Applying text mining services to facilitate discovery and linking of Wheat Scientific Information and Data

Agroknow

We find, connect and deliver agriculture & food information worldwide

(for the past 8 years)

So yes, we consider ourselves pioneers


We help organizations to build
a better future for all of us
using open agriculture and
food data

We work with major institutions and initiatives

OUR FIRST RULE AT WORK

We look at each project separately, focusing on the needs of people in the institution and helping them deliver their vision

But how we do that?

We are a flexible, intuitional team of experts that delivers

The **Strategist**

Nikos Manouselis

Expert in strategy design

The **Architect**

Giannis Stoitsis

Designs the most efficient solutions

The **Perfectionist**

Babis Thanopoulos

Expert in customer needs identification and analysis

The **Doer**

Kostas Kastrantas

Expert in data and technology solutions

The rest of the Dream Team

Blending domain experts, computer engineers and information

Mihalis Papakonstadinou

Panagis Katsivelis

Thodoris Kontogiannis

Jenny Kesoglidou

Vassilis Tsantilis

Panagiotis Siokas

Antonis Koukourikos

Christina Vichou

Panagiotis Zervas

Maritina Stavrakaki

Pythagoras Karampiperis


3
Success Stories


AGRIS


8 million
bibliographic records
350.000
visitors per month
Serving 200 data providers
20% of the total FAO visits

AGRIS

SEARCH TECHNOLOGIES

Google Scholar

ProQuest

USDA

Wolters Kluwer
When you have to be right

ExLibris
The bridge to knowledge

EBSCO
INFORMATION SERVICES

OpenAIRE

We aggregate and serve agricultural scientific information to maximise the impact of AGRIS


TAPipedia
A G20 initiative

tap
a G20 initiative

Login | Register

tap TROPICAL AGRICULTURE PLATFORM

Home About Resources TAPipedia Network TAP Framework Contact

TAPipedia

Exchanging knowledge and supporting capacity development for agricultural innovation systems

PARTNERS

GWPP

Global Water Pathogen Project

GWPP
Global Water Pathogen Project

UNESCO, P&G and Gates Foundation are supporting the Global Scientific Knowledge Hub for Water Pathogens

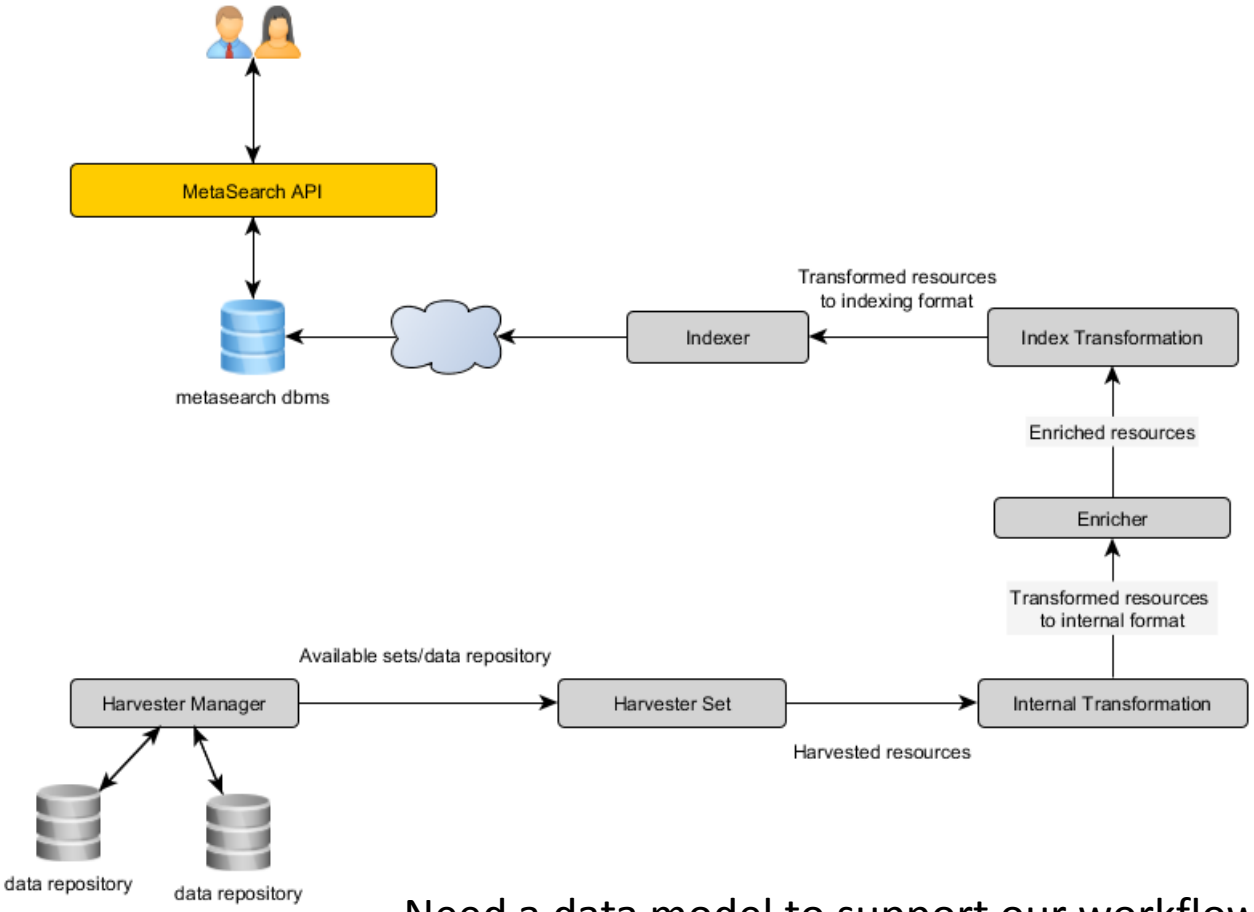
Specific Use Cases

- Let us consider a real-world problem.
- Consider an organization (perhaps yours?) having research information and data in different databases.
- How could we connect these data silos?
- Need to:
  - Define a workflow,
  - Design a data model,
  - Implement it!

Proposed Solution

- Develop a layer, running on top of the data silos:
  - Harvesting the data stored,
  - Aligning them in a uniform internal format,
  - Enriching/Interlinking them with external systems,
  - Indexing the enriched data,
  - Providing them back through a search api.

Proposed Solution – Complete Picture

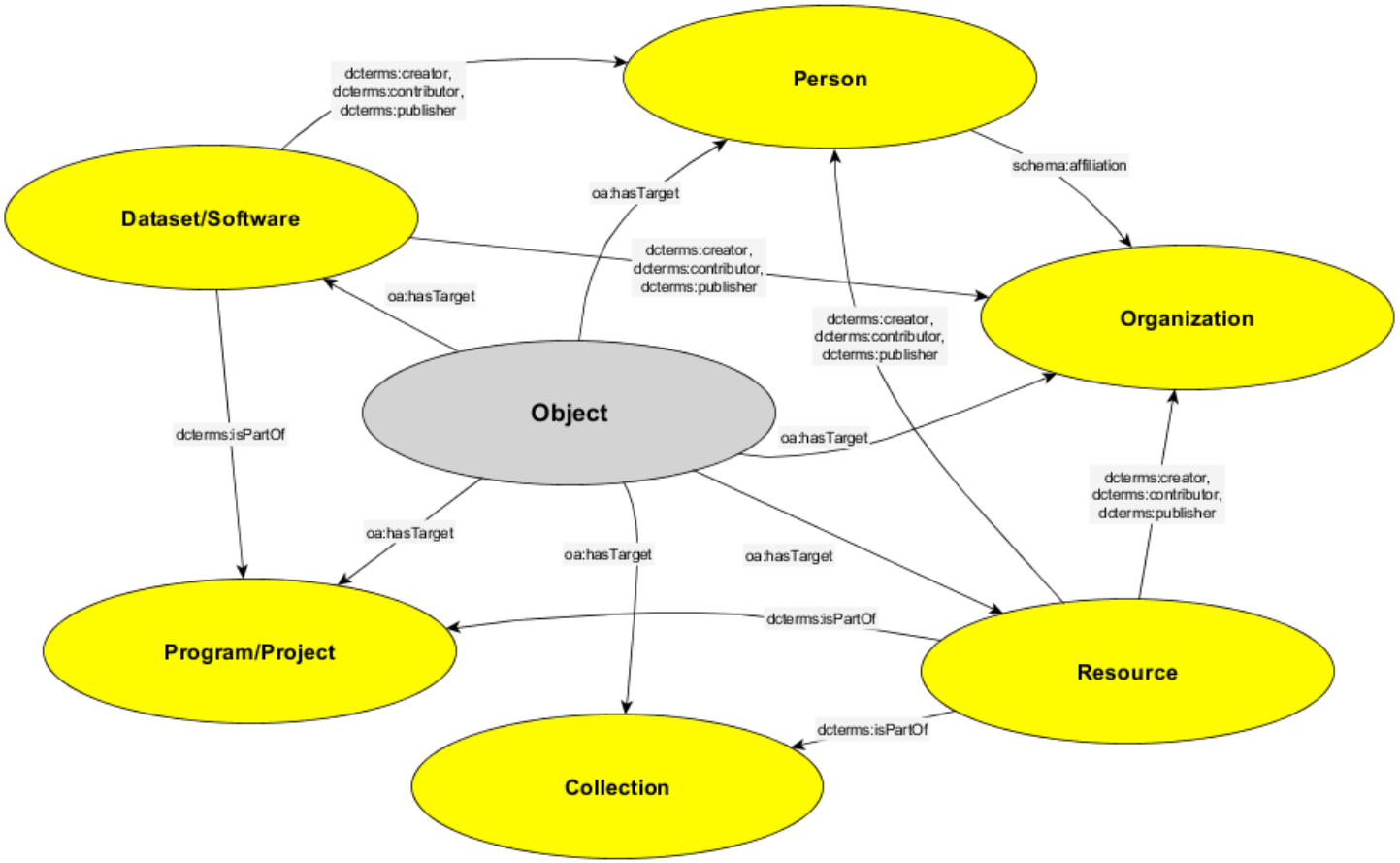


Need a data model to support our workflow

Data Model

- Everything is an object.
- Each object has a specific type, with different properties.
- Of course everything is interlinked.
- Any new content type can be added on the second level with specific properties.

Data Model



Important: Enrichment Process

Enrichment Process

- The process:
  1. use raw information stored in resources (title, abstract, author/publisher list, full-text etc.),
  2. recognize entities using various endpoints,
  3. interlinking them with both internal entities and external systems.

Enrichment Process Example – Initial State

The potential for wheat production in Africa: analysis of biophysical suitability and economic profitability

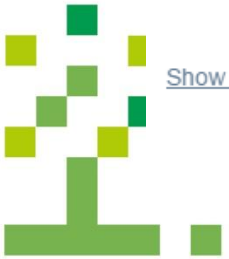
Title:	The potential for wheat production in Africa: analysis of biophysical suitability and economic profitability
Author:	Asfaw Negassa; Shiferaw, B.; Koo, J.; Sonder, K.; Smale, M.; Braun, H.J.; Gbegbelegbe, S.; Zhe Guo; Hodson, D.P.; Wood, S.; Payne, T.S.; Abeyo Bekele Geleta
Year:	2013
Copyright:	CIMMYT manages Intellectual Assets as International Public Goods. The user is free to download, print, store and share this work. In case you want to translate or create any other derivative work and share or distribute such translation/derivative work, please contact CIMMYT-Knowledge-Center@cgiar.org indicating the work you want to use and the kind of use you intend; CIMMYT will contact you with the sutable license for that purpose.
Program:	Genetic Resources Program; Socioeconomics Program; Global Wheat Program
Pages:	viii, 64 p.
Place:	Mexico, DF (Mexico)
Publisher:	CIMMYT
Citation:	The potential for wheat production in Africa: analysis of biophysical suitability and economic profitability. 2013. Asfaw Negassa; Shiferaw, B.; Koo, J.; Sonder, K.; Smale, M.; Braun, H.J.; Gbegbelegbe, S.; Zhe Guo; Hodson, D.P.; Wood, S.; Payne, T.S.; Abeyo Bekele Geleta. : viii, 64 p.. Mexico, DF (Mexico). CIMMYT.

Show full item record

Enrichment Process Example – Entity Recognition

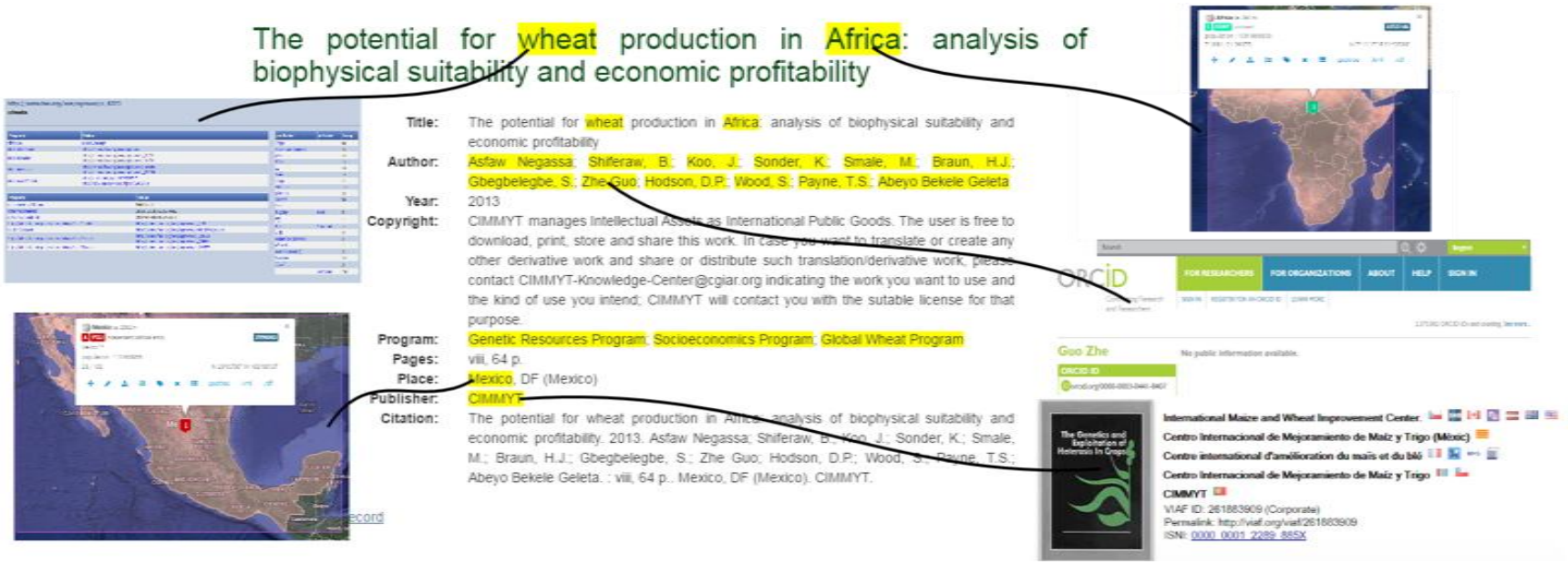
The potential for wheat production in Africa: analysis of biophysical suitability and economic profitability

Title:	The potential for wheat production in Africa: analysis of biophysical suitability and economic profitability
Author:	Asfaw Negassa; Shiferaw, B.; Koo, J.; Sonder, K.; Smale, M.; Braun, H.J.; Gbegbelegbe, S.; Zhe Guo; Hodson, D.P.; Wood, S.; Payne, T.S.; Abeyo Bekele Geleta
Year:	2013
Copyright:	CIMMYT manages Intellectual Assets as International Public Goods. The user is free to download, print, store and share this work. In case you want to translate or create any other derivative work and share or distribute such translation/derivative work, please contact CIMMYT-Knowledge-Center@cgiar.org indicating the work you want to use and the kind of use you intend; CIMMYT will contact you with the sutable license for that purpose.
Program:	Genetic Resources Program; Socioeconomics Program; Global Wheat Program
Pages:	viii, 64 p.
Place:	Mexico, DF (Mexico)
Publisher:	CIMMYT
Citation:	The potential for wheat production in Africa: analysis of biophysical suitability and economic profitability. 2013. Asfaw Negassa; Shiferaw, B.; Koo, J.; Sonder, K.; Smale, M.; Braun, H.J.; Gbegbelegbe, S.; Zhe Guo; Hodson, D.P.; Wood, S.; Payne, T.S.; Abeyo Bekele Geleta. : viii, 64 p.. Mexico, DF (Mexico). CIMMYT.



Show full item record

Enrichment Process Example – Entity Interlinking



Some statistics using this process (and only 3 endpoints):

3.9K
Resources integrated

33.9K
Entities extracted

270K
enrichments

Useful Endpoints (1/2)

- **FREME API**, used for topics extraction and annotation (against AGROVOC), and entity recognition (person, organization, location),
- **Geonames API**, for location extraction and interlink,
- **OpenAIRE mining service** (part of the OpenMinTed project), can be used to mine projects from text, data citation, classification, etc.

Useful Endpoints (2/2)

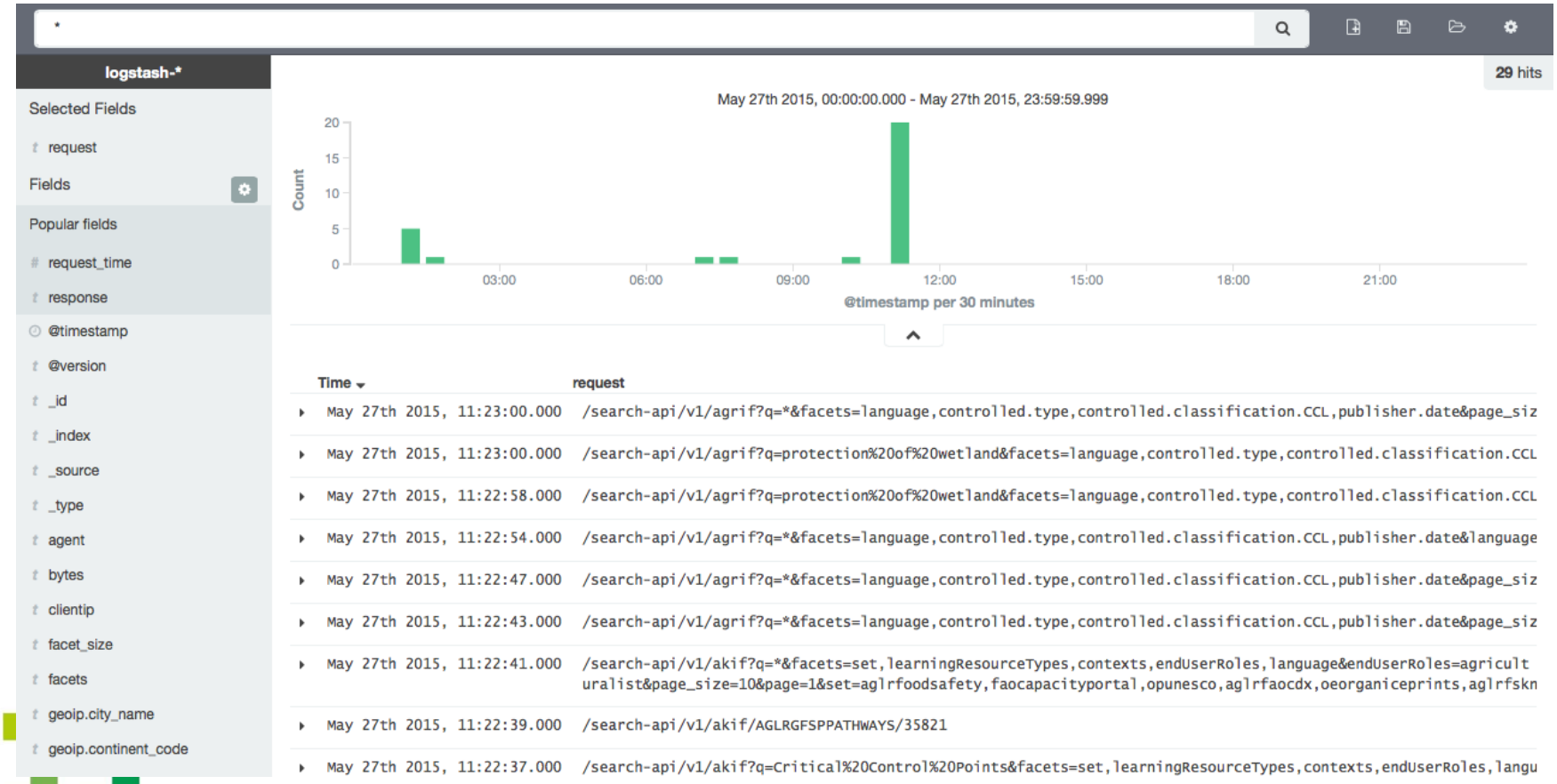
- **CropOntology**, can be used to extract wheat trait entities,

- **PDF Text Extraction (and annotation) service**, used to extract text from pdf files and annotate it using various endpoints (1st prize in 1st AgroHackathon, Montpellier, 29/6-1/7/2016)

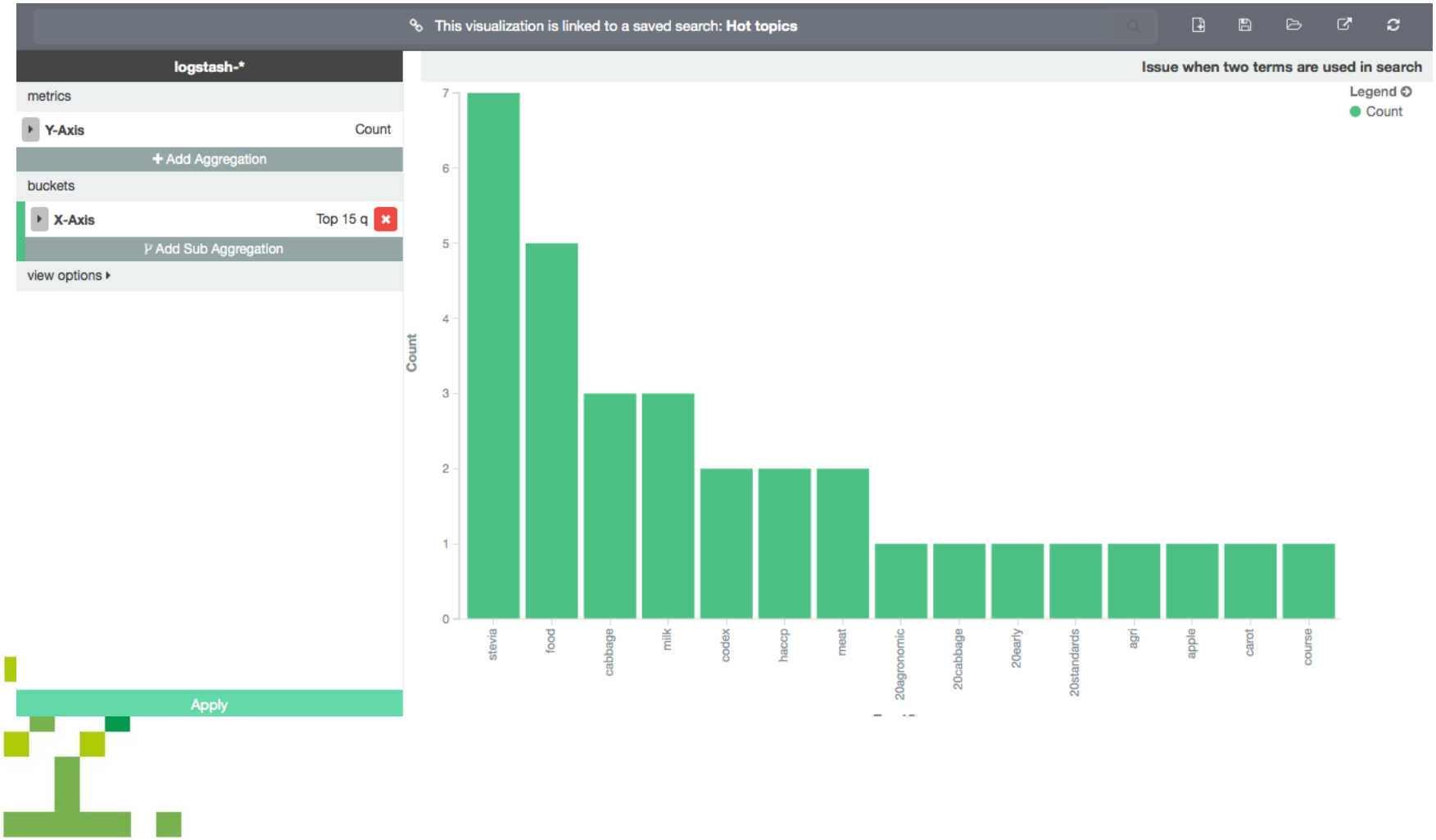
Outcomes Using these Technologies

- AKIF Search API
- CIMMYT MetaSearch API

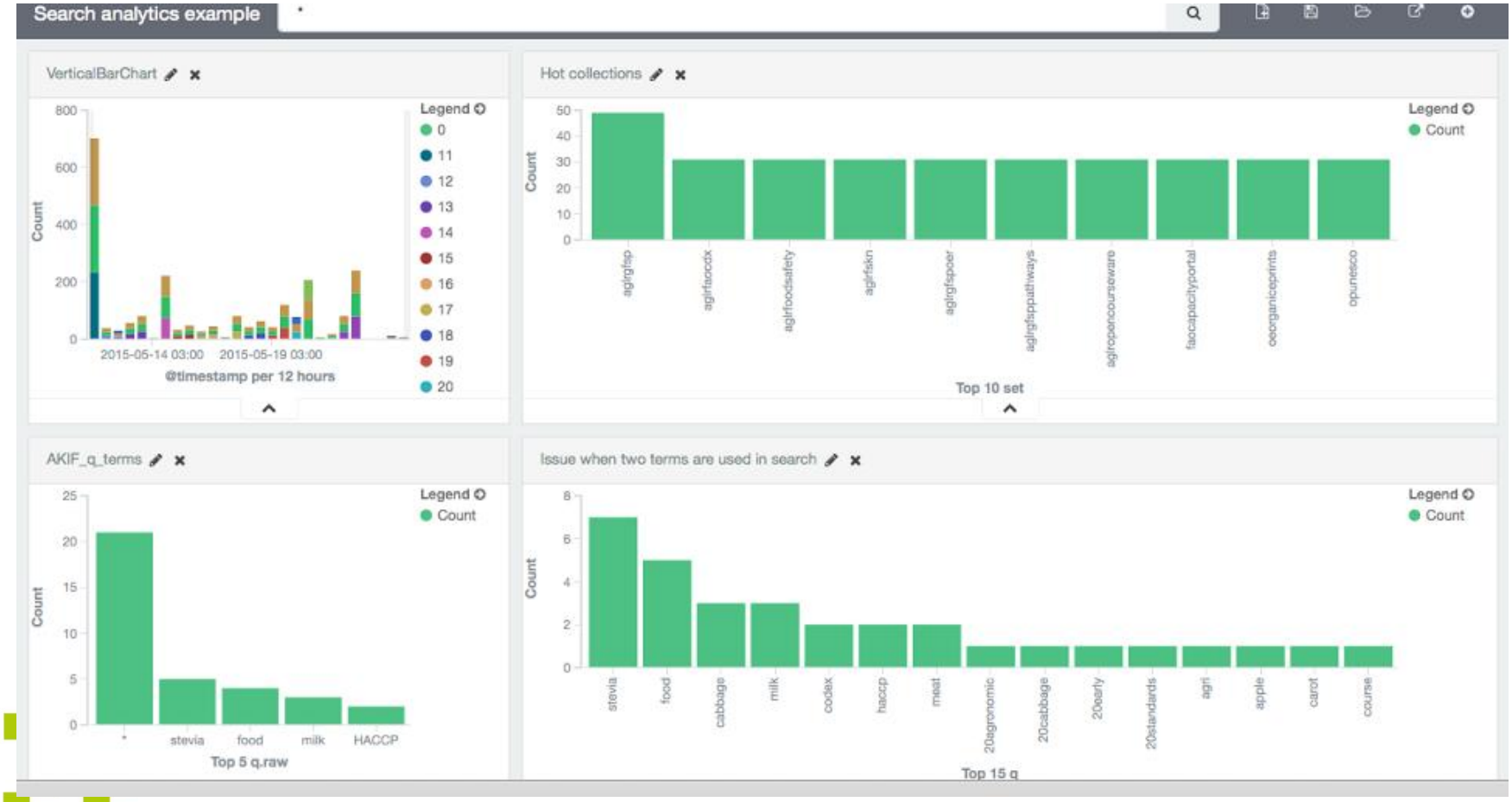
Analytics (1/3)



Analytics (2/3)



Analytics (3/3)




Thank you