thoughts on a public tagging schema

This document was originally written for when I ran the Public Tag Repository. This is now run by users, so I am no longer an authority for it. I am briefly editing the page and leaving it as a record for some of my thoughts on tagging if you are interested. You can, of course, run your own tag repositories and do your own thing additionally or instead.

A newer guide and schema for the PTR is here.

seriousness of schema

This is not all that important; it just makes searches and cooperation easier if most of us can mostly follow some guidelines.

We will never be able to easily and perfectly categorise every single image to everyone's satisfaction, so there is no point defining every possible rule for every possible situation. If you do something that doesn't fit, fixing mistakes is not difficult.

If you are still not confident, just lurk for a bit. See how other people have tagged the popular images and do more of that.

you can add pretty much whatever the hell you want, but don't screw around

The most important thing is: **if your tag is your opinion, don't add it**. 'beautiful' is an unhelpful tag because no one can agree on what it means. 'lingerie', 'blue eyes', and 'male' or 'female' are better since reasonable people can generally agree on what they mean. If someone thinks blue-eyed women are beautiful, they can search for that to find beautiful things.

You can start your own namespaces, categorisation systems, whatever. Just be aware that everyone else will see what you do.

If you are still unsure about the difference between objective and subjective, here's some more examples:

- objective tags: (add these!)
  - firetruck
  - hors d'œuvre
  - high heels
  - character:jean-luc picard
  - person:okita anri
  - title:the tragical history of hamlet, prince of denmark
  - page:17
- subjective tags: (don't add these!)
  - awesome
  - faggot level:super-gay
  -
  - 4 stars
  - this is boring, why did anyone upload this here
  - moran communist and ONE TERM PRESIDENT!!! SARAH PALIN 2012! FOR JESUS CHRIST

Of course, if you are tagging a picture of someone holding a sign that says 'beautiful', you can bend the rules. Otherwise, please keep your opinions to yourself!

numbers

Numbers should be written '22', '1457 ce', and 'page:3', unless as part of an official title like 'ocean's eleven'. When the client parses and sorts numbers, it does so intelligently, so just use '1' where you might before have done '01' or '001'. I know it looks ugly sometimes to have '2 girls' or '1 cup', but the rules for writing numbers out in full are hazy for special cases.

(Numbers written as 123 are also readable by many different language-speakers, while 'tano', 'deux' and 'seven' are not.)

plurals

Nouns should generally be singular, not plural. 'chair' instead of 'chairs', 'cat' instead of 'cats', even if there are several of the thing in the image. If there really are *many* of the thing in the image, add a seperate 'multiple' or 'lineup' tag as apppropriate.

Ignore this when the thing is normally said in its plural (usually paired) form. Say 'blue eyes', not 'blue eye'; 'breasts', not 'breast', even if only one is pictured.

acronyms and synonyms

I personally prefer the full 'series:the lord of the rings' rather than 'lotr'. If you are an advanced user, please help out with tag siblings to help induce this.

character:anna (frozen)

I am not fond of putting a series name after a character because it looks unusual and is applied unreliably. It is done to separate same-named characters from each other (particularly when they have no canon surname), which is useful in places that search slowly, have thin tag areas on their web pages, or usually only deal in single-tag searches. For archival purposes, I generally prefer that namespaces are stored as the namespace and nowhere else. 'series:harry potter' and 'character:harry potter', not 'harry potter (harry potter)'. Some sites even say things like 'anna (disney)'. It isn't a big deal, but if you are adding a sibling to collapse these divergent tags into the 'proper' one, I'd prefer it all went to the simple and reliable 'character:anna'. Even better would be migrating towards a canon-ok unique name, like 'character:princess anna of arendelle', which could have the parent 'series:frozen'.

Including nicknames, like 'character:angela "mercy" ziegler' can be useful to establish uniqueness, but are not mandatory. 'character:harleen "harley quinn" frances quinzel' is probably overboard.

protip: rein in your spergitude

In developing hydrus, I have discovered two rules to happy tagging:

1. Don't try to be perfect.
2. Only add those tags you actually use in searches.

Tagging can be fun, but it can also be complicated, and the problem space is gigantic. There is always works to do, and it is easy to exhaust oneself or get lost in the bushes agonising over whether to use 'smile' or 'smiling' or 'smirk' or one of a million other split hairs. Problems are easy to fix, and this marathon will never finish, so do not try to sprint. The ride never ends.

The sheer number of tags can also be overwhelming. Importing all the many tags from the boorus is totally fine, but if you are typing tags yourself, I suggest you try not to exhaustively tag

everything in the image. You will save a lot of time and ultimately be much happier with your work. Anyone can see what is in an image just by looking at it--tags are primarily for finding things. Character, series and creator namespaces are a great place to start. After that, add what you are interested in, be that 'blue sky' or 'midriff'.

newer thoughts on presentation preferences

Since developing and receiving feedback for the siblings system, and then in dealing with siblings with the PTR, I have come to believe that the most difficult disagreement to resolve in tagging is not in what is in an image, but how those tags should present. It is easy to agree that an image contains a 'bikini', but should that show as 'bikini' or 'clothing:bikini' or 'general:bikini' or 'swimwear:bikini'? Which is better?

This is impossible to answer definitively. There is no perfect dictionary that satisfies everyone, and opinions are fairly fixed. My intentions for future versions of the sibling and tag systems is to allow users to broadly tell the client some display rules such as 'Whenever you have a clothing: tag, display it as unnamespaced' and eventually more sophisticated ones like 'I prefer slang, so show pussy instead of vagina'.

siblings and parents

Please do add siblings and parents! If it is something not obvious, please explain the relationship in your submitted reason. If it *is* something obvious (e.g. 'wings' is a parent of 'angel wings'), don't bother to put a reason in; I'll just approve it.

My general thoughts:

• siblings

In general, the correctness of a thing is in how it would describe itself, or how its creator would describe it.

For shorthand, I will say 'a'->'b' to mean 'a' is replaced by 'b'.

For instance, japanese names are usually written surname first and western forename first, so let's go 'character:rei ayanami'->'character:ayanami rei' but leave 'person:emma watson' and other western names as they are.

Unless it is too obscure, replace the english version of a word with any more proper or original foreign name. But stick to something a westerner can read. Do things like 'series:the melancholy of haruhi suzumiya'->'series:haruhi suzumiya no yuuutsu' or 'series:princess mononoke'->'series:mononoke hime'. There's even an argument for things like 'series:harry potter and the sorcerer's stone'->'series:harry potter and the philosopher's stone'.

Accents and other unusual/unicode characters are great in tags if they reflect the official marketed name, and should be preferred, but make sure there's an ascii->unicode sibling to make it easy for most users to type. 'series:pokemon'->'series:pok�mon' is excellent, as it both reflects official branding and also helps anyone who can't easily produce '�' on their keyboard find it.

I don't care about popularity as much as accuracy. Given 'series:pretty cure' and 'series:precure', I would prefer 'series:pretty cure' because it is the 'full and proper' rendering, even though there are more instances of 'precure' on the boorus.

Do correct for common plural mistakes. ear->ears, women->female, and so on.

And feel free to replace any 'character (series)' booru artifacts as with the 'anna (frozen)' example above. 'character:anna (frozen)'->'character:princess anna of arendelle' is great wherever it makes sense.

But please **do not** go 'blah'->'character:blah' unless the name is popular and unique. No one is going to be confused by 'ayanami rei'->'character:ayanami rei', but going 'archer'->'character:archer' is going to create a lot of false positives. There's a similar problem with something like 'character:mercy'->'character:angela "mercy" ziegler'--although the left hand side is namespaced, there are still plenty of *characters* named 'mercy', so a sibling that converts all Mercys to Overwatch's Mercy is not appropriate.

If the character name is the same as the series name, make the unnamespaced version go to the series version. For instance, set 'harry potter'->'series:harry potter', since we don't know which one it is and 'character:harry potter' ⊂ 'series:harry potter'. (If a picture of just Hermione that for some reason was not providing namespace information had 'hermione granger' (the character) and 'harry potter' (the series), we wouldn't want to infer 'character:harry potter' by accident.

In general, swap out slang for proper terms. 'lube'->'lubricant', 'series:zelda'->'series:the legend of zelda'.

parents

Be shy about adding character:blah->series:whatever unless you are certain the character name is unique. 'character:harry potter'->'series:harry potter' seems fairly uncontroversial, for instance, but adding specific sub-series just to be completionist, such as 'character:miranda lawson->series:mass effect: redemption' is asking for trouble.

Remember that parents define a relationship that is always true. Don't add 'blonde hair' to 'character:elsa', even though it is true in most files--add 'animal ears' to 'cat ears', as cat ears are always animal ears, no matter what an artist can think up.

Also, tag parents are only worth something if the parent is useful for searching. Adding 'medium:blue background'->'blue' isn't useful since 'blue' itself is not very valuable, but 'fishnet stockings'->'stockings' is useful as both tags are common and used in searches by plenty of people.

You can create a complicated tree like the firearms diagram on my parents page, but if it only adds seven tags that you probably wouldn't ever use yourself, you probably wasted your time.