiences in creating them.*

### CIAO's Background

CIAO is the largest library of international affairs content on the web. Originally funded by the Mellon Foundation, CIAO became self-sustaining through library subscriptions after three years of operation. CIAO was built in a partnership with Columbia's libraries, the University Press, and its academic computing and information systems group (AcIS). Subject specialists, computer scientists and librarians all had a hand in its initial development. Today, such expertise is drawn on to further realize the service's goals of promoting a wide range of grey and published literature in international affairs. Currently over 200 institutions partner with CIAO, primarily contributing working papers, conference proceedings, reports, books, policy briefs and journals. CIAO boasts more than 800 subscribers, among them government agencies, militaries, academic institutions and businesses. In any given month, over 2000 pages of material from dozens of contributors will be posted on CIAO. Such a large and mature repository poses significant challenges with regard to data management, archiving and customization.

### Many Organizations, Many Standards

At CIAO's inception in 1997 a variety of file formats were commonly in use. CIAO's production staff was likely to receive files from Word, WordPerfect, Quark, and a smattering of non-standard text editors. In keeping with our desire to make CIAO usable to as extensive an audience as possible, all files were converted to faster loading html. Initially, CIAO adhered to HTML 2.0 specifications and when additional HTML specifications came out, adjustments were made to new, but not existing, content.

Today the bulk of CIAO's contributors deliver content in PDF (Portable Document Format) or Microsoft Word. For some of CIAO's subscribers, particularly those overseas and from secure locations, low bandwidth continues to be an issue that we design around by producing HTML abstracts for the majority of PDFs. In addition, html abstracts afford us the opportunity to more comprehensively describe the content using our metadata. Where possible we add author and title information to the PDFs, allowing our search engine to take advantage of that metadata as it indexes the site.

### Metadata - A Primer

Several definitions of metadata exist. The W3C defines metadata as "machine understandable information for the web." For our purposes, metadata is simply "data about data" -- particularly information like keywords, document type, title, abstract, location, ISBN, etc.

---

<sup>1</sup> Meagan Cooke and Sean Costigan can be reached at mc1430@columbia and sc717@columbia.edu.