



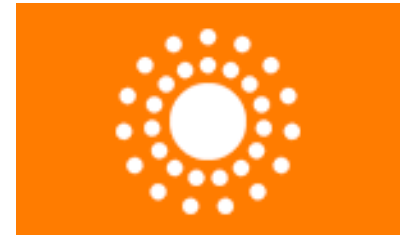
DCMI Kernel Metadata Working Group

4 Oct 2006 – DC2006, Manzanillo, Mexico

¿Quién? ¿Qué? ¿Cuándo? ¿Dónde?
(¿Cómo? ¿Por qué?)

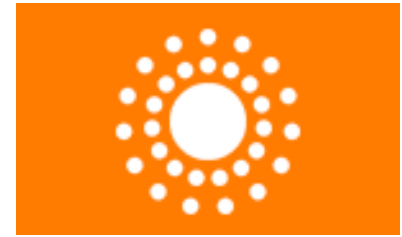
Who? What? Where? When?
(How? Why?)

Kernel WG Agenda



- Welcome and Introduction
- Current status and summary
- Review what Kernel metadata is
 - Plus application in the Tiny HTTP URL Mapping Protocol (THUMP)
- Pete Johnston:
 - commentary on Kernel Application Profile (KAP) conformance to DCMI Abstract Model
- Alistair Miles: the SKOS perspective
- Jane Greenberg: the Tools WG perspective
- Wrap up discussion and completing the KAP

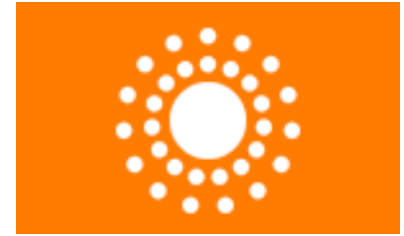
Kernel Working Group



- Established Oct 2002
- Current mailing list: 55 subscribers
- Charter: to provide a forum...
 - to explore the ultra-simple ERC "kernel" approach to metadata
 - to identify applications of the compact ERC record format
 - to refine ERC value rules and minimum element requirements
 - to develop an XML representation of kernel elements
- DC 2001, Tokyo paper

erc: Kunze, John A. | A Metadata Kernel for Electronic Permanence
| 20011106 | <http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Kunze/>

DCMI Kernel Results



- Draft DCMI Kernel metadata specification
 - Taught in graduate school metadata classes
- Support in two open-source search engines
 - Amberfish and Isite2
- Perl module for producing metadata
- Draft Kernel Application Profile
- Working Group to become Task Force/Community
 - Top priority: complete Kernel Application Profile
- Volunteers?

A Tiny Retrieval Protocol: THUMP + Kernel Metadata

June 2006 – JCDL Metadata Tools Workshop

Kevin Gamiel, Renaissance Computing Institute John
Kunze, California Digital Library Nassib Nassar,
Renaissance Computing Institute

Overview

Objects, surrogates, and metadata

Simple protocols aren't

Simple metadata ain't

Making it minimal: Kernel/ERC

Thinking tiny: THUMP

Applications: persistence and discovery



FEMA

Disaster Communities

Acting Under Secretary R. D.

Search FEMA

Go

» Advanced Search

Disaster

- [Current Disasters](#)
- [Declared Disasters Archive](#)
- [Hazards](#)
- [Current FEMA News](#)
- [Individual Assistance](#)
- [Public Assistance](#)

SERVICES

Ecological Applications, 13(6), 2003, pp. 1626–1639
© 2003 by the Ecological Society of America

SPECIES-RICH PLANTINGS INCREASE BIOMASS AND NITROGEN ACCUMULATION IN A WETLAND RESTORATION EXPERIMENT

JOHN C. CALLAWAY,^{1,4} GARY SULLIVAN,² AND JOY B. ZEDLER³

¹Department of Environmental Science, University of San Francisco, 2130 Fulton St.,
San Francisco, California 94117 USA

²The Wetlands Initiative, 53 West Jackson Boulevard, Suite 1015, Chicago, Illinois 60604-3703 USA

³Botany Department and Arboretum, 430 Lincoln Drive, University of Wisconsin, Madison, Wisconsin 53706 USA

Abstract. Our test of the hypothesis that biomass and nitrogen would increase with more species-rich plantings simultaneously vegetated a salt marsh restoration site and demonstrated that on average, randomly chosen, 6-species plantings accumulated more biomass and nitrogen than the mean for 0- and 1-species assemblages, with the mean for 3-species assemblages being intermediate. In addition, we found that individual species (from the pool of eight native halophytes) differed in their functional capacity, with *Salicornia virginica* (Sv) and *Jaumea carnosa* contributing the greatest biomass when planted alone, while *Triglochin concinna* had the highest tissue N concentrations. When planted alone, Sv accumulated comparable amounts of biomass and nitrogen as in the multispecies plots, indicating that individual species can have a large effect on particular functions. Soil TKN in the surface 0–5 cm was greater in 6-species plots than unplanted plots in 1999, while both 3- and 6-species plots were greater than unplanted plots in 2000; however, there were no differences at 5–20 cm depth and no species-specific effects. Root and shoot biomass both increased with species richness, with total biomass of 6-species plots averaging 995.6 ± 120.5 g/m² in 2000, compared to the mean for 1-species plots (572.1 ± 90.3 g/m²) and unplanted plots (164.5 ± 24.7 g/m²). Still, at the age of three years, root biomass was only about one-third that of the species-rich reference site, and shoot biomass was one-half to one-fifth the maxima reported for reference salt marshes. Species-specific effects were found for Sv, which had high biomass of both roots and shoots in the multispecies plots (55% of aboveground biomass in 3-species plots and 41% in 6-species plots) and the highest pool of N (52% of the N pool in 3-species plots and 42% in 6-species plots), even though only one-eighth of the initial plantings were Sv. However, when plots with this species were excluded from the analysis, the species-richness effect persisted. Thus, ecosystem function, as measured by biomass and N accumulation, increased with species richness regardless of dominance by the highly productive Sv. We conclude that manipulating the richness and composition of plantings offers ecosystem restorationists an effective tool for accelerating the rate of functional development.

Key words: biodiversity; diversity; ecosystem functions; nitrogen; richness; *Salicornia virginica*; salt marsh; wetland restoration.

INTRODUCTION

The last decade has seen an explosion of interest in the relationship between species diversity and ecosystem function, with many experiments supporting positive relationships between species richness and productivity, despite controversies over the interpretation of results (Schulze and Mooney 1993, Naeem et al. 1994, Huston 1997, Tilman et al. 1997, 2001, Schwartz et al. 2000, Tilman 2000, Engelhardt and Ritchie 2001). To date, researchers have asked how the loss of species diversity affects ecosystem functioning (Naeem et al. 1994, Chapin et al. 2000, Tilman et al. 2001), focusing on the concern of global impacts to species diversity (Pimm et al. 1995, Gaston 2000). But as natural habitats

continue to be lost, conservationists become increasingly dependent on restoration efforts for improving the status of degraded ecosystems (Daily 1995, Dobson et al. 1997). Hence, we asked a corollary question: Does increasing the number of species accelerate the development of functions in restored ecosystems?

Improving our ability to restore functional ecosystems requires that we identify the factors that constrain ecosystem development (Simenstad and Thom 1996). Attempts to understand natural ecosystem development date to Odum's seminal paper (1969), and the development of ecosystem functions has been linked to the accumulation of soil organic matter (Jenny 1941, Crocker and Major 1955) and the soil nutrient pool (Chapin et al. 1986, Vitousek et al. 1993, Schlesinger et al. 1998, Crews et al. 2001). However, few investigators have followed the cycling and retention of carbon and nutrients in restoration sites over long time periods (Bishel-Machung et al. 1996, Simenstad and

Manuscript received 29 April 2002; revised 18 December 2002; accepted 30 January 2003; final version received 26 February 2003. Corresponding Editor: I. C. Burke.

⁴ E-mail: callaway@usfca.edu

A In the News

aster

ndays, Oct 28 **New**

shatta, Oct 27

, Oct 27

CLICK HERE TO VIEW
ELECTION RESULTS

Object Surrogates

Surrogates provide a time-honored way of avoiding the inconvenience of directly handling objects.

- Surrogates usually much smaller (eg, a catalog card)
- A surrogate may be unencumbered and in a language you understand even if the object isn't
- Surrogates can be much more *uniform* (for easier processing) than objects
- A surrogate is essentially a metadata record for an object
- Every system has surrogates, even if dynamically generated

Reminder: What is metadata for?

- Metadata is a surrogate-based tool to help us find, use, and manage information *objects*, *resources*, or **stuff**.

Where metadata meets protocol

Metadata definition 1: “data about data”

- Too broad and too narrow, e.g., a book review? a catalog record for a statue?

Metadata definition 2: “structured data about stuff”

- “stuff” avoids having to say a statue is data
- “structured” data assists *automation* by making it easy to recognize and record individual data elements
- The more uniform, the more leverage for *interoperation*

Automation + Interoperation \Rightarrow Protocol

Simple protocols aren't

In the beginning, ... application protocols layered on TCP/IP

- Email set the standard for simplicity (RFC 822 headers)
- HTTP, NTTP, gopher, etc. followed its lead; OSI protocols withered

Second system syndrome (expanding functionality):

- Z39.50, CORBA, SOAP, and others

Regret period (contracting complexity):

- OpenSearch, RSS, and in DL world, SRW/SRU, OAI

How are we doing?

- Tues 13 June: “low barrier” OAI failures attributed to errors in XML coding, schemas; poor, inconsistent, and expensive metadata; with surrogates too non-uniform to be of much use [CL & CL]

How might this contraction phase bottom out? (hint: with a THUMP)

Simple metadata ain't

Dublin Core metadata tried to be simple

- Goal: “specification shouldn’t register on a bathroom scale”

Goal achieved, but DC spec. was a bit under-specified, practical applications must add:

- definition of *record*
- concept of minimal object description
- layout rules for author names and dates
- *meta-metadata*, eg, provenance, commitment statements

Simple becomes not so simple

Simple Metadata: Dublin Core

15 elements thought to apply to almost any object – discovery as goal

<i>Content</i>	<i>Intellectual Property</i>	<i>Instantiation</i>
Coverage	Contributor	Date
Description	Creator	Format
Type	Publisher	Identifier
Relation	Rights	Language
Source		
Subject		
Title		

Despite DCMI efforts to correct known problems, the simplest protocol with the simplest metadata – OAI – reports an overall 36% failure rate, 77% due to metadata/encoding and protocol errors.

Simple Dublin Core metadata

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF PUBLIC "-//DUBLIN CORE//DCMES DTD
    2002/07/31//EN"
    "http://dublincore.org/documents/2002/07/31/dcmes-
    xml/dcmes-xml-dtd.dtd">
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
    ns#"
        xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description
    rdf:about="http://www.nap.edu/books/0309064996/html/">
    <dc:title>The Digital Dilemma</dc:title>
    <dc:creator>National Research Council</dc:creator>
    <dc:date>2000-06-22</dc:date>
  </rdf:Description>
</rdf:RDF>
```

Same record with Dublin Kernel

Here's the same information, still machine-readable, as an Electronic Resource Citation (ERC) with Kernel metadata:

```
erc:
who:   National Research Council
what:  The Digital Dilemma
when:  2000
where: http://books.nap.edu/html/digital%5Fdilemma
```

Motivators for the ERC

- Meet the need for a simple and manipulable record
- Direct human contact with metadata is inevitable
- Record should place minimal strain on people
- Succinct, transparent, trivially parseable (2 lines of Perl code)

Making it minimal: Kernel/ERC

Electronic Resource Citation (ERC) – back to basics

- An ERC record is a sequence of elements in email header format:
 ⇒ label, colon, value
- Long values are continued on indented lines
- A blank line ends a record

Based on cross-domain kernel distilled from Dublin Core

- **who** – a responsible person or party
- **what** – a name or other human-oriented identifier
- **when** – a date important in the object's lifecycle
- **where** – a location or a machine-oriented identifier

The ERC notion of “story”

The same record as before, in its most compact form:

```
erc: National Research Council  
    | The Digital Dilemma | 2000  
    | http://books.nap.edu/html/digital%5Fdilemma
```

Either ERC form starts by telling the story of an *expression* of the resource, applying who-what-when-where questions to it.

- All 4 kernel elements are required
- Absent values must be explained; 7 flavors of “empty”
- Element ordering is rigid in compact form (positional semantics)
- Arbitrary additional elements may occur after the 4 elements

Other *segments* in the ERC may introduce other stories, such as,

- `erc-about`, `erc-support`, `erc-from`

A 2-story ERC record

erc:

who: Tomlinson, Richard

what: Adjustable knock down chair

when: (:unkn)

where: [http://espacenet.com/dips/bnsviewer%{
? CY=ec & LG=en & DB=EPD & PN=US5498054
& ID=US+++5498054A1+I+ %}](http://espacenet.com/dips/bnsviewer%{? CY=ec & LG=en & DB=EPD & PN=US5498054 & ID=US+++5498054A1+I+ %})

erc-support:

who: European Patent Office

what: (:permuc) Permanent, Unchanging Content

Note to ops staff: verify date.

when: 20010621

where: <http://ark.espacenet.com/ark:/23003/US5498054>

Mapping ERC to Dublin Core

Kernel Element	Equivalent DC Element
<i>erc</i>	
who	Creator/Contributor/Publisher
what	Title
when	Date
where	Identifier
<i>erc-about</i>	
who	<i>None</i>
what	Subject
when	Coverage (temporal)
where	Coverage (spatial)

ERC special values

Controlled element values have the form, “(:ccode)”

- e.g., missing: (:unkn) Anonymous, (:unas) Unassigned
- e.g., general: (:791) Bee Stings

Sort-friendly values keyed off of initial comma

who: , van Gogh, Vincent

who: ,Howell, III, PhD, 1922-1987, Thurston

who:, Mao Tse Tung

what:, Health and Human Services, United States Government
Department of, The,

and their equivalents in natural word order:

Vincent van Gogh

Thurston Howell, III, PhD, 1922-1987

Mao Tse Tung

The United States Government Department of Health and
Human Services

ERC dates and expansion blocks

ERC value with an “expansion” block — “%{“ and “%}”

```
where: http://foo.bar.org/node%{  
      ?db= foo  
      &start = 1  
      &end = 5  
      &buf = 2  
      &query = foo + bar + zaf  
      %}
```

is equivalent to the correct and intact URL,

where:

```
http://foo.bar.org/node?db=foo&start=1&end=5&buf=2&query=foo+bar+zaf
```

Dates are in TEMPER format

1996-2000	(range of four years)
1952, 1957, 1969	(list of three years)
1952, 1958-1967, 1985	(mixed list of dates & ranges)
20001229-20001231	(range of three days)

Kernel/ERC summary

ERC is a cheap, general-purpose metadata container

- Kernel metadata is designed to be a low-barrier way to support orderly management of collections
- Might help resource discovery and description too
- Succinct, trivial to parse, extensible yet predictable in the kernel elements

See <http://dublincore.org/groups/kernel/> for more

How to transmit an ERC? One way is with THUMP.

Thinking tiny: THUMP

The HTTP URL Mapping Protocol (THUMP)

- A set of URL-based conventions for retrieving information and conducting searches
- Can be used for focused retrievals or for broad database searches
- Based on commands put in the query string after ‘?’

`http://example.foo.com/?in(books)find(war and peace)show(full)`

THUMP requests

The HTTP URL Mapping Protocol (THUMP)

- A protocol based on HTTP and URLs
- A request is passed to a server with HTTP GET (or POST)

Shortest request is a URL ending in '?', as in

```
http://example.foo.com/object321?
```

Which is shorthand for the common request:

```
http://example.foo.com/object321?show(brief)as(anvl/erc)
```

Naked '?' and '??' designed to support the known-item query convention arising in the ARK persistent id scheme

THUMP responses

Responses consist of HTTP response headers, one record set header, and one or more ERC records

```
1  C: [opens session]
   C: GET http://ark.cdlib.org/ark:/13030/ft167nb0vq? HTTP/1.1
   C:
   S: HTTP/1.1 200 OK
5  S: Content-Type: text/plain
   S: THUMP-Status: 0.5 200 OK
   S:
   S: set-start: California Digital Library | THUMP 0.5 | 20060606161407
   S:           | http://ark.cdlib.org/ark:/13030/ft167nb0vq?
10 S:           | http://dublincore.org/groups/kernel/erc
   S: here: 1 | 1 | 1
   S:
   S: erc:
   S: who:   Stanton A. Glantz and Edith D. Balbach
15 S: what:  Tobacco War: Inside the California Battles
   S: when:  20000510
   S: where: http://ark.cdlib.org/ark:/13030/ft167nb0vq
   S: [closes session]
```

Broad searching in THUMP

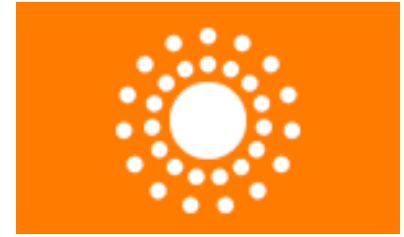
General form of broad query

`Key ? in(DB) find(QUERY) list(RANGE) show(ELEMS) as(FORMAT)`

Many details to be worked out; watch for

<http://www.ietf.org/internet-drafts/draft-kunze-thump-01.txt>

Kernel WG Wrapup



- Wrap up discussion and completing the KAP