nycfacets
by pediacities

Metadata, Extrametadata & Crowdknowing

Fostering 'Big Open Data' in government through Open Collaboration

Ontolog - "Big Open Data" session 2
May 17, 2012

Joel Natividad, co-founder
@jqnatividad

CROWDKNOWING

Human-powered,
Machine-accelerated,
Collective Knowledge Systems

0. Huge Open Data

1. Extract Metadata

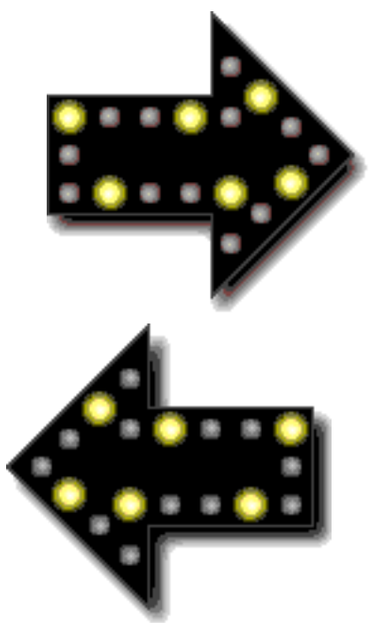
2. Derive ExtraMetadata
   (Semantics + Statistics + Algorithm + Crowd)

3. Do Federated Queries on both the Metadata AND the Data

Crowdknowing

Crowdknowing

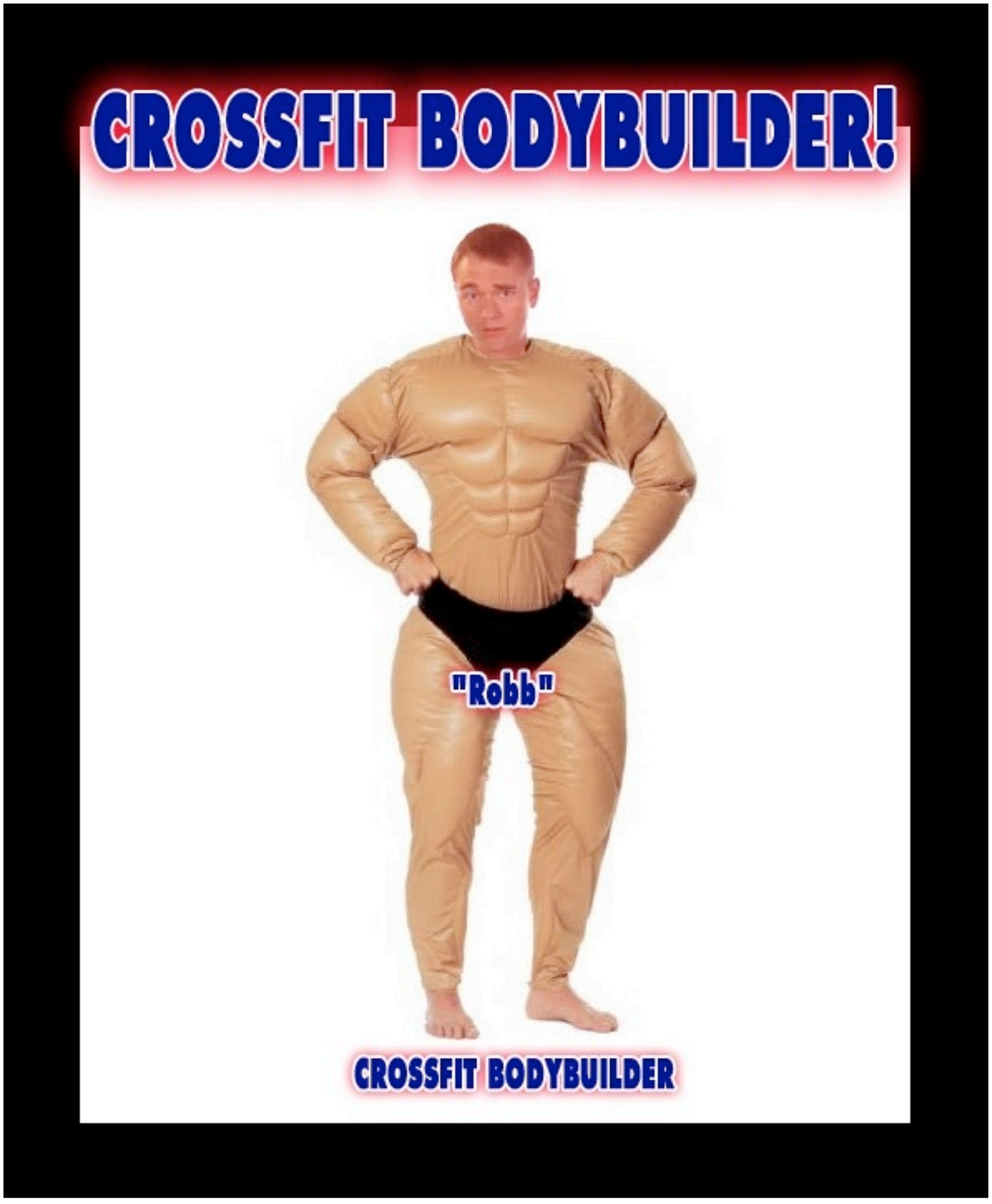
Human-powered, Machine-accelerated, Collective Knowledge Systems



Curation, Comments,
Feedback, Bug Reports,
Likes, Shares, Profile, Votes,
Subscribes, Tagging,
etc. etc. etc.

Ontology, Inferencing, Semantic Mapping, Query Federation, Statistics, Pattern Recognition, Multivariate Analysis & Forecasting, Automated linking, Feeds, Notifications
etc. etc. etc.

a Semantic Data Dictionary



Semantic Steroids

- Searchable
  - Faceted Search
  - Drilldown
- Interlinked
- Semantic Browsing
- Queryable
- Query Results Formats

~3.5M facts

~950 datasets/views

NYCFacets Spider v0.5

- Crawls NYC Open Data Catalog every weekend
- RESTFul API
- Extracts metadata & derive extrametadata
- Pumps the data into NYCFacets

Metadata

Top Level Metadata

- Name/ID
- Category
- Dataset Type
- Attribution
- Owner ID, etc.

Detail Metadata

- Column Names
- Datatype
- Width, etc.

Wifi Hotspot Locations (ehc4-fktp)

Location of wifi hotspots in the city with basic descriptive information.

Details | Columns | Download

Name	Wifi Hotspot Locations
ID	ehc4-fktp
Category	Mass Media
Dataset Type	Map
Publication Group	
Attribution	Department of Information Technology and Telecommunications (DoITT)
OID	341764
Rows Updated By	
Table ID	243853
View Type	tabular
Owner ID	5fuc-pqz2
Owner Name	NYC OpenData
Owner Role	administrator

Stats	
View Count	11079
Number of Downloads	1483
Number of Rows	1126
Number of Columns	10
Number of Comments	5
Average Rating	60

Dates	
Date Created	Fri Oct 7 14:59:44 2011
Date Published	Fri Oct 7 15:00:01 2011
Rows Updated At	Wed Dec 31 19:00:00 1969
View Last Modified	Tue Mar 13 14:18:16 2012

★ PEDIACITIES RANK

RELATED DATASETS

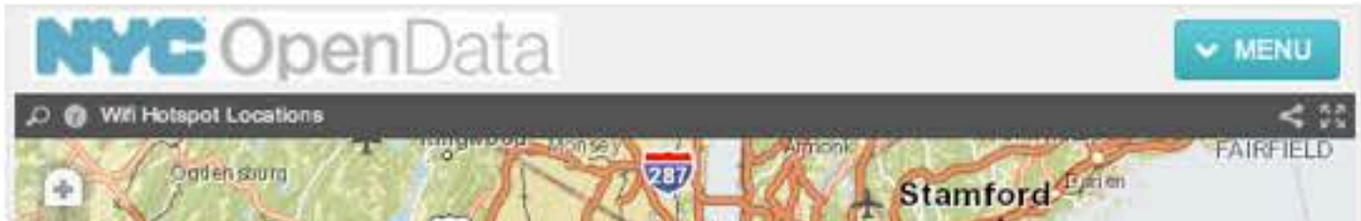
by Category

by Attribution

by Owner

TAGS

geographic , location , map , cartography , services , wifi , wireless , hot spot , access , telecommunication



ExtraMetadata?

- Derived using
  "Semantics, Statistics, Algorithm & the Crowd"
- "Supercharacterize" each dataset
  not just the schema, but by sampling the underlying data as well
- Score each dataset - Pediacities Rank
- Virtuous Feedback Loop
  micro-conversations/contributions around the Data

ExtraMetadata

Top Level ExtraMetadata

- Number of Rows
- Pediacities Rank
  - Freshness Score
  - Sparseness Score
  - Social Score
  - Views Score
  - Download Score
  - Rating Score

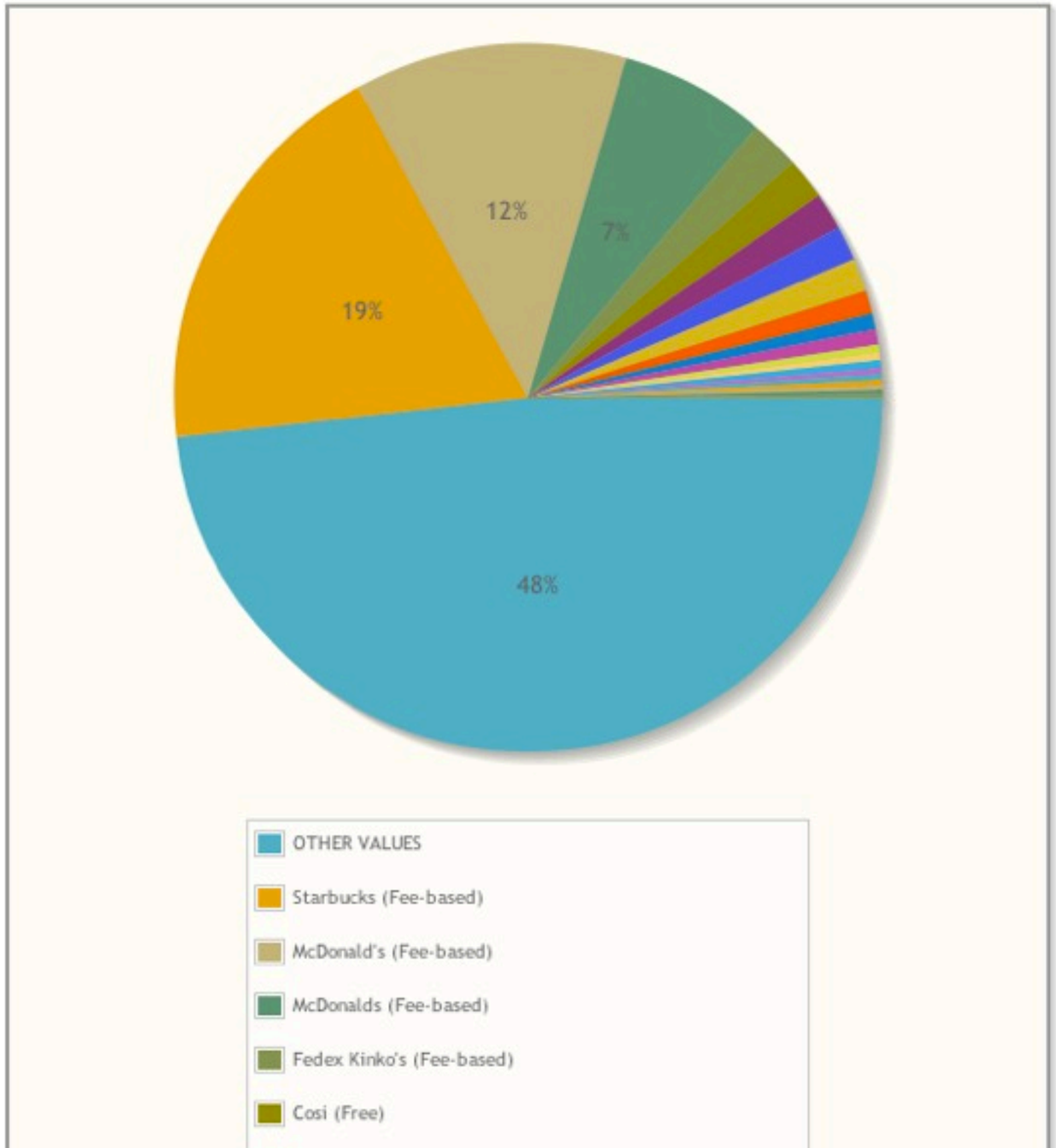
Detail ExtraMetadata

- Top Values
- Descriptive statistics
  - Nulls/Non-nulls
  - Smallest Value
  - Largest Value
  - “Uniqueness”
- Simple Visualization

Name (ehc4-fktp,2979143)

Top Values | Top Values (table)

Please hit **Refresh** if the page is empty.



COLUMN METADATA

Wifi Hotspot Locations

Name:	NAME
Parent ID:	ehc4-fktp
ID:	2979143
Data Type:	text
Field Name:	name
Position:	3
Render Name:	text
Table Column ID:	1560341
Column Width:	100

COLUMN STATS

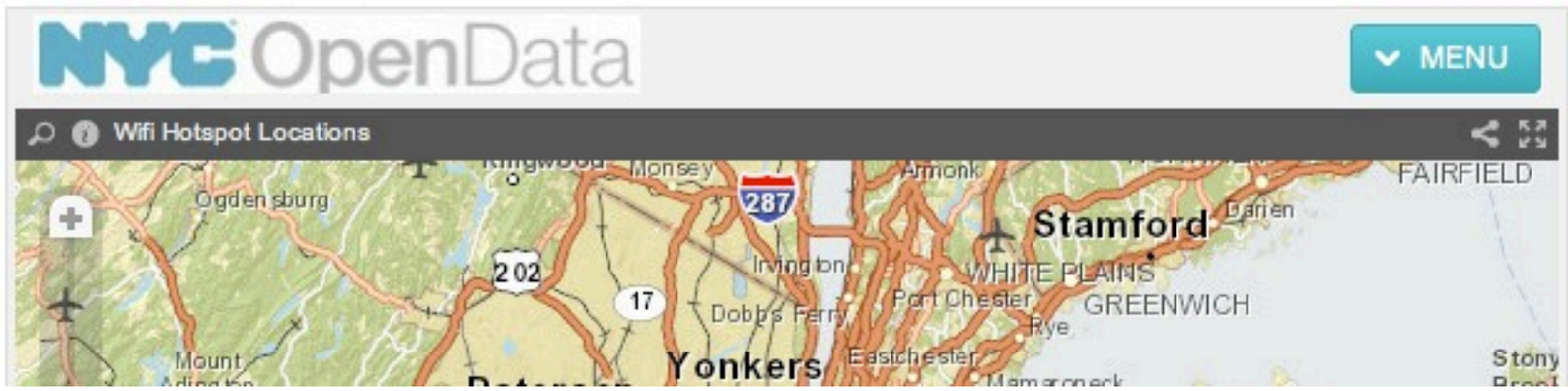
Data Type:	text
Non Null:	1126
Smallest:	'sNice (Free)
Largest:	Zocalo at Grand Central Terminal (Fee-based)
Null:	0

Wifi Hotspot Locations (ehc4-fktp)

Location of wifi hotspots in the city with basic descriptive information.

Details | Columns | Download

Position	Name	DataType	FieldName	Width	TopVals	Uniqueness
1	OBJECTID	number	objectid	100	20	1
2	Shape	location	shape	100	20	1
3	NAME	text	name	100	20	0.64
4	ID	number	id	100	20	0.75
5	ADDRESS	text	address	100	20	0.72
6	CITY	text	city	100	20	0.51
7	ZIP	text	zip	100	20	0.66
8	PHONE	text	phone		0	0
9	TYPE	text	type		0	0
10	URL	text	url		0	0



★ PEDIACITIES RANK

RELATED DATASETS

by Category

by Attribution

by Owner

TAGS

geographic , location , map , cartography , services , wifi , wireless , hot spot , access , telecommunication

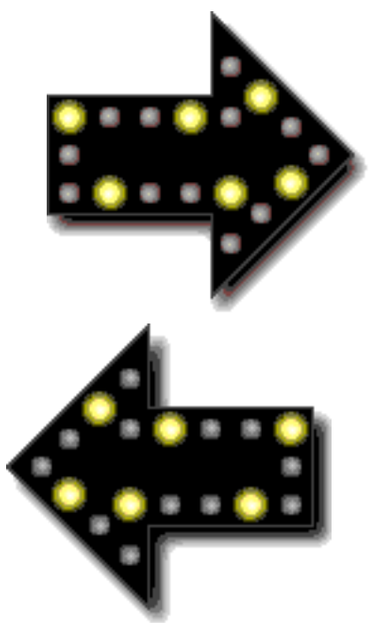
“Crowd”

Microconversations/contributions

- Overall Rating
- Comments (comment rating)
- Bug Reports (data quality)
- Likes/Shares
- Downloads

Crowdknowing

Human-powered, Machine-accelerated, Collective Knowledge Systems



Curation, Comments,
Feedback, Bug Reports,
Likes, Shares, Profile, Votes,
Subscribes, Tagging,
etc. etc. etc.

Ontology, Inferencing, Semantic Mapping, Query Federation, Statistics, Pattern Recognition, Multivariate Analysis & Forecasting, Automated linking, Feeds, Notifications
etc. etc. etc.

nycfacets
by pediacities

- More Datasources!
- Not just Metadata!
- Federated Queries!
- SPARQL endpoint
- Bugzilla Integration
- Collaborative Ontology Modeling
- Feeds
- Microcontributions
- Gamification
- In time for NYCBigApps 4.0

We need your help & feedback

A Smart Data Exchange for All Data NYC

Find out more at
http://nyc.pediacities.com/facets

@jqnatividad @samimirzabaig @pediacities @ontodia

CREDITS

- Flickr User Weston Price, Paleo-Caveman-Omnivore-LowCarb-Meat-Diet-Info (http://www.flickr.com/photos/paleo-atkins-meat-diet-info/with/6718805047/)
- Flickr User Gao Yi (http://www.flickr.com/photos/gaoyi/178514677/)