The ADEPT Digital Library Architecture

Greg Janée
gjanee@alexandria.ucsb.edu

James Frew
frew@bren.ucsb.edu

Outline

- Goals
- Architecture
  - components, data model, services, interfaces
- Item-level metadata: buckets
  - constraint types
  - metadata mapping
  - standard buckets
- Collection discovery
- Current directions
- Status

Goals

- Digital library for georeferenced information
  - distributed
  - heterogeneous
  - rich services
  - scalable
    - **many providers**
    - **collections, large and small**
- Standard components, interfaces

Components/services

collection registry

collection-level search

thesaurus

shared vocabularies

library

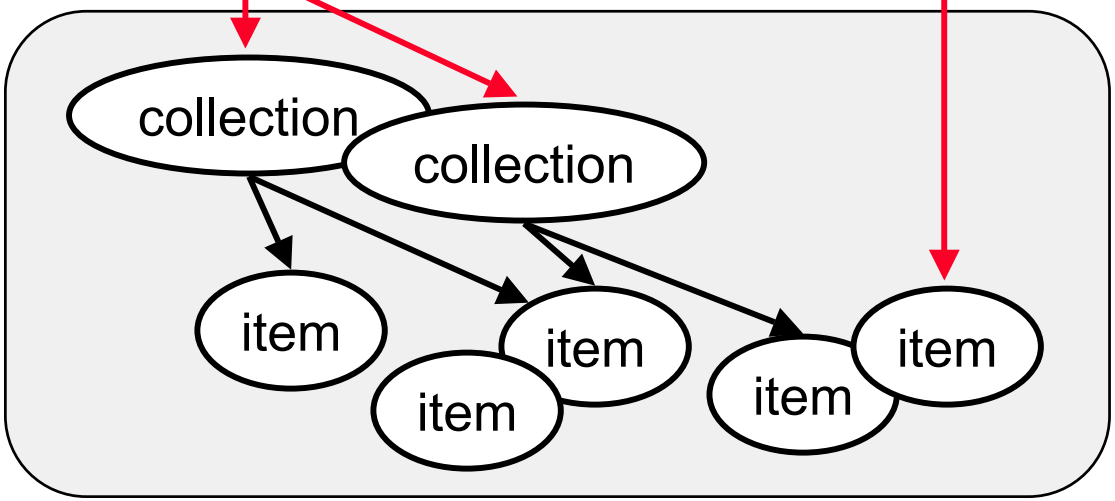
item-level search,
metadata management

content

data access

gazetteer

maps placenames
to locations



map

background imagery,
layering capability

*many interconnections
between services*

Data model

- Collection
  - name
  - static, dynamic metadata
  - set of items
  - functional behaviors
- Item
  - identifier
  - bucket view
    - **searchable metadata mapped to standard, typed buckets**
  - browse view
    - **content abstracts**
- Item, cont'd
  - access view
    - **multiple access points**
      - file-like
      - human interface
      - programmatic service
      - offline
  - other views
    - **collection- and/or item-specific**
    - **FGDC, MARC, etc.**
  - content

Library services

- configuration
- collection-metadata
  - retrieve
- item-metadata
  - retrieve views
- query
  - standard query language
- result-set
  - access server-cached query result sets
- harvest
  - collection
- collection-management
  - {create, delete, replace} static metadata
- item-management
  - {create, delete, replace} views
- reference
  - remote collection

Library server architecture

user interface

metadata mapper

harvest loader

item tracker

client interface (XML / Java,HTTP,RMI)

middleware

access control; query fan-out; query result caching & ranking
collection referencing & registration

collection interface (XML / Java)

internal collections

generic database driver

Z39.50 driver

proxy driver

collection aggregator

Bucket motivation

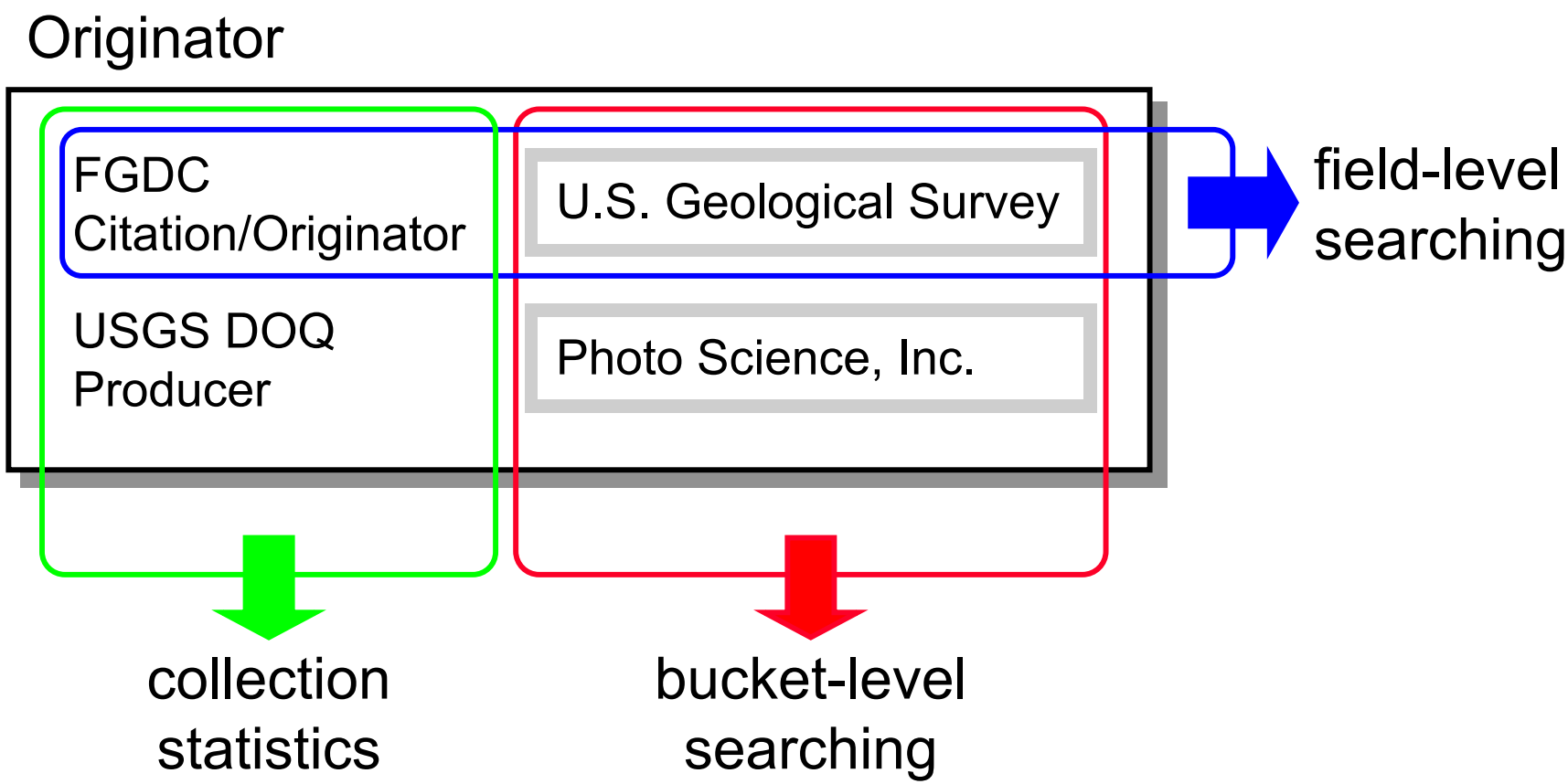
- Goals
  - heterogeneous metadata
  - uniform client services
- Typed searches
  - spatial search requires it
  - new issues
    - **validation**
    - **boolean combinations**
    - **ranking**
    - **...**

Constraint types

- Spatial
  - overlaps, contains, ...
  - lat/lon polygon, box
- Temporal
  - overlaps, contains, ...
  - date range
- Numeric
  - <, =, >, …
  - real number
  - optional unit of measure
- Textual
  - contains phrase, ...
  - word list
- Hierarchical
  - is a
  - thesaurus term
- Identification
  - matches
  - string, optionally namespace-qualified

Booleans: AND, OR, AND NOT

Bucket mapping



Standard buckets

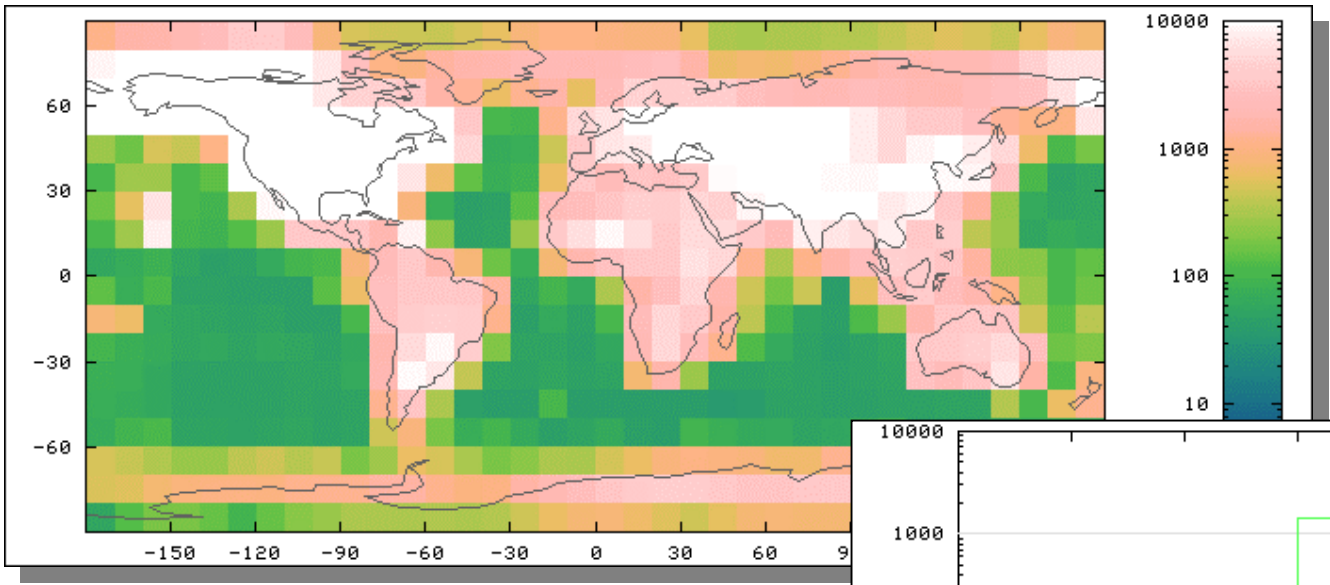
ADEPT

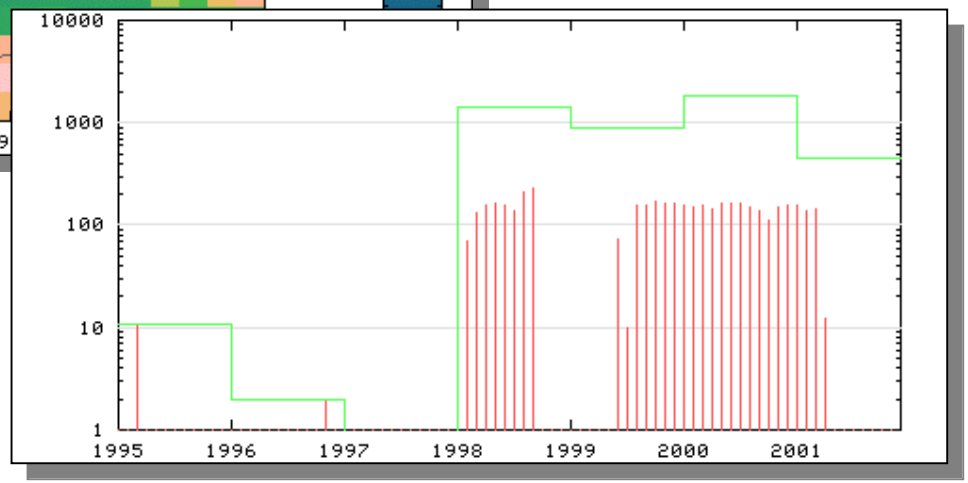
- Subject-related text
  - Title
  - Assigned term
- Originator
- Geographic location
- Coverage date
- Object type
- Feature type
- Format
- ...
- Identifier

Dublin Core

- DC.Subject
  - DC.Title
  - DC.Subject (qualified)
- DC.Creator + DC.Publisher
- DC.Coverage.Spatial
- DC.Coverage.Temporal
- DC.Type
- -
- DC.Format
- -
- DC.Identifier

Collection-level metadata





Object Type	Count
cartographic works	324,876
maps	324,876
images	2,014,799
photographs	484,083
aerial photographs	484,083

Collection discovery

- Collection registry polls known library servers
- Relevance model
  - binary
  - more is better
- Query language
  - range searching over space, time, vocabulary terms
  - subset of item-level query language
- Limitations
  - no joint constraint conditions
  - no text statistics à la STARTS
  - multiple, overlapping vocabularies

Current directions

- Lowering the barrier
  - metadata management services
  - OAI harvest loader
  - improved packaging
- Service aggregation via harvesting
- Content-based searches, ranking
  - text IR, image texture
- Collection discovery
- Integration with access mechanisms
- Client development
  - custom
  - embedded

Summary

- Distributed, service-based architecture
  - two search levels
  - heterogeneous, native metadata
  - rich, uniform services
- Status
  - basis of UCSB MIL operational library
    - **http://webclient.alexandria.ucsb.edu**
  - downloadable
    - **http://www.alexandria.ucsb.edu/middleware**
  - initial full version late 2002