

No Bull, No Spin:  
comparing tags  
with other types of photo metadata

Cathy Marshall

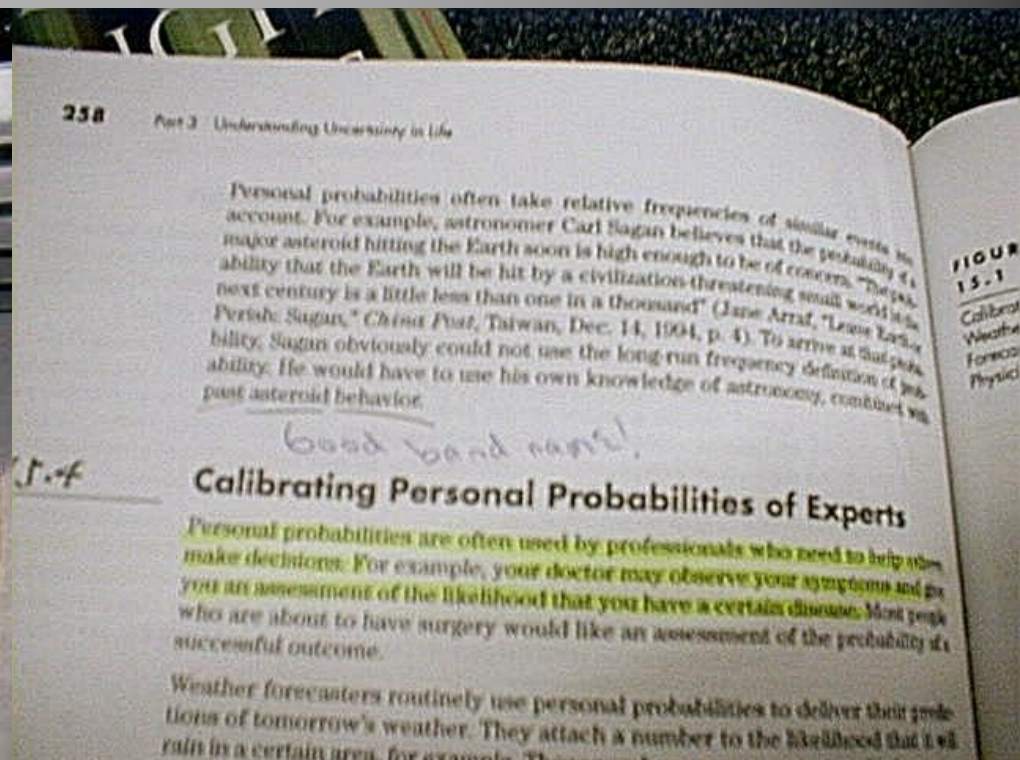
MSR-SVC Wednesday lab meeting  
29 April 2009

# What do you do over the holidays?



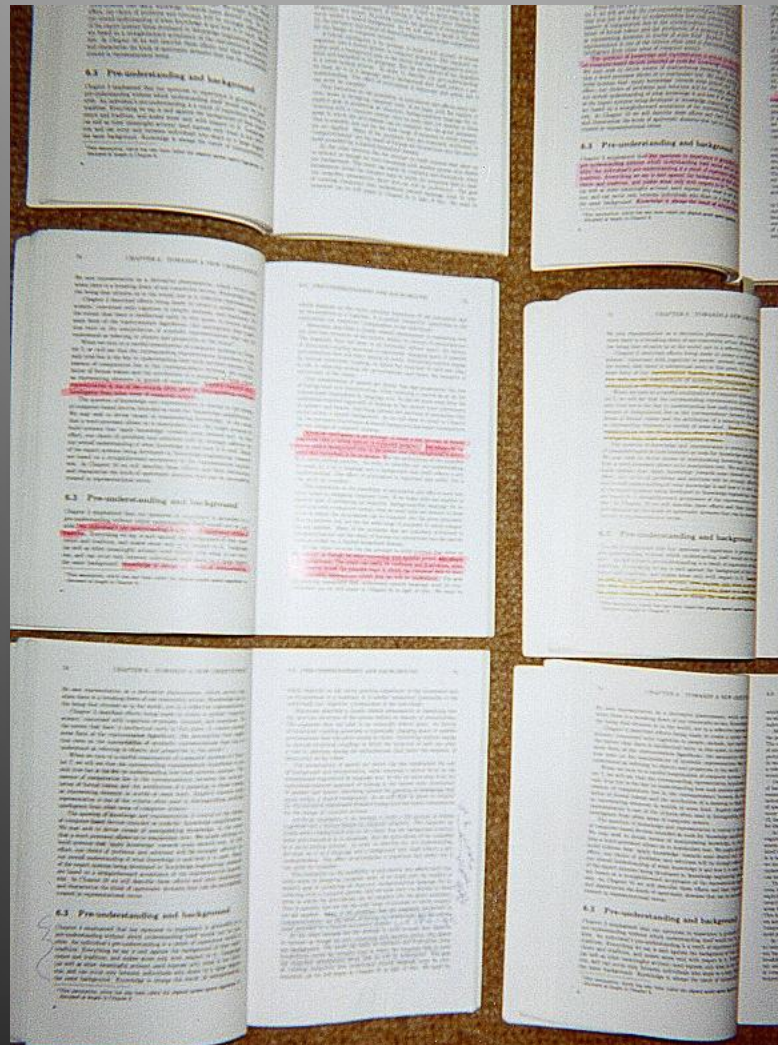


# Over the holidays, I do feral research



In 1996, I spent my holiday in a university book store, looking at annotations in 1000s of used textbooks...

in 1997, I looked at annotations' social effects by comparing different copies of the same used textbook



so usually  
these projects  
are time-  
consuming,  
compulsive,  
off-topic...

Three years ago, I had a yen.

I wanted to put the folksonomy question to rest.  
That is, I wanted to answer the loaded question

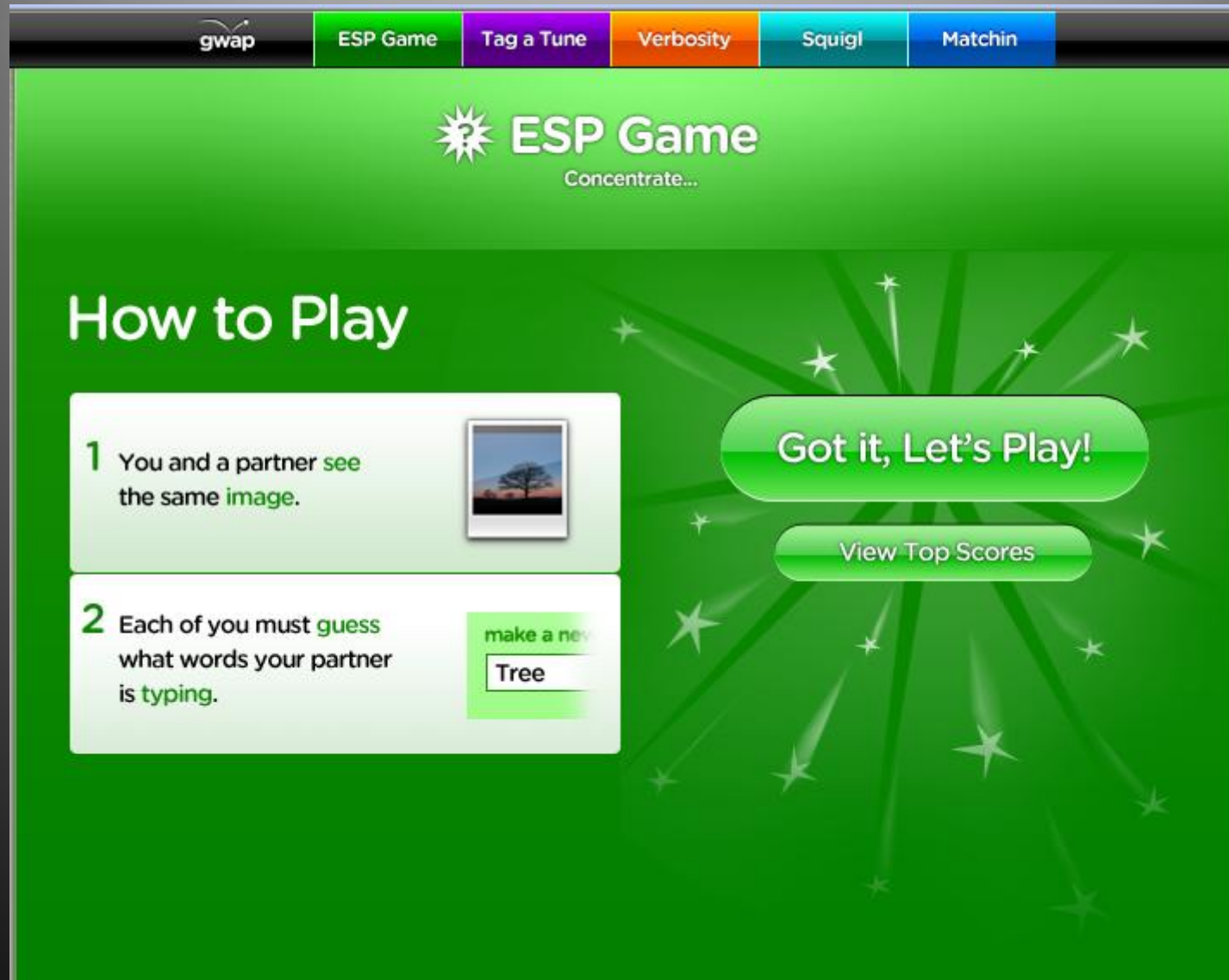
Do folksonomies work?

*I was skeptical about the value of public tags  
&*

*I wanted to convince others to be skeptical too.*



about then Luis van Ehn introduced the ESP Game to show how crowdsourcing + tags can be rolled up into a game to solve the image retrieval problem. It was all over the news... better than sliced bread...



# from my blog, 4 May 2006

“

Here's one that was already tagged *girl* when we got there. We added *pink*. When in doubt, add a color. ... Don't try to be too specific about the color though: you'll never match on *aqua*, only *blue* and *green*. And no matter how little *red* there is, go for it: it's a very popular tag.



Ah, here's one from the news. Someone's already given this photo the helpful descriptor *man*. He's someone in the Spanish-speaking world, perhaps a soccer star. ... But to us, he's just a man with *hair*. Yep. Hair was the salient feature we picked out for this guy. That'll help someone who's looking for men who aren't bald.

We don't recognize this guy either. He's a TV newsman on MSNBC. He's already tagged with *man old gray*. He doesn't look that old to me, but then again, we know this kind of thing is relative. He obviously looks old to some of my [ESP Game] compatriots ... My partner and I add *smile* to the list.



”

but just blogging about the ESP game  
didn't seem like it would cut it.

Sure, it was snarky,  
but it wasn't convincing

I was missing a good lens through  
which to look at tags

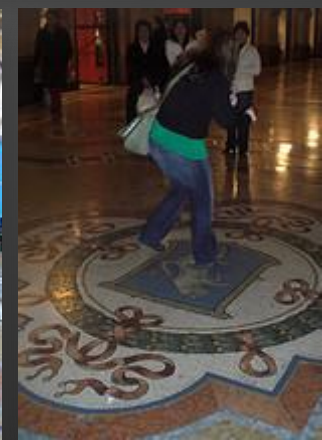


around that time, the *Daily Show* alerted me to a weird and tasteless pop culture fad...



A bit of research on the Internet told me they were called "Bumper Nuts"

# My search on 'bull balls' turned up something else...







in an aha! moment,

I realized Flickr gave me an opportunity to  
look at lots of different peoples' tags of

*the same image*



so in 2007, over the xmas holidays,  
I did an informal study of how people  
had tagged their photos of the bull  
mosaic...





# TEKKA

ENJOYING NEW MEDIA • SOFTWARE AESTHETICS

[home](#) | [writing](#) | [about Tekka](#)

## DO TAGS WORK?

Cathy Marshall

Tag! You're it!

It seems that everywhere I go on the Web these days is tagged.

When I login to Flickr, the first photo I see— an artistic effort that looks to be a study of a sturdy dandelion growing from a crack in the curb—is tagged *flowers*, *spring*, *scenery*, *lapsana*, *great*, *fotolog*, *ThinkFlickrThink*, *SuperShot*, and *1mill*.



photo: Junichiro Aoyama

I like the photo, although it calls to mind dozens of inspirational posters, the province of the over-earnest, the literal, and the hapless striver. Flowers? I can see only one. A lapsana. Who knew? Is it scenery? Well, it's not a dog, although I think of scenery as grander in scale and as less ambiguous in its beauty than a lone dandelion. Is it great? According to his comment, Blue Cockatoo thinks so. How thick is the stem? Oh, I'd say it's about 1mill.

Heckuva job, tags!

that was pretty satisfying, but really I  
wanted to publish a more rigorous  
argument based on a case study of the  
bull mosaic...

My holiday project for 2008!

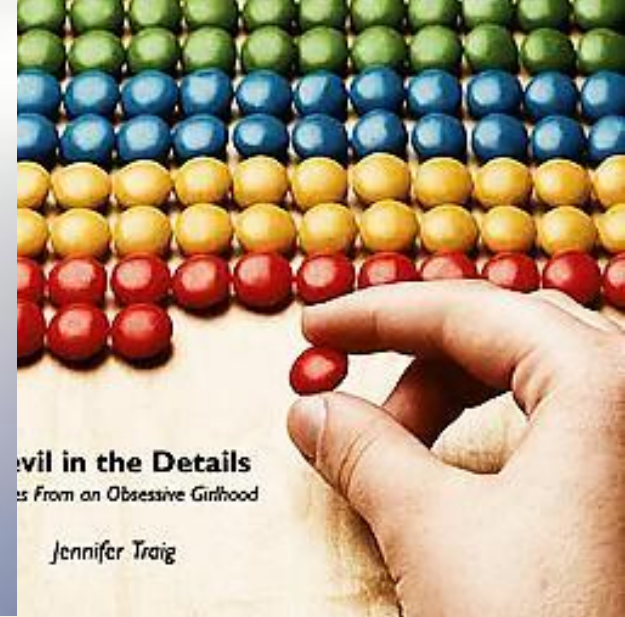
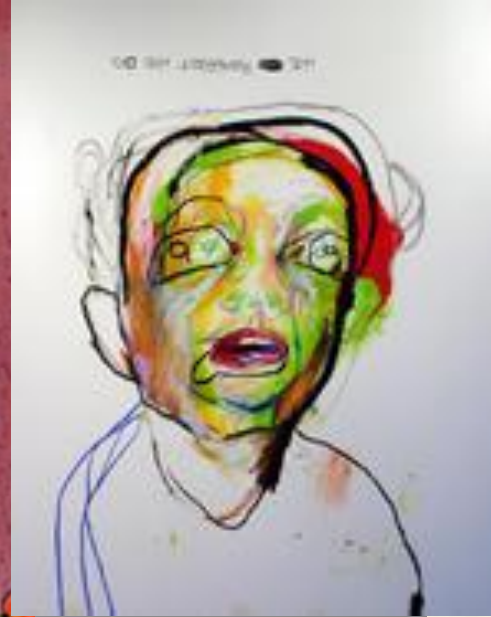
# assemble a dataset and answer 3 simple questions about tags

- Are tags effective\*  
especially as compared with  
other kinds of metadata?
- How are tags different from  
other kinds of metadata?
- Any chance of improving  
tags by using the results of  
this study?



First, assemble the dataset  
(this time recording the process and  
rationale)

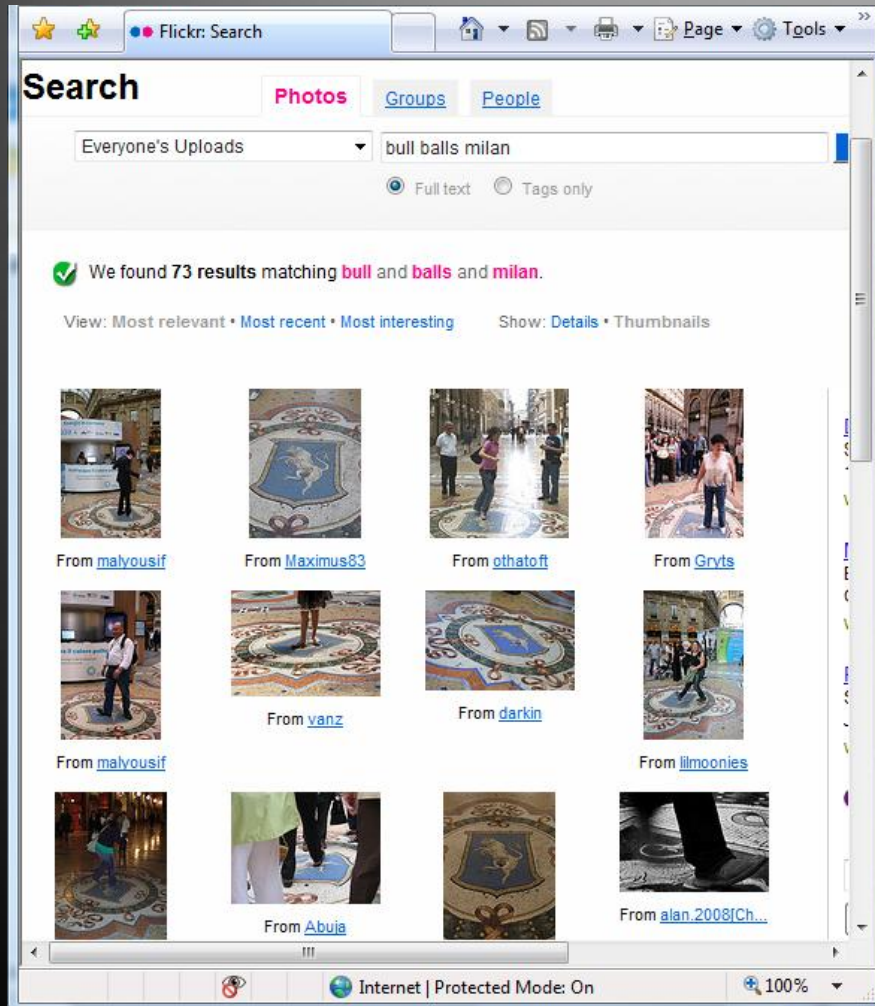




Good thing I'm (a little) OCD...

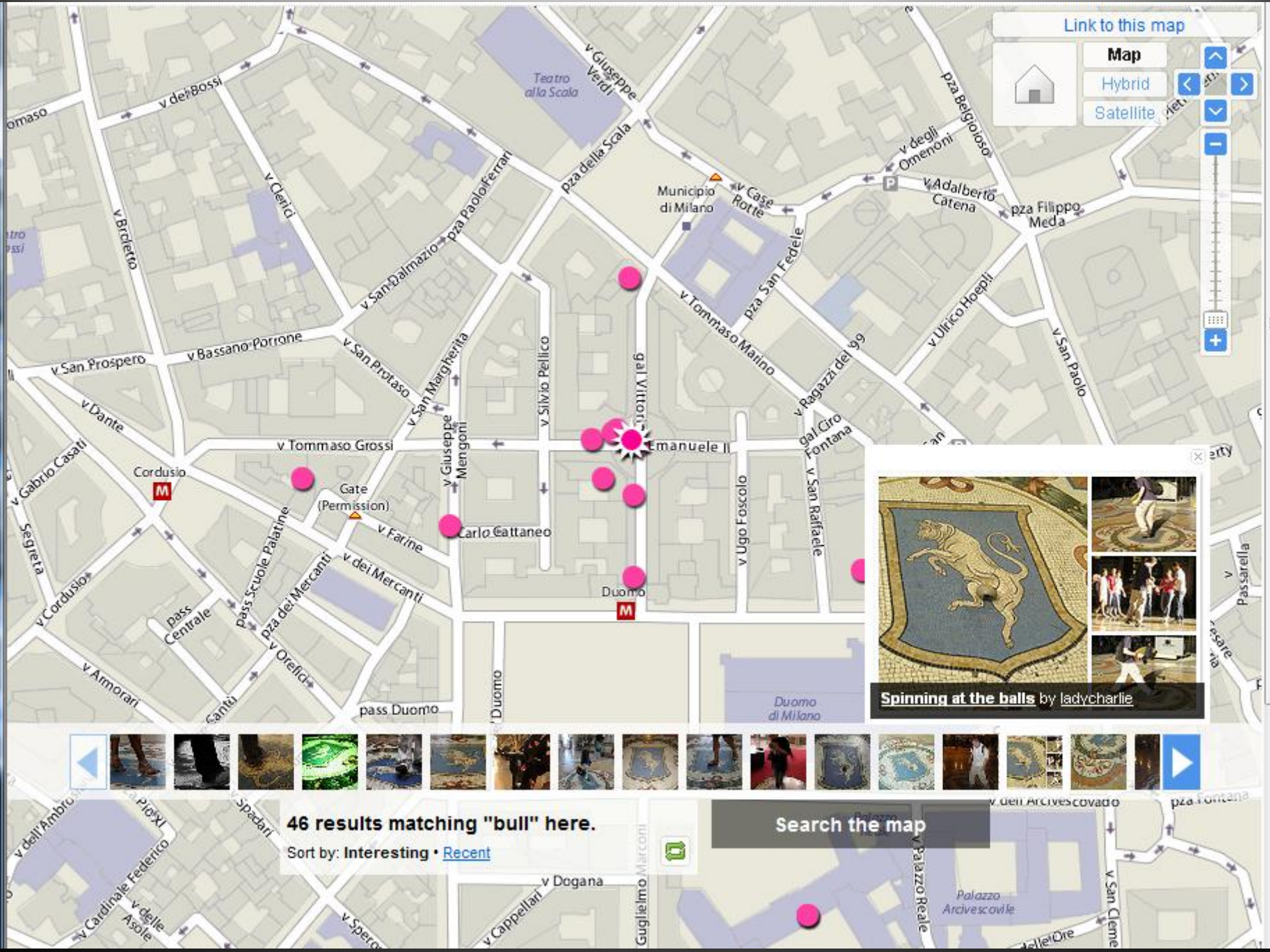


# Find as many qualified photos as possible using different strategies...



- Using queries
  - 34119 candidates
  - 540 new qualified
- Browsing photostreams
  - 4250 candidates
  - 62 new qualified
- Using geotags
  - 2555 candidates
  - 1 new qualified





Link to this map

Map

Hybrid

Satellite

**Spinning at the balls** by ladycharlie

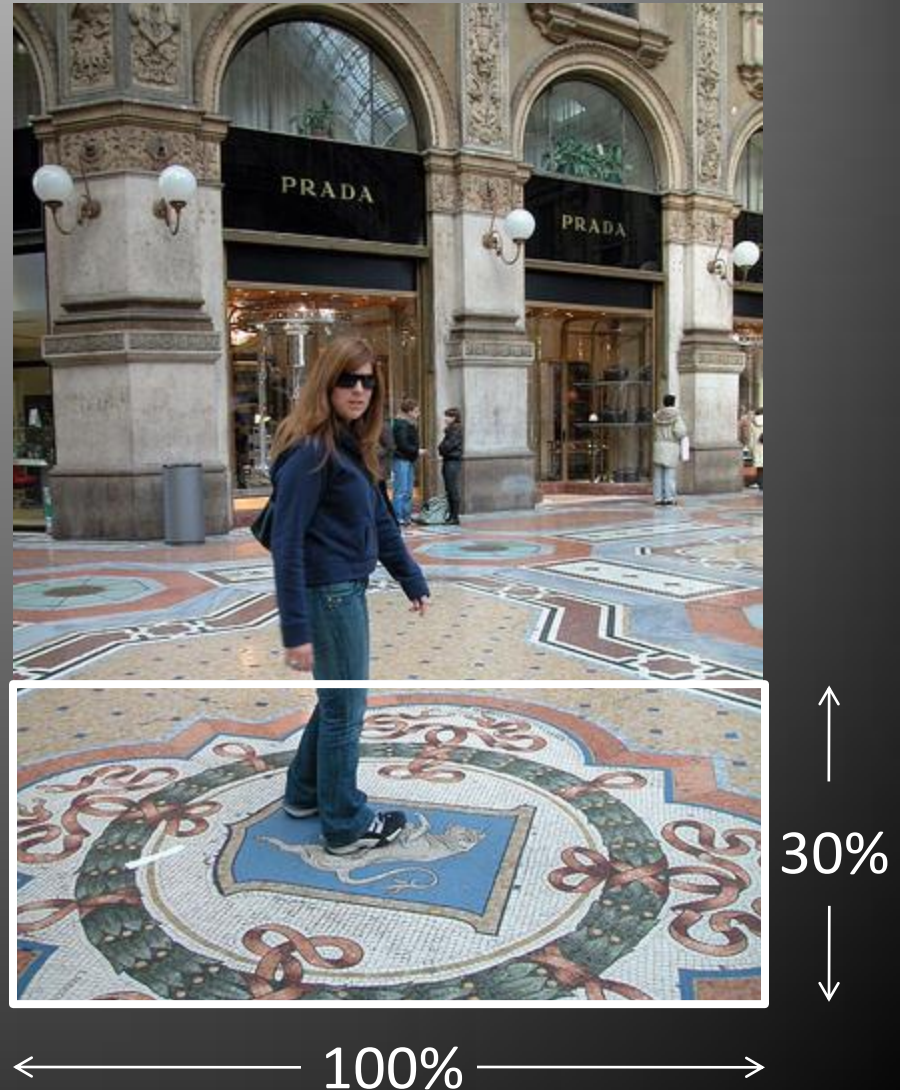
46 results matching "bull" here.

Sort by: Interesting • Recent

Search the map

# relevance criteria

- Unambiguously correct subject (the bull mosaic)
- Mosaic covers 20% of height and 75% of width
- Mosaic is centered horizontally
- Metadata alone is insufficient to establish relevance—identification must be visual





Using those criteria  
and brute force gathering methods

I found **603** photos of the Milanese  
bull mosaic

*a sufficiently large a number for a  
case study*

- person (306 photos)



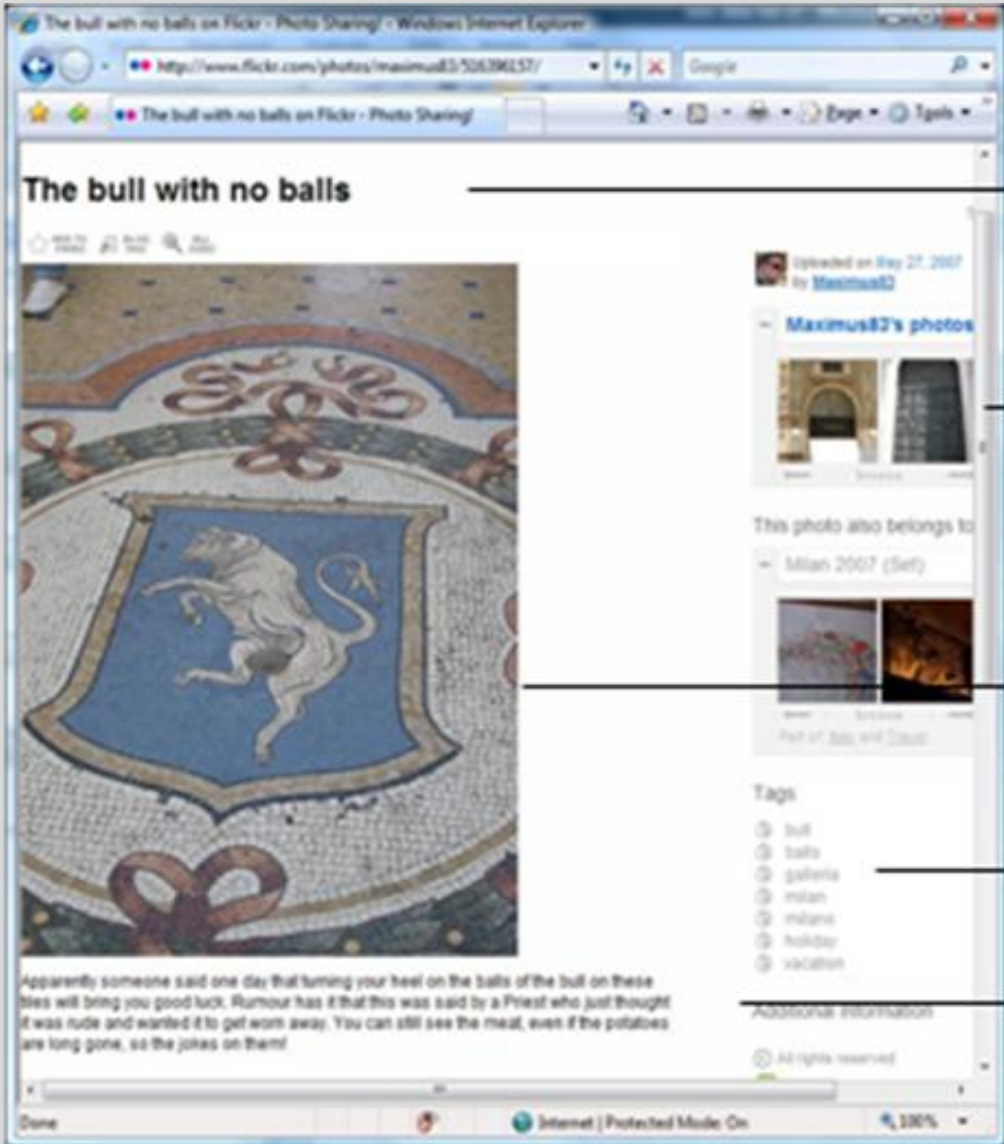
- mosaic (186 photos)



- action (111 photos)



# I recorded user metadata for each



The screenshot shows a Flickr photo page for a bull's head mosaic. Annotations with arrows point to various elements:

- title**: Points to the title "The bull with no balls" at the top left of the photo area.
- photostream**: Points to the "Maximus82's photos" section on the right side of the page.
- photo**: Points to the main image of the bull's head mosaic.
- tag set**: Points to the "Tags" section on the right, which lists: bull, balls, galeria, milan, milano, holiday, and vacation.
- narrative caption**: Points to the text at the bottom of the page: "Apparently someone said one day that turning your heel on the balls of the bull on these tiles will bring you good luck. Rumour has it that this was said by a Priest who just thought it was rude and wanted it to get worn away. You can still see the meat, even if the potatoes are long gone, so the jokes on them!"

next, build a profile of this dataset:

how does it compare with its larger kin?

what kinds of conclusions can we draw  
about the different types of metadata?



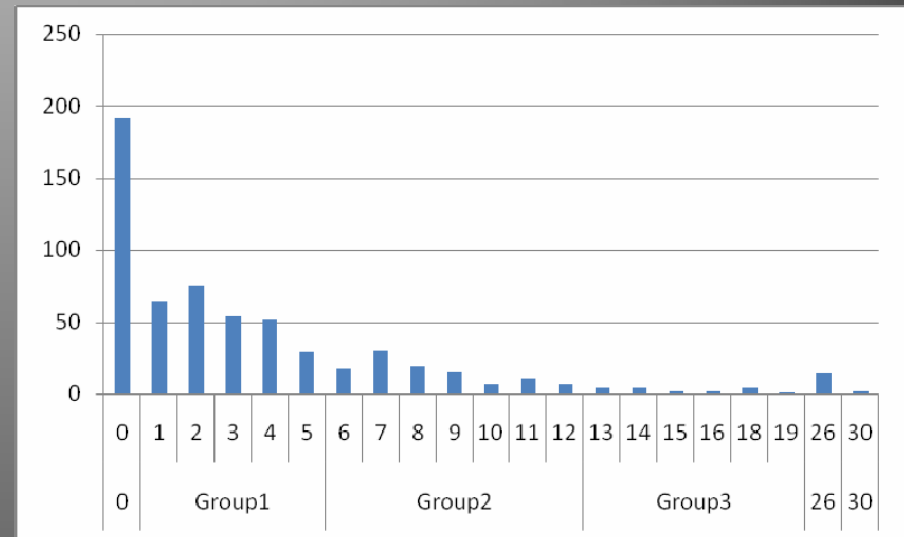
# Photos with each type of metadata

<b>type</b>	<b>#</b>	<b>with tags</b>	<b>with a title</b>	<b>with a caption</b>
<i>action</i>	111	80 (72%)	109 (98%)	58 (52%)
<i>mosaic</i>	186	139 (75%)	182 (98%)	114 (61%)
<i>person</i>	306	193 (63%)	304 (99%)	189 (62%)
<i>total</i>	603	412 (68%)	593 (98%) 501 (83%)	361 (60%)

# What can we learn about tags?

(how does this dataset compare to the literature?)

- Mean number of tags per photo is 3.8 (slightly higher than the number reported by Ames and Naaman, CHI 2007—2.2)
- Short tag sets are much more common than long ones. (similar results in Sen et al., CSCW 2006)
- Frequency distribution looks like Sigurbjörnsson and van Zwol, WWW 2008.
- *Basically this small dataset is comparable to the tag profiles in the literature.*
- But... if we look at them closely, it's clear that different tagging strategies are at work.
  - some focus on what's in a single photo
  - some span multiple photos (e.g. itinerary)
  - some distinguish local context



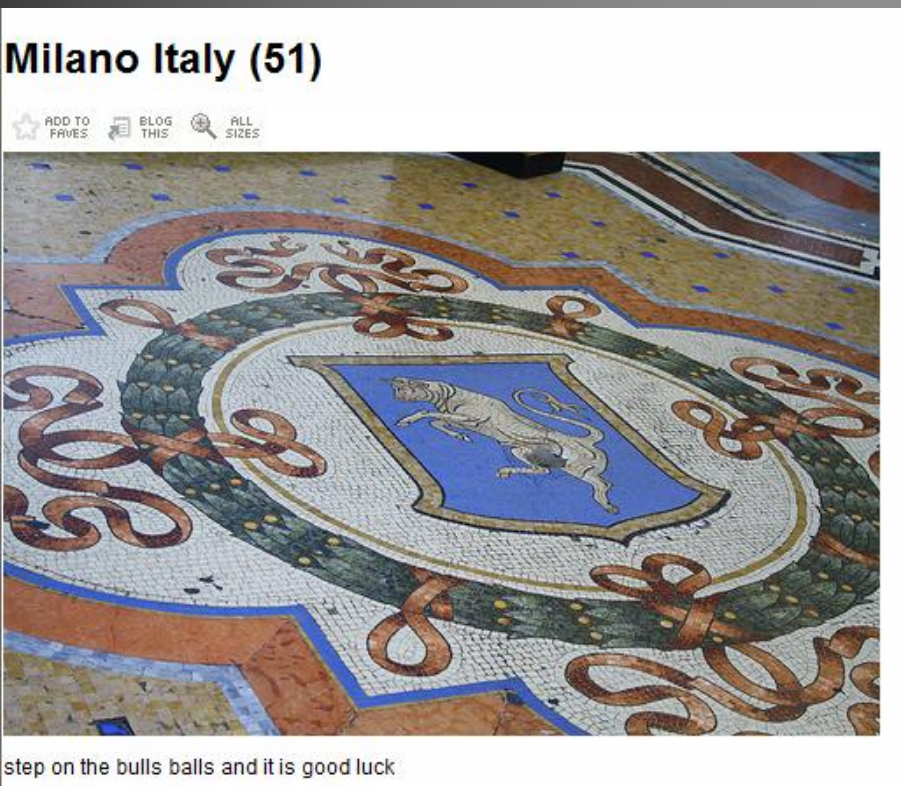
*{Arriving Home, Dinner With Lia, Milan or Bust} (PID 120)*

—or—

*{Switzerland, Italy} (PID 60)*

# How about the captions?

- Mean length is around 21 words
- Generally the same story is told with varying amounts of detail.



The Galleria Vittorio Emanuele fulfills numerous social functions...The floor mosaics are a vastly underrated source of pleasure, even if they are not to be taken too seriously... Be sure to follow tradition and spin your heels once or twice on the more-"delicate" parts of the bull beneath your feet in the northern apse; the Milanese believe it brings good luck.  
(normative 60 words, from Fodor's)

The floor mosaic represents arms of four Italian cities, Milan - red cross on white ground; Rome - the wolf, Florence - the lily; and Turin. - the bull. Follow tradition and spin your heels three times on the most "delicate" parts of the bull. Legend says that revolving three times on one leg placed exactly in that position brings good luck and fortune.  
(64 words, from the user)

# And what about titles?

- Mean length: 4 words
- Median length: 3 words
- Max length: 20 words
- An example 4 word title:  
*Crushing the bulls balls*
- An example 7 word title:  
*Spinning on the bull's balls in Milan*
- An example 20 word title:  
*if you put your right legs in the hole and turn right for 3 times, it'll bring you good luck*

*Spinning*

*Spinning*

*Spinning 3 times on the Bulls Balls.*

*Spinning after a wish*

*Spinning at the balls*

*Spinning for luck*

*Spinning on bull testicles in Milan*

*Spinning on bulls balls, Milan*

*Spinning on Taurus the Bull*

*spinning on the bulls balls*

*Spinning on the Bull for Good Luck in Milan*

*Spinning on the bull for luck*

*Spinning on the bull in Milan...*



It was the titles that first tipped me  
off...

it's not just tags that are similar.  
I should be looking at relative word  
frequencies

# Different kinds of words/strategies

- *Place*: refers to where photo was taken (e.g. milan, galleria)
- *Artifact*: refers to subject of the photo (e.g. bull, balls, mosaic)
- *Context*: refers to photographer's personal situation (e.g. switzerland, Lia, mother)
- *Story*: refers specifically to the legend (spin, good, luck, three)



# Top 10 words used in tags & their frequencies (English only)

Tag word	Frequency (items w/tags)	Word category
Milan	85%	place
Italy	66%	place
Galleria (&variants)	26%	place
bull	25%	artifact
Emanuele	14%	place
Vittorio	13%	place
Europe	12%	place
200x	11%	context
travel	9%	context
luck	9%	story



as you might guess from the frequencies, the most common tag set for these photos is:

{Milan, Italy}

26/66 photos with two tags had that tag set!

Well-aligned with Sigurbjörnsson and van Zwol,  
WWW2008

Only a quarter of the tag sets (17% for all items)  
reflect the artifact and 9% (6% for all items)  
refer to the story.

# Top 10 words used in titles & their frequencies (English only)

title word	Frequency* (items w/titles)	Word category
bull	42%	artifact
Milan	27%	place
luck	26%	story
Galleria (&variants)	22%	place
balls	17%	artifact
Vittorio	15%	place
spin/spun	15%	story
Emanuele	14%	place
good	13%	story
Il	9%	place

# Top 10 words used in captions & their frequencies (English only)

Caption word	Frequency (items w/captions)	Word category
bull	65%	artifact
luck	51%	story
good	45%	story
spin	45%	story
Milan	31%	place
heel/heal	30%	story
balls	25%	artifact
Galleria (&variants)	24%	place
around/round	20%	story
testicle	18%	artifact

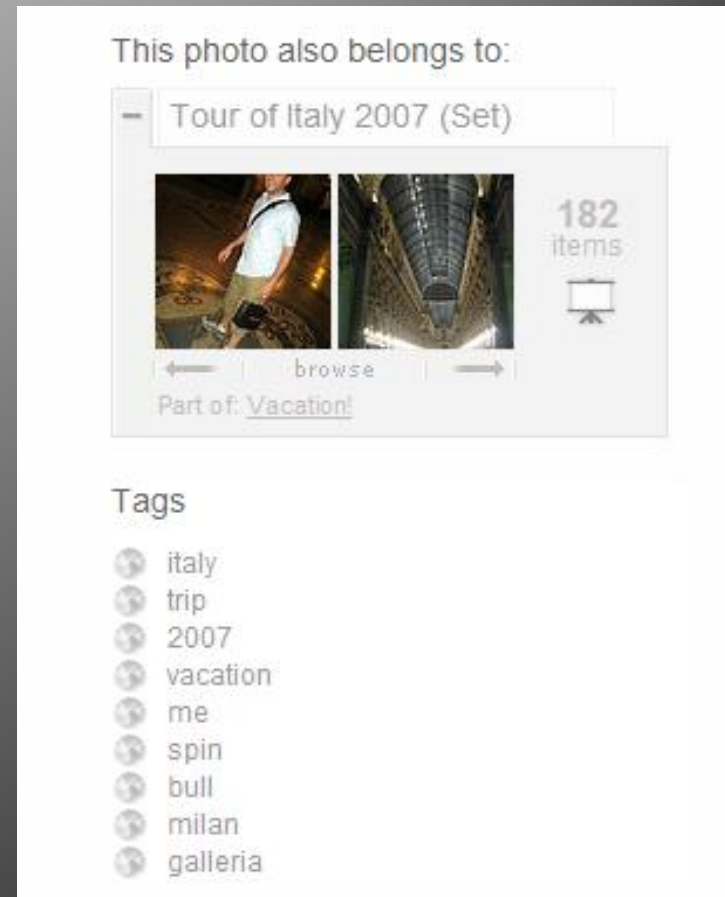




next... ruminate

# some thoughts about public tags

- people *do* seem to use tags differently than other forms of metadata
  - tags focus on place and are general in scope
  - captions focus on the story and are similar
  - titles integrate artifact, story, and place
  - tags are more apt to express personal context
- results confirm that tag recommendation should be approached with caution [Naaman and Nair, *IEEE Multimedia*, 2008]
  - based on frequency, {*Milan*, *Galleria*, *bull*, *balls*, *luck*, *spin*} would be a good tag set, but taggers might be avoiding particular terms (and I learned several new colloquialisms, in fact)
- scope of location-based tags is problematic (*Galleria* v. *Milan* v. *Italy* v. *Europe*...)



# implications



- captions and titles seem to be a better basis for retrieval and description from public collections...
  - elicit tags in a different way, one that encourages tag coverage (e.g. verbs)
  - use model metadata to get image query surrogate; subsequent queries use image similarity.
- geotagging + other processing might replace current tags (see, e.g. Crandall et al., WWW09)
- Flickr (and its ilk) might be a good starting point for excerpting art and architecture collections; metadata quality must be divorced from image quality



this time I've been much  
more polite and cautious,  
and it has diluted my argument.

But... the case study supports my  
original claim:

*“Tags don't have the folksonomic  
power people say they do”*

# what to do next holiday season?

- check results against datasets built from other types of images? (automate)
- check other types of metadata (e.g. comments) for comparable effects
- develop tag quality assessment mechanisms or heuristics?
- develop prototype tag elicitation interface?
- build a search interface that uses image surrogates?
- **stop worrying about tags and learn to do something constructive, say, knit?**



*drawing: Natalie Dee*

of course I'd welcome collaboration!





A photograph showing the lower legs and feet of a person standing on a highly decorative mosaic floor. The person is wearing brown strappy sandals. The mosaic features a central shield-shaped medallion with a blue background and a white figure, surrounded by a wide border of intricate geometric and floral patterns in shades of blue, red, and green. The text is overlaid in the center of the image.

questions?  
suggestions for related research?  
puzzled looks?