openMINTED

@openminted_eu

Applying Text Mining Services to Facilitate Discovery and Linking of Wheat Scientific Information and Data

Agroknow

Applying Text Mining Services to Facilitate Discovery and Linking of Wheat Scientific Information and Data

The scope of this presentation is to demonstrate the process of applying text mining services to support discovery and inter-linking of wheat scientific information. This is supported by a set of useful endpoints that can be used for in this process

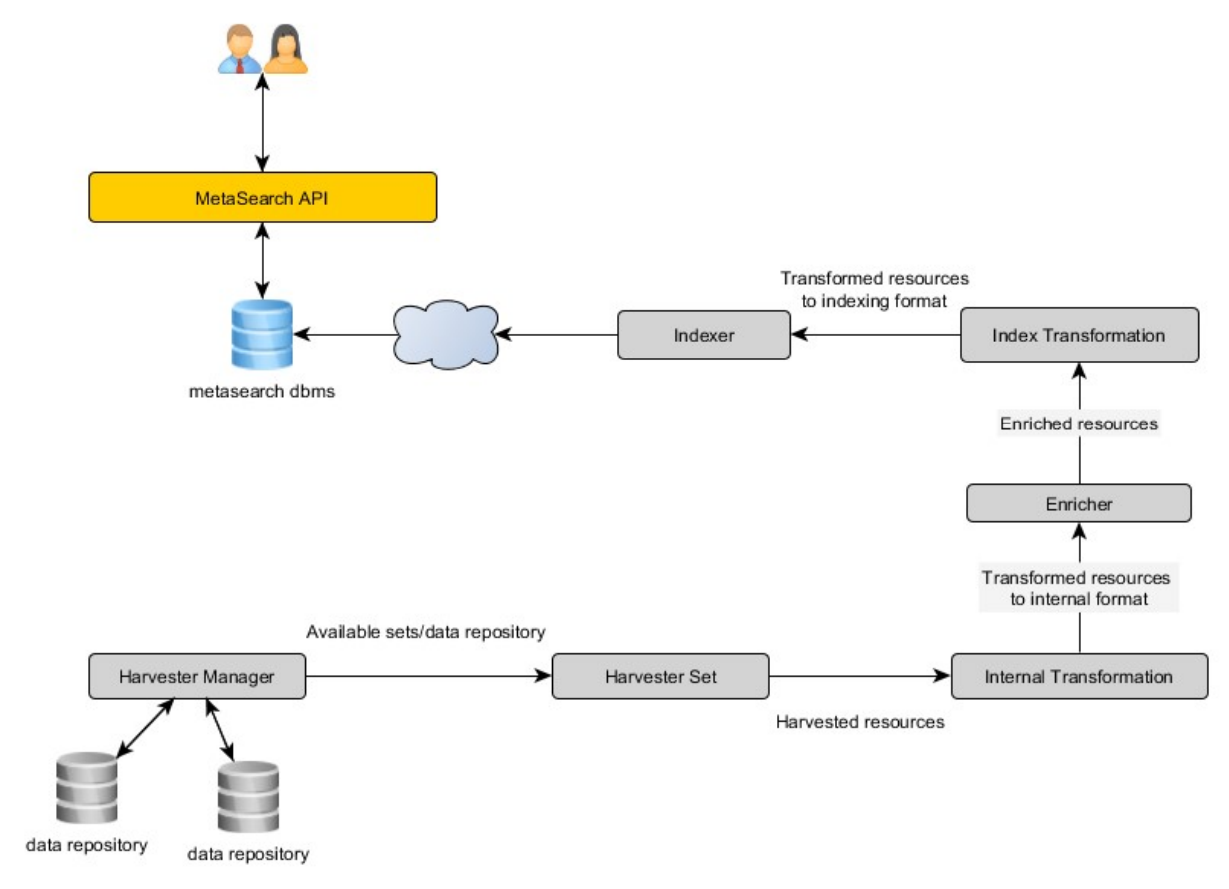
Specific Use Cases

- 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

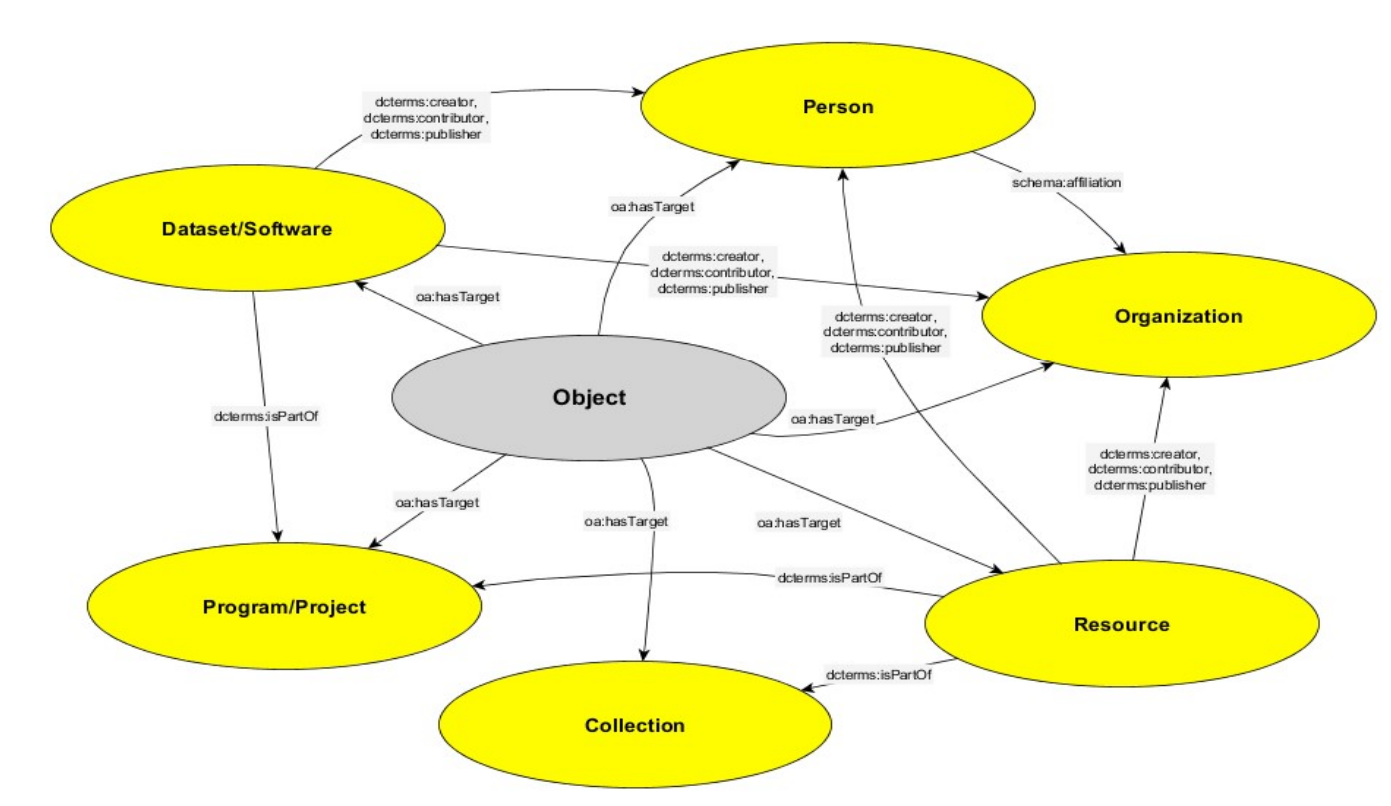


Need a data model to support our workflow

Data Model (1/2)

- 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 (2/2)



Enrichment Process

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

Enrichment Process – 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

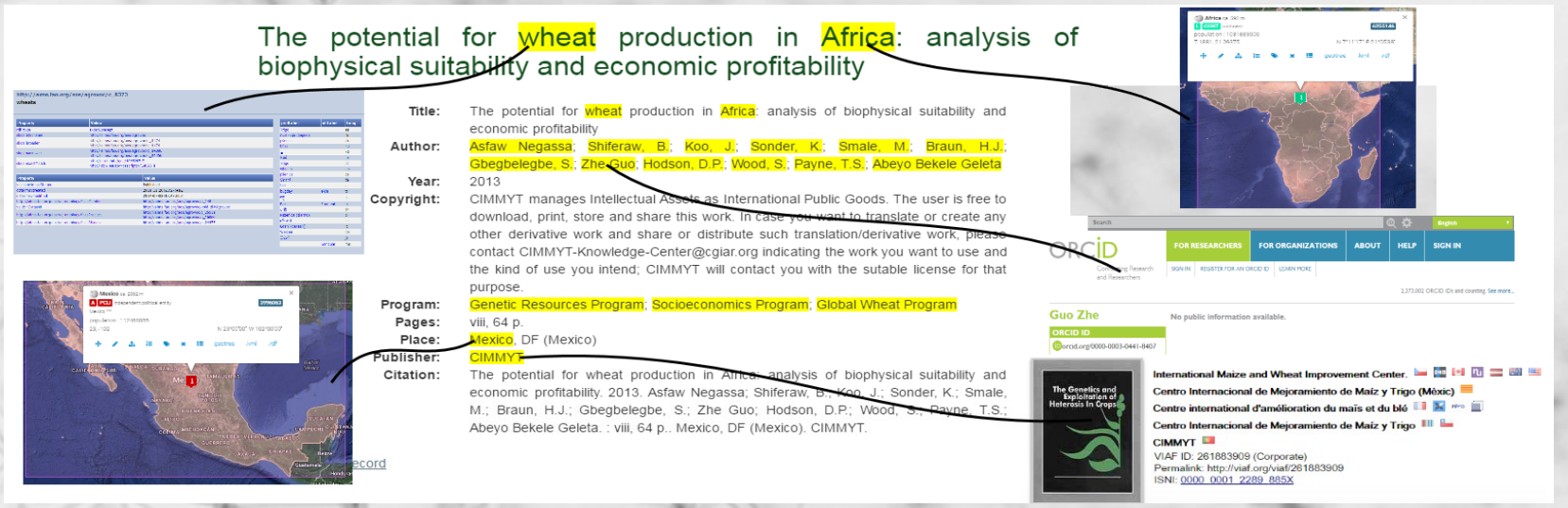
Enrichment Process – Entity Recognition

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 – 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.

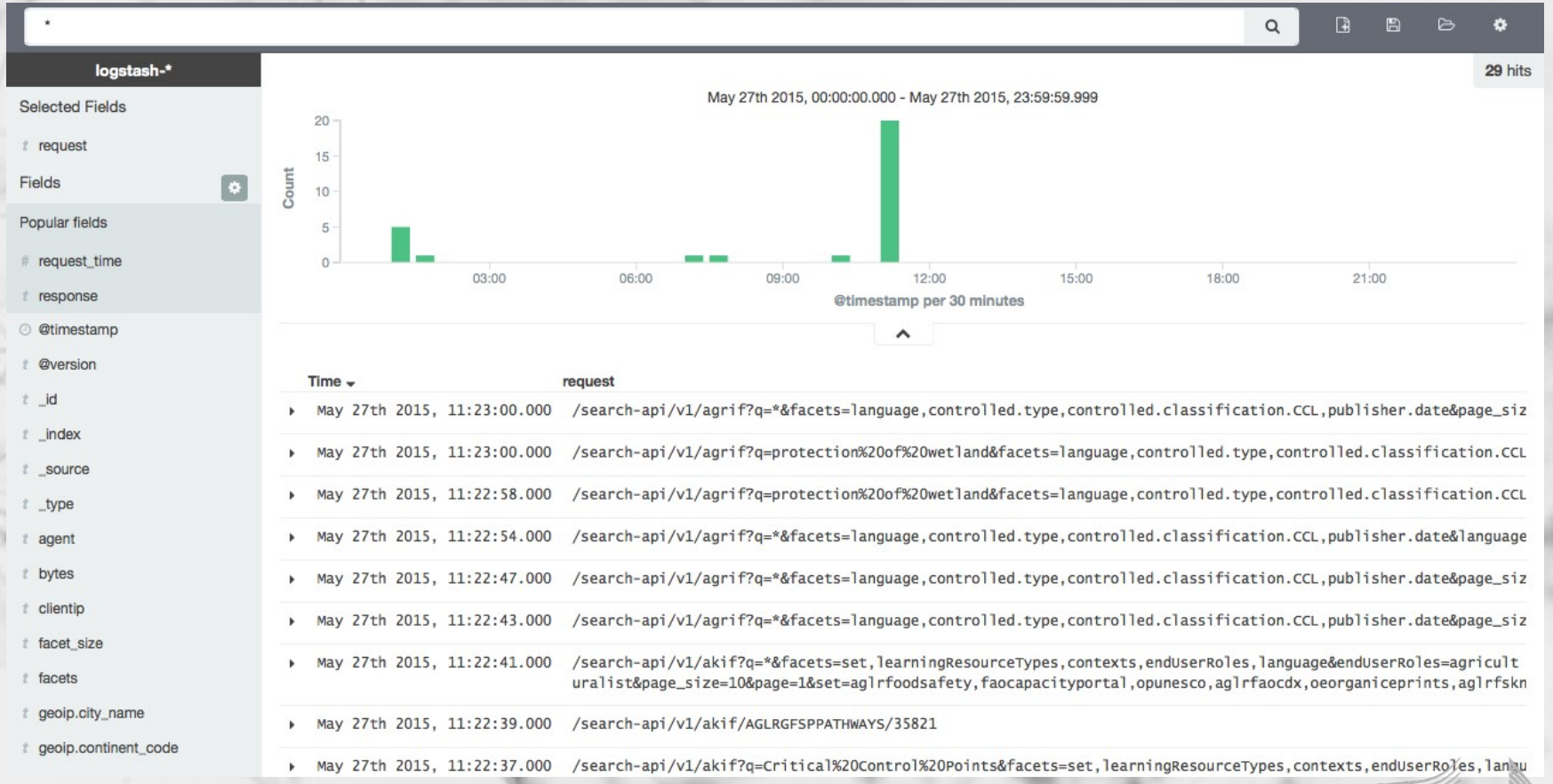
Useful Endpoints (2/2)

- **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)

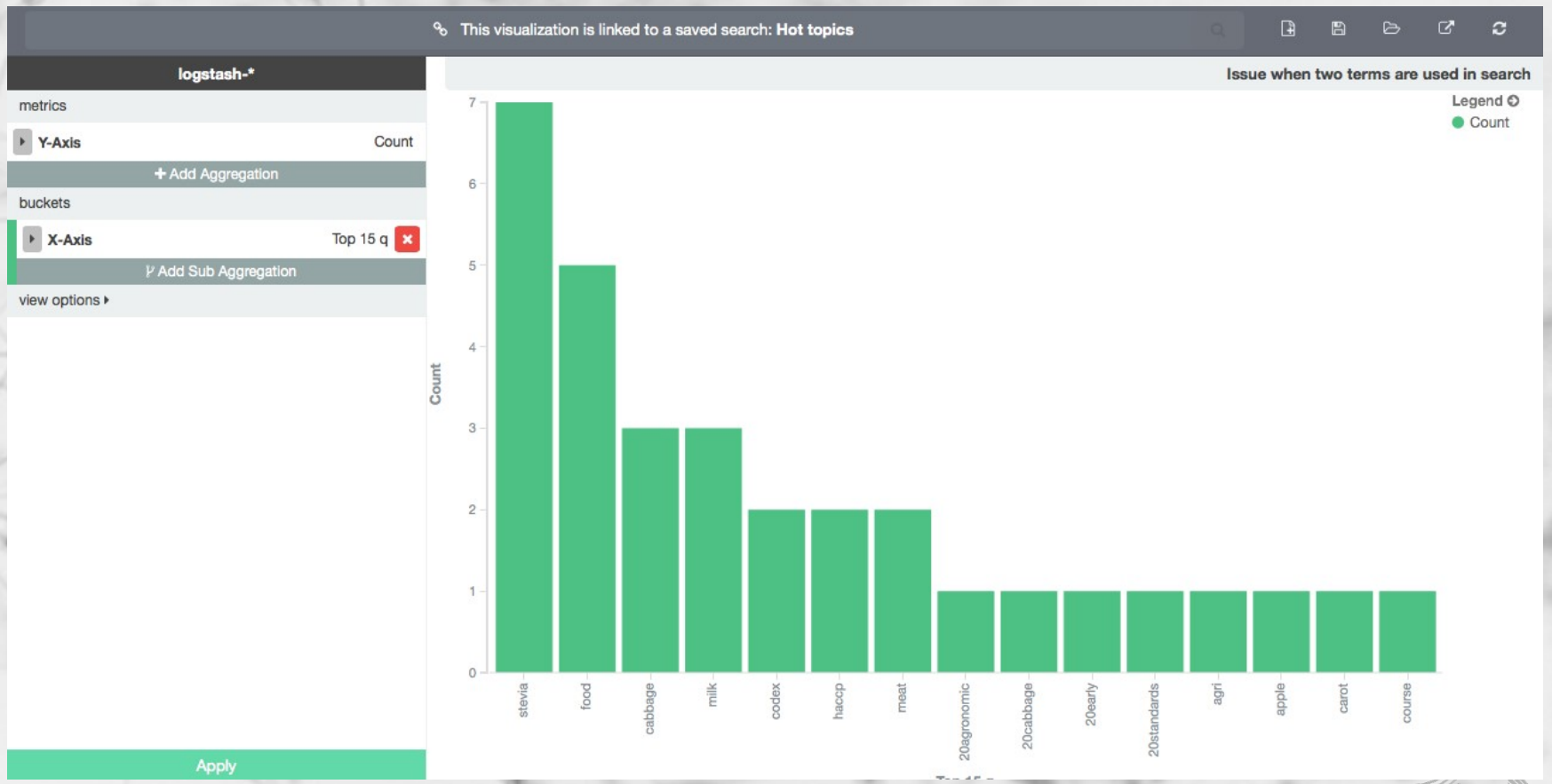
Outcomes Using these Technologies

- **AKIF Search API**
- **CIMMYT MetaSearch API**

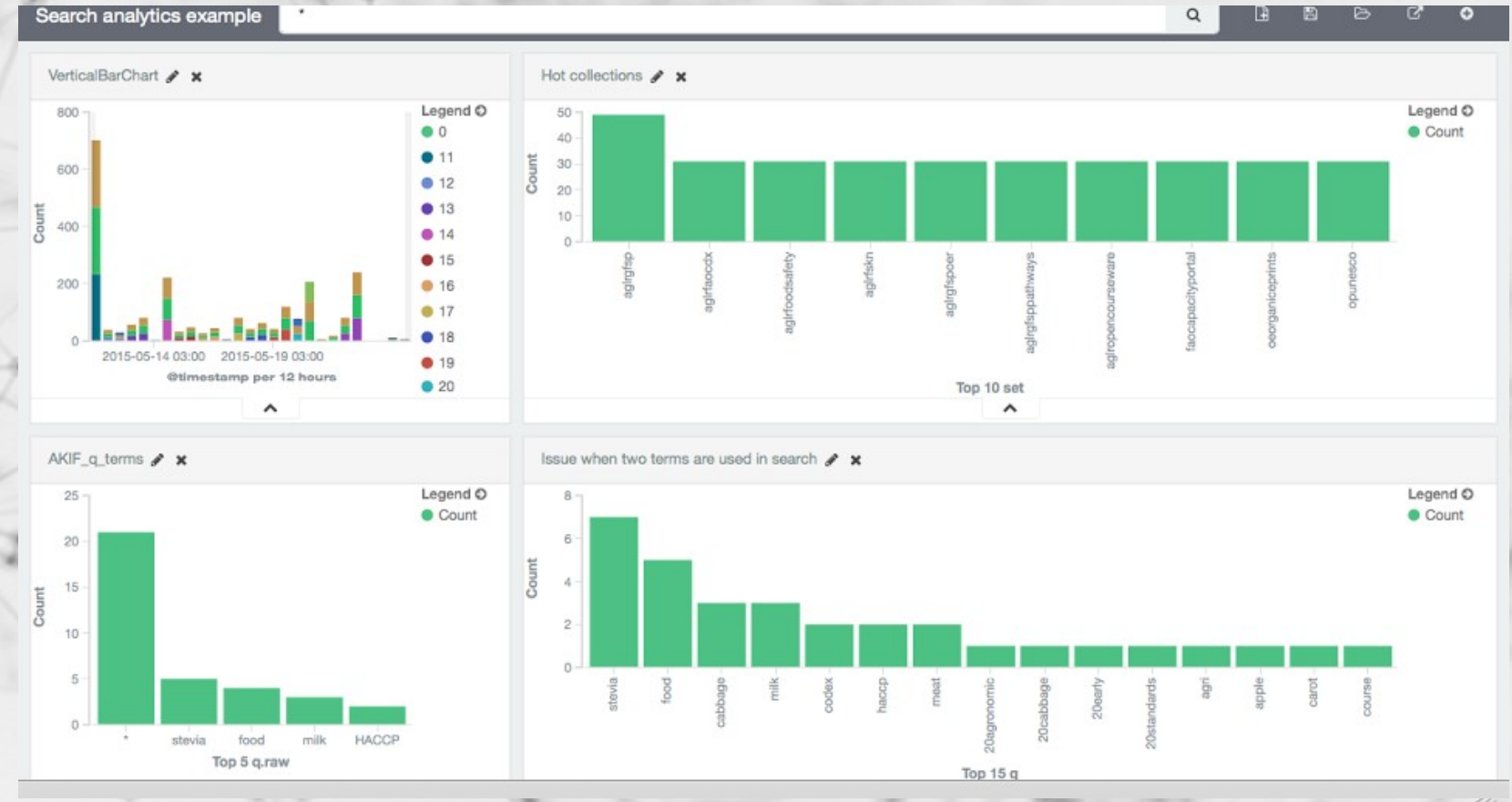
Analytics (1/3)



Analytics (2/3)



Analytics (3/3)



openMINTED

twitter.com/openminted_eu

facebook.com/openminted

bit.do/openmintedlinkedin

vimeo.com/openminted

bit.do/openmintedplus

Contact us

www.openminted.eu