



## HarvestChoice:

Developing Biblio-spatial Integrations for Search

Phil Pardey, Chad Fennell, Grant McKenzie, John Butler



# **Outline**

- HarvestChoice Project Background
- Portal Architecture Overview
- GeoNetwork & the HC Implementation



Evaluates technologies and strategies to inform policy and investment choices designed to raise the productivity of agricultural systems most beneficial to the poor.



- Multi-year project funded by the Bill and Melinda Gates Foundation
- Additional support from USAID, IPM CRSP, HarvestPlus, International Food Policy Research Institute (IFPRI), Generation Challenge Programme, University of Minnesota, others...
- Web: harvestchoice.org



# **Activities**

- Macro trends on human welfare and crop systems
- Cross-country synthesis of micro (household/farmscale census/survey) data; hh characterization, food consumption, crop production, technology adoption, marketing
- Application of evaluation tools for assessing the productivity, market, and welfare consequences of change (through, e.g. policy change and investments)
- Assessment of the commercialization potential of new technologies



# **Activities**

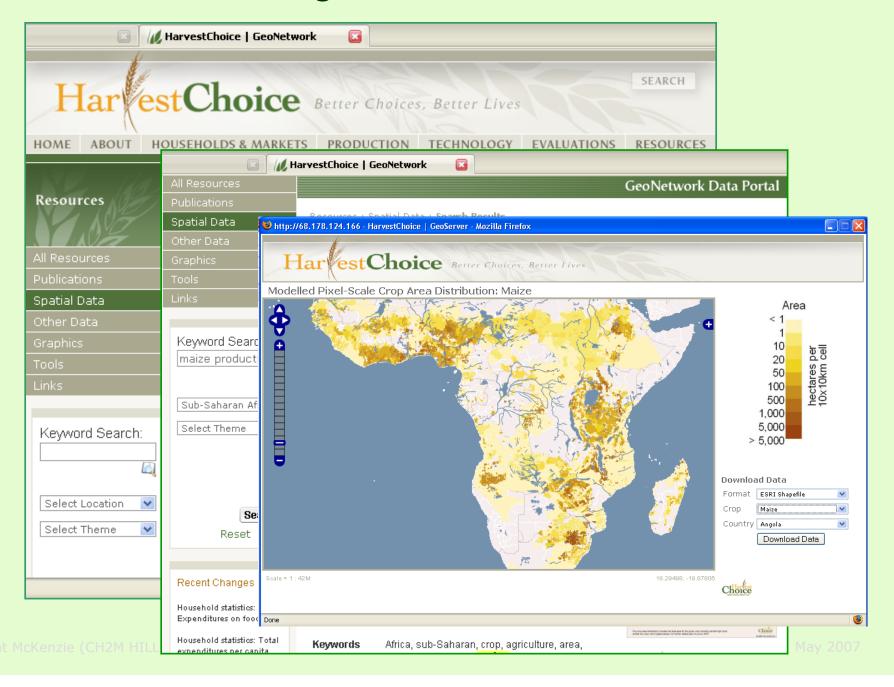
- Development of an advanced Web portal for access, query and retrieval of integrated, bibliographic, map, tabular, graphic resources
- Harmonization of GIS/Spatial outputs across all Gates grantees



## **Web Portal Objectives**

- Blend spatial data, flat files and curated bibliographic collection via a Drupal/Solr platform integrated with a GeoNetwork platform for accessing spatial and mapped data
- Facilitate online data and bibliographic submissions to the web site collection
- Share primary data, simulation results, and other information online, using open access principles
- Move FAST compressed development timeline

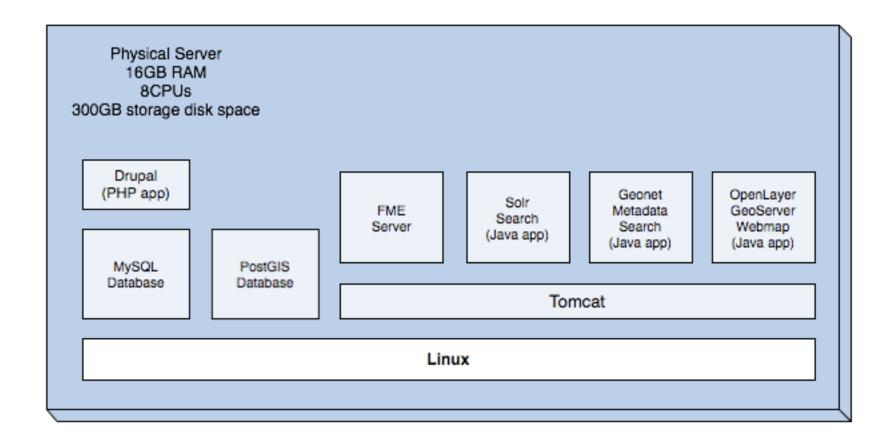
### www.harvestchoice.org





# Web Portal Architecture

# Project Stack





An "open source enterprise search server based on the Lucene Java search library, with XML/HTTP and JSON APIs, hit highlighting, faceted search, caching, replication, and a web administration interface. It runs in a Java servlet container such as Tomcat."

http://lucene.apache.org/solr/#intro

# Why Drupal?

- "Drupal is the Answer to Everything…"
  - Will Snow (Stanford University), DLF Spring '08
- Undemanding Technology Stack
- Active Open Source Developer Community
  - Drupal 5.x = 2,000 + Modules
    - Core & Contributed (Comments, Solr, OAI, Shiboleth)
    - Flexible API to Reach Into and Add Onto
- "Eliminate the Programmer" Dries Buytaert
  - GUI Web Form Construction, Query Generators, Output Formatters



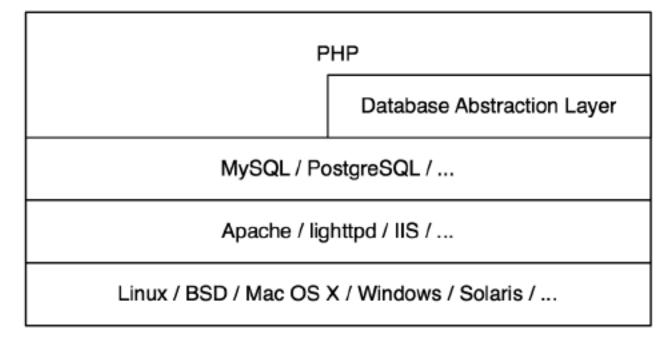
# **Drupal Stack**

Language

Database

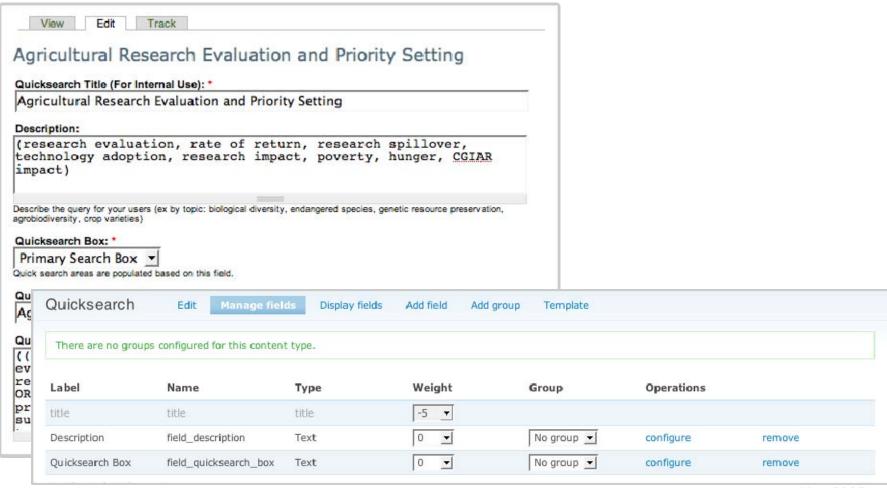
Web Server

Operating System

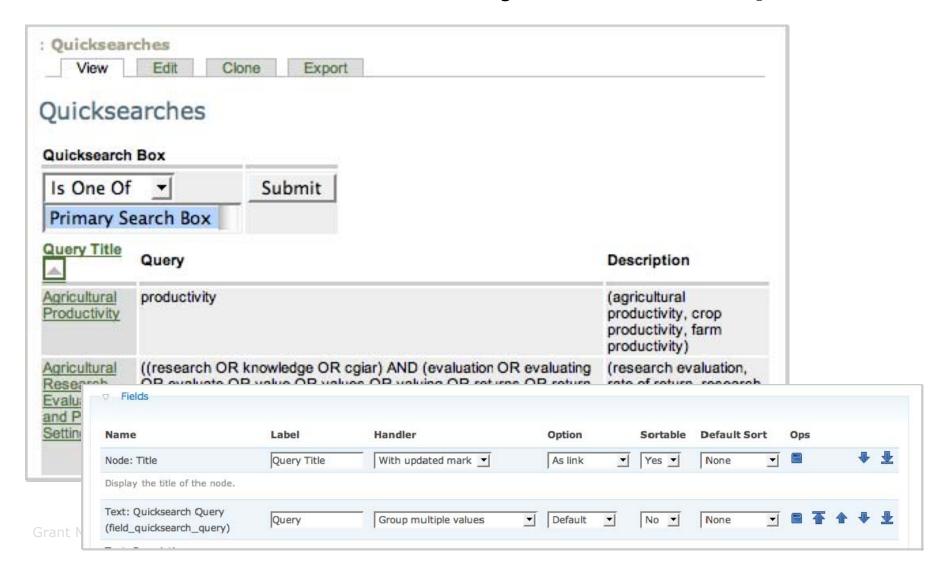


http://www.ddj.com/linux-open-source/199201603

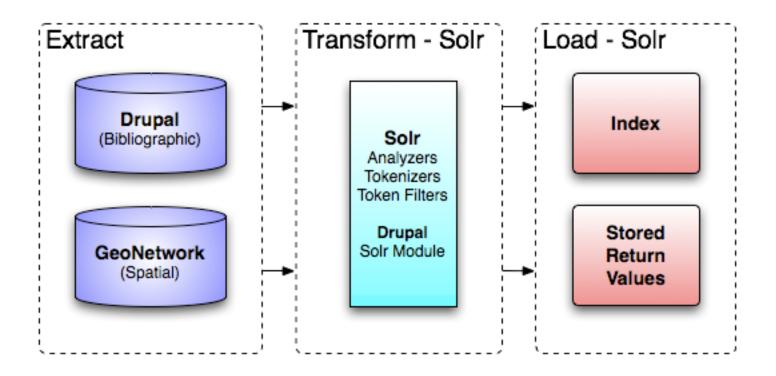
# Eliminating the Programmer with "CCK" - Web Form Construction UI



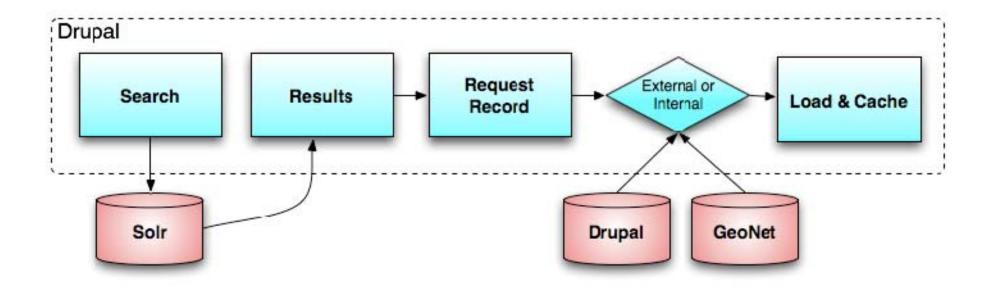
# Eliminate the Programmer with "Views" - Query and Output



# Indexing (Extract, Transform, Load "Light")



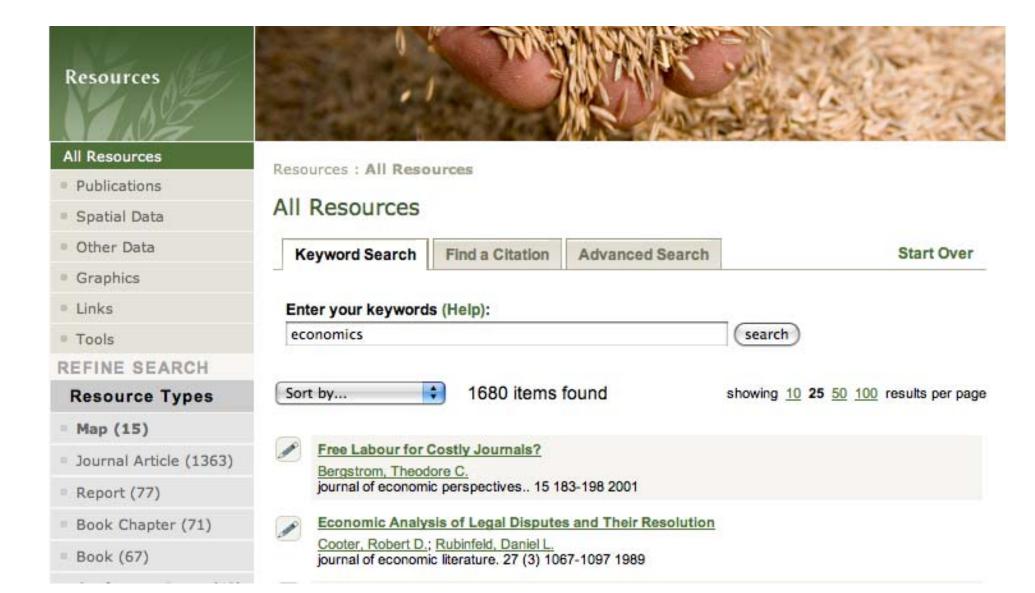
# Search & Retrieval



# Next Steps for This Model

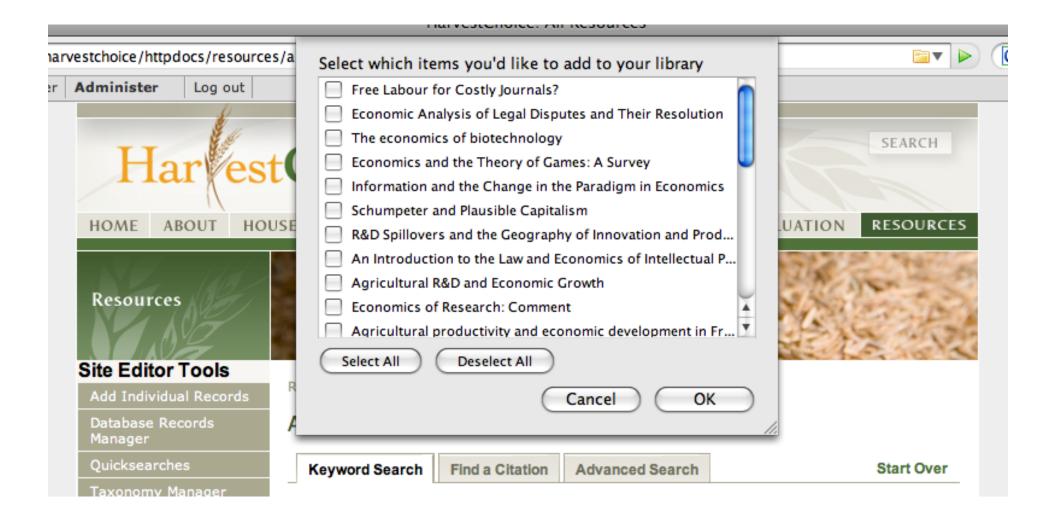
- Search Sprint Minnesota May 2008
  - New Pluggable Search API for Drupal
- End-User Content Contribution
  - Workflow, Duplicate Records, Quality Control
- Revise Data Model
  - Migrate Module Functionality to More Generalized Model
  - Extract, Transform, Load Formalize & Extend
- Share the Costs and the Benefits
  - University of Rochester eXtensible Catalog

# Let's Search!



# Single Citation

Resources : All Resources All Resources Advanced Search **Keyword Search** Find a Citation Start Over - Find an Individual Citation -Journal: Year: YYYY-YYYY (Second Year Is Optional) Volume: Issue: Author: Single Author in Format: Nagy, Joseph G Title: search



Spatial Data

Other Data

Graphics

Links

Tools

#### REFINE SEARCH

#### Resource Types

Map (15)

#### Subjects

- Household (14)
- Surveys (14)
- Africa (10)
- Sub-saharan (10)
- Sub-saharan Africa (5)
- + show more

#### Recently Added

Identifying and managing the conflicts between agriculture and biodiversity

#### All Resources

**Keyword Search** 

Find a Citation

**Advanced Search** 

Start Over

#### Enter your keywords (Help):

economics

search

Map (Resource Type)

Sort by...

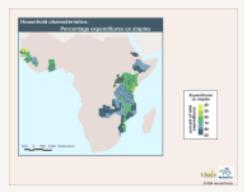
15 items found

showing 10 25 50 100 results per page



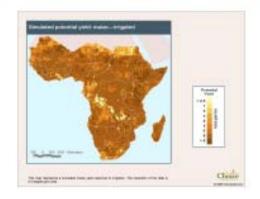
#### Household statistics: Percentage of expenditures on staples

This dataset is a part of products provided by HarvestChoice (http://harvestchoice.org) that undertakes broad-ranging evaluation of technologies and strategies to inform policy and investment choices designed to raise the productivity of the agricultural systems most beneficial to the poor. The dataset describes the percentage of expenditures on staples in each household at subnational administrative units - 2nd or 3rd levels depending on countries. Aggregated value (unit: percentage) and its standard error for a given administrative unit are presented in AVERAGE and STANDARD\_E columns, respectively. We have obtained data from a variety of household surveys performed in Sub-Saharan African over recent decades. So far, data have been compiled for



15 countries in SSA and South Asia, but we expect to add more as more data become available. Typically, these surveys aimed to provide information on specific elements of households, such as economic, nutritional, social indicators of households in the various countries. In addition, many of these surveys also provided consumption and production information of agricultural products. This dataset is a part of the HarvestChoice Household Characteristics Database series.

### Household statistics: Percentage of expenditures on staples



View Full Record

View / Download Data

#### Summary

Reference System	GCS_WGS_1984		
Abstract	This dataset is a part of products provided by the HarvestChoice (http://harvestchoice.org) project that undertakes broad-ranging evaluation of technologies and strategies to inform policy and investment choices designed to raise the productivity of the agricultural systems most beneficial to the poor. This dataset describes the simulated maize yield (t/ha) in the irrigated/high-input cropp system in Sub-Sahara Africa at 30' grid resolution. As a preliminary analysis, the irrigated/high-input cropping system was arbitrarily characterized with nitrogen fertilization at 100 kg[N]/ha, irrigation at 30 mm[water]/ha/week, and access to improved germplasm with medium maturity duration. This dataset is a part of the HarvestChoice Crop Systems Evaluation series.		
Contact	Jawoo Koo		
Organization	IFPRI / HarvestChoice		



# GeoNetwork & the HarvestChoice Implementation

### **GeoNetwork Opensource**



#### **Primary Contributors**

- Food and Agriculture Organization of the UN (FAO)
  - The UN World Food Programme (WFP)
  - The UN Environmental Programme (UNEP)



### What is GeoNetwork?

Web Based Geographic Metadata Catalog System.



Grant McKenzie (CH2M HILL)

### **GeoNetwork Architecture**

- Java language (platform independent)
- Servlet environment (Tomcat or Jetty)
- XML data representation
- XSL templates for presentation
- JDBC for database access (MySQL, PostgreSQL, Oracle)
- Apache Lucene search capabilities
- XML Schema for metadata editing



### **GeoNetwork Standards**

#### Metadata

- ISO 19115/19139
- FGDC Federal Geographic Data Committee
- Dublin Core

#### **Data**

OGC Web Map Services (WMS) – GeoServer / Intermap

#### Search

- OGC Catalogue Service for the Web (CSW)
- z39.50 with GEO profile

#### **GeoNetwork Metadata Exchange Format**

- Spatial Data
- Metadata
- Preview Images
- Basic Privileges
- Other information

### **GeoNetwork Harvesting**

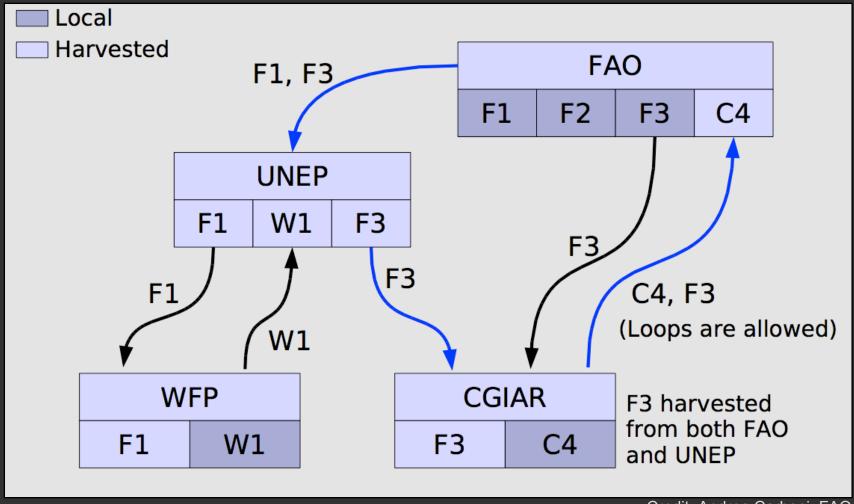
• ... the process of collecting remote metadata and storing it locally.

#### **GeoNetwork Harvesting Capabilities**

- A web DAV server
- A CSW 2.0.1 server
- Old GeoNetwork 2.0 nodes
- An OAI-PMH Server



### **GeoNetwork Harvesting**



Credit: Andrea Carboni, FAO

### **GeoNetwork & HarvestChoice**

Applications	FME Server	GeoNetwork	OpenLayers	
			GeoServer	
Languages	PHP 5 / XSLt			
Web Server	Apache & Apache Tomcat (Java)			
Database	PostgreSQL / PostGIS			
Operating System	Red Hat Enterprise Linux			

Grant McKenzie (CH2M HILL)