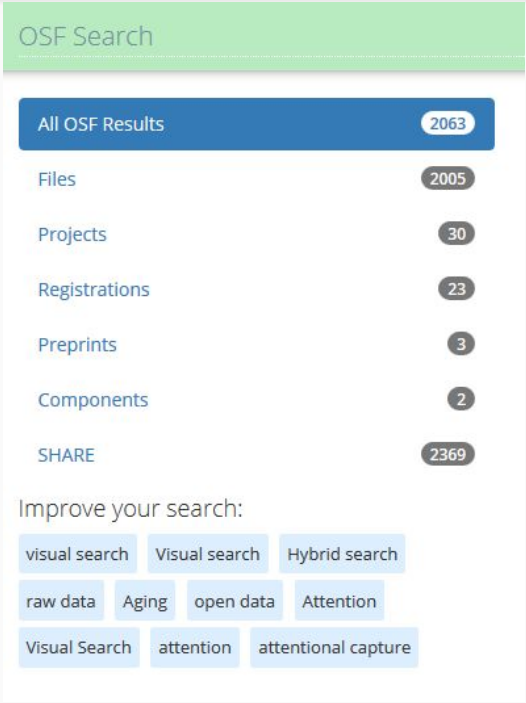
Improving Data Discoverability with Folksonomy to Taxonomy Integration

Alena Begler
PhD student at Graduate School of Management SPbU
Lab Manager at cognitive psychology scientific group
alena.begler@gmail.com

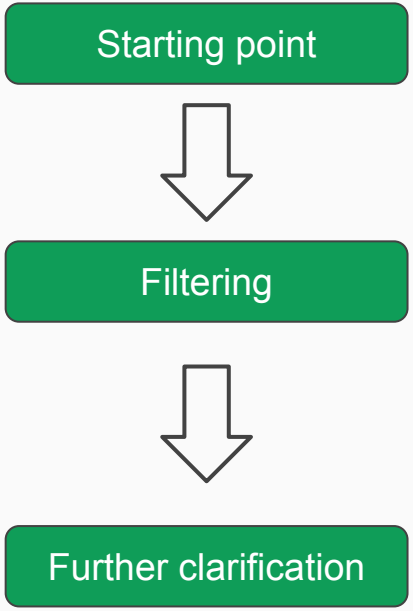
What we have already



1. Search field

2. Categories (sometimes)

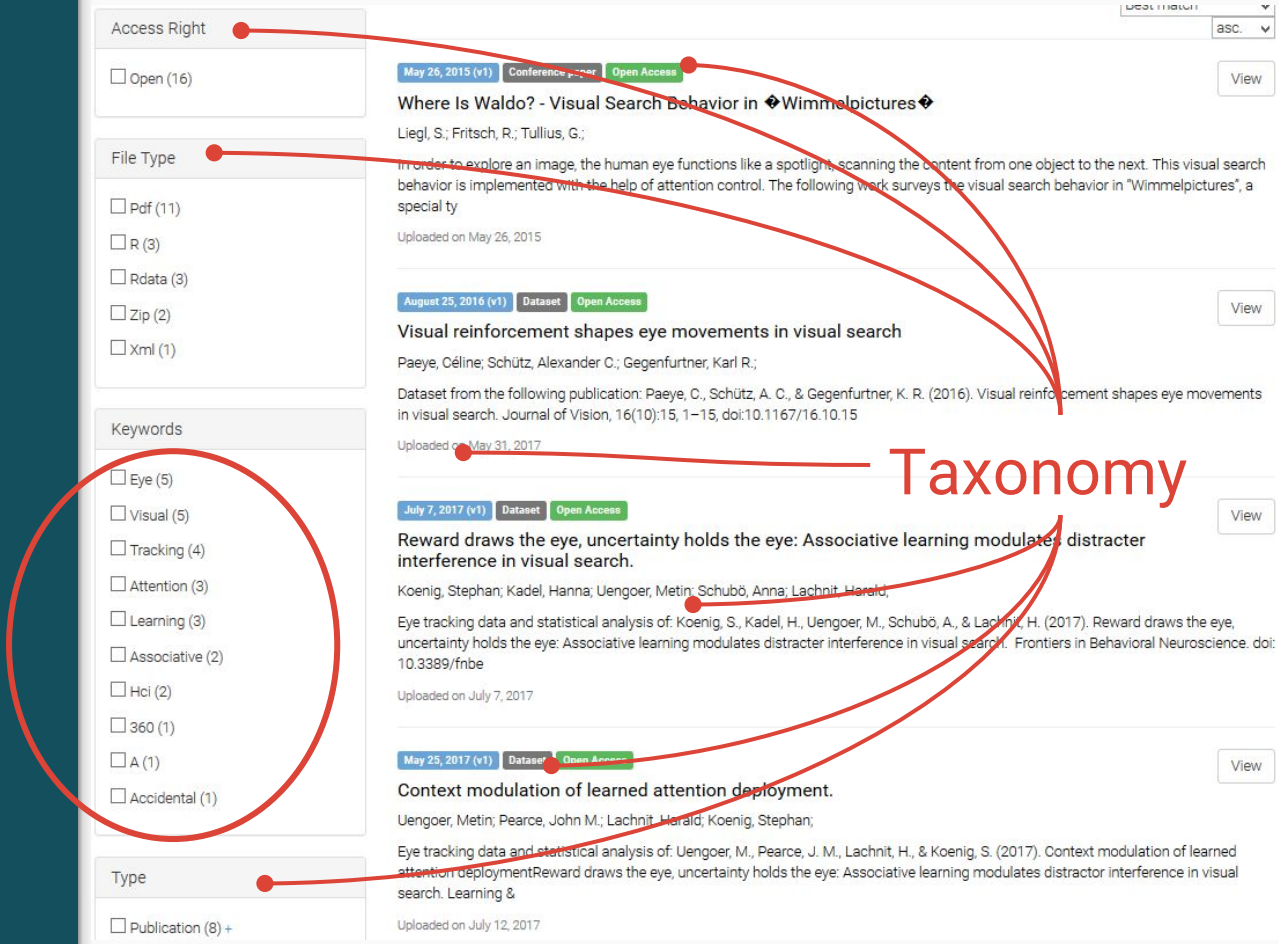
3. User tags



How it works

Search results at Zenodo

Folksonomy



The problem

Search results for “visual search”:

Repository	Items founded	“Useful” items	Tags provided
OSF	20 projects, 1999 files	9	3
figshare	0 projects, 123 datasets	10	10*
Dryad	5 projects, 0 files	3	3
Zenodo	4	4	3

*figshare has mandatory tagging for public projects

Tags problems

- Majority of tags are unique
  (116 tags overall, 86 of them are unique, only 4 more than 2 times)
- No differentiation in the field of science
  (psychology, biology, machine learning)
- Tags not always useful (i.e. “task” or “result”)

Users need guidance to provide good tags

Taxonomy-directed folksonomies

- myEdna (Hayman, S., & Lothian, N. (2007). Taxonomy directed folksonomies)
- TaxoFolk (Kui, C., & Tsui, E. (2011). TaxoFolk: A hybrid taxonomy-folksonomy structure for knowledge classification and navigation)

Metadata taxonomies

For general metadata aspects

e.g. Dublin Core properties:
Content: Title, Subject, Description, Type, Source, Relation, Coverage
Intellectual Property: Creator, Publisher, Contributor, Rights
Instantiation: Date, Format, Identifier, Language

For different types of content

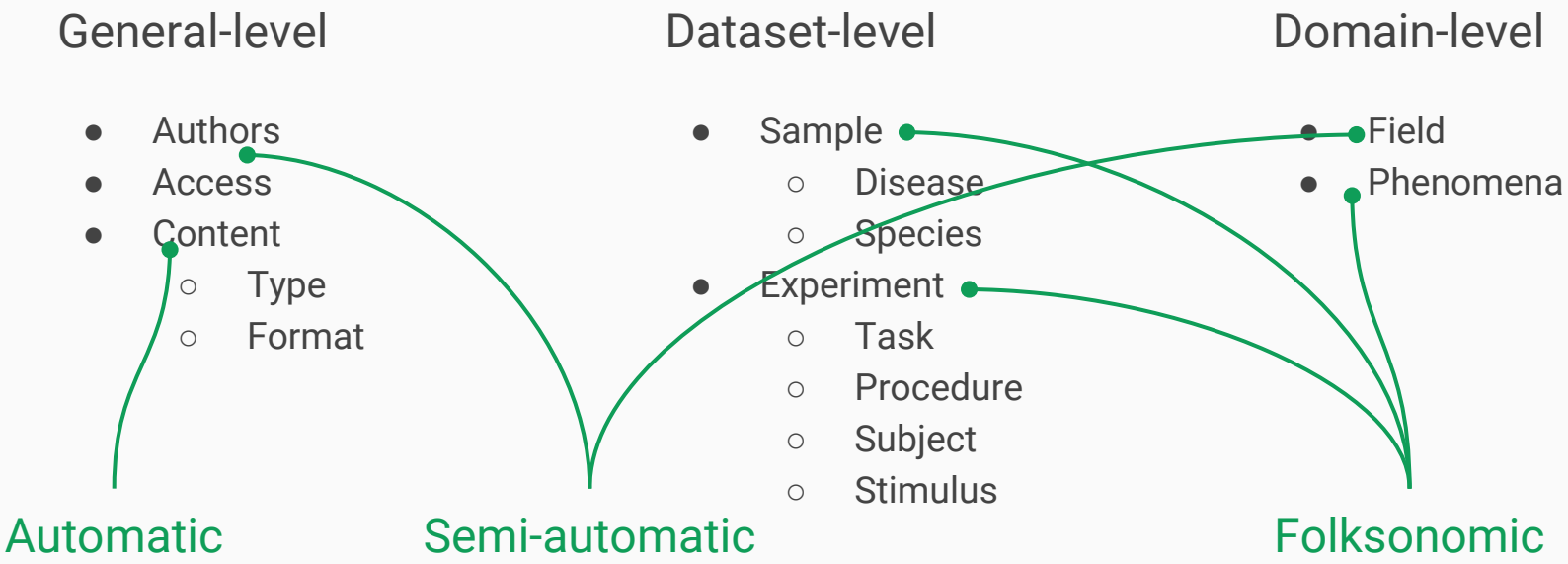
e.g. BrainMap classes:
Paper-level: Citation, Submitter, Abstract, Prose Description, Subject Groups, Conditions, Sessions, Imaging, Results Synopsis, Feedback
Experiment-level: Context, Contrast, Paradigm Class, Behavioral Domain, Locations

For domain-specific information

1. Research areas classifications (WoS, OECD)

2. Inside a research area (MIBBI standards for biological studies)

Facets example



Tags integration

General Dataset/Sample Dataset/Experiment Domain-specific

Access	Sample	Task	Procedure	Subject	Stimulus	Phenomena	Uncategorized
dataset	adhd	visual search	associative learning	humans	camouflage	attention	concept
open data	awd	discrimination learning	cognitive control	ideal observer	Animal coloration	attentional	Conceptual Proces
raw data	deuteranomaly	distractor exclusion	decision experiment		Anti-predator colorat	attentional temp	context
		distractors distribution	fear conditioning		background comple:	color vision	coordination
		ensemble representation	search display		Defensive coloration	color vision defic	Crypsis
		learning of probability	real world behavior		line orientation	decision-making	difficulty
		head orientation	Visual Attention Modulation		Multiple Object Trac	search strategie	duration
		negative cues	Bayesian inference		Multiple Target Sear	summary	e.g
		oculomotor capture	eye movements		orientation discrimin	visual attention	finding
		pereptual load	eye tracker		stimulus correlation:	visual perceptior	Head Direction Pe
		priming of pop-out	head movements			visual statistics	literacy

Computational experiment?

What do you think?