Arrangements for Data Sharing and Reuse in Citizen Science Projects

"Data from the People for the People" Session
SCIDATACON 2014 @ New Delhi

Tyng-Ruey Chuang
Academia Sinica
Taipei, Taiwan

(CC) BY 4.0
Outline

- Lessons from Wikipedia
- Issues when using social media
- Experiences in making arrangements
- You gotta have apps!
- Toward best practices
Collaboration
Wikipedia

English
The Free Encyclopedia
4 000 000+ articles

Español
La enciclopedia libre
900 000+ artículos

Русский
Свободная энциклопедия
875 000+ статей

Italiano
L'enciclopedia libera
939 000+ voci

Português
A enciclopédia livre
742 000+ artigos

日本語
フリー百科事典
815 000+ 記事

Deutsch
Die freie Enzyklopädie
1 430 000+ Artikel

Français
L'encyclopédie libre
1 270 000+ articles

Polski
Wolna encyklopedia
908 000+ haseł

中文
自由的百科全書
500 000+ 篇目

Search · Suchen · Rechercher · Zoeken · Ricerca · Szukaj · Buscar · Поиск · 検索 · Busca · Sök · 搜尋 · Tim kiém · Пошук · Cerca · Sök · Haku · 講談 · Keresés · 찾기 · Cari · Ara · Căutare · تبحث · Hľadať · Seg · Serĉu · Претрепа · Paleška · Poišči · Cari · 查覈 · Търсене · iązdey · Bilatu · Suk · Bilinga · Traži · ถอย
Collaborative Works

• Works created and used by multiple members at the same time.

• Fluid in member composition; indefinite in the time span of one's participation.

• Materials contributed by collaborators and/or taken from other sources.

• Producing works of some social value.
Public Licenses

- "Some Rights Reserved"; Other rights to use the work are granted to the public with a license.

- The license is worldwide, royalty-free, non-exclusive, and perpetual.

- The licensor makes clear what is to be licensed, and under what license.

- A licensee is any person who agrees to the license, and uses the work.
Creative Commons Licenses

[Images of Creative Commons license icons for various combinations of BY, NC, ND, SA]
Search by License

http://flickrcc.bluemountains.net
公眾領域
PUBLIC DOMAIN
Collaborative Works + Public Licenses

• Participants make available some of their works.

• Wanted: The right to make modifications, and the obligation to share the results likewise.

• Arrangement: All participants agree to use a particular public license for the output of their collaboration.

• Whoever agrees to the license can participate!
Citizen Science Projects

• Not all contributors are equal?
  - scientists
  - participants

• Not all contributions are good?
  - data quality
  - personal data
  - sensitive information
  - BTW, Is this data or creative work?
Making Arrangements

• Put them into writing:
  – public licensed/domain; use of personal data; dealing with sensitive information
  – principles of data quality, integrity, and availability

• Community feedback and overseeing

• Use common legal, technical, and practical tools to easy the task

• Good provenance makes useful data products
Social Media
Terms of Service

• The pre-conditions one must agree to before using the provider's service:
  - rights to user contributions
  - use of personal data
  - level of service quality

• The service providers makes the rules!
  - (Citizen) Scientists are just users

• Users can make parallel arrangements for themselves.
Issues of Metadata and Provenance

Flickr and Facebook STILL Strip EXIF Data

Posted on April 22, 2010 by Jonathan Bailey • 31 Comments

In November 2008 I wrote a post that highlighted how Flickr was removing EXIF data in images. I received a comment to that post...
Your (embedded) metadata? But I run the services!

http://www.embeddedmetadata.org/social-media-test-results.php

### Social Media sites: photo metadata test results

We have uploaded a test image with embedded metadata and have checked if any of the embedded values are displayed and if so, which ones. As a next step we have tried to save the picture from a Social Media web site or system on our computer and then we checked what metadata fields are still embedded.

Find the test results in the grid below - in short: a green light indicates good results, a yellow one not so good ones and the red light that the handling of metadata should be improved, grey means: has not been tested.

You can read the full details of the tests on this page. It lists all details about what was tested and what the icons below stand for.

<table>
<thead>
<tr>
<th>Social Media site/system</th>
<th>Summary</th>
<th>Displays correctly?</th>
<th>Displays 4Cs?</th>
<th>Save As embedded?</th>
<th>Download embedded?</th>
</tr>
</thead>
<tbody>
<tr>
<td>500px - <a href="http://www.500px.com">www.500px.com</a></td>
<td>Most relevant fields shown, copyright was overridden, in downloaded file we only found Exif metadata. Save As is disabled.</td>
<td><img src="https://example.com" alt="Exif" /> <img src="https://example.com" alt="IPTC" /> <img src="https://example.com" alt="IPTC" /></td>
<td><img src="https://example.com" alt="Exif" /> <img src="https://example.com" alt="IPTC" /> <img src="https://example.com" alt="IPTC" /></td>
<td><img src="https://example.com" alt="Exif" /> <img src="https://example.com" alt="IPTC" /> <img src="https://example.com" alt="IPTC" /></td>
<td><img src="https://example.com" alt="Exif" /> <img src="https://example.com" alt="IPTC" /> <img src="https://example.com" alt="IPTC" /></td>
</tr>
<tr>
<td>Dropbox - <a href="http://www.dropbox.com">www.dropbox.com</a></td>
<td>No metadata shown, all embedded fields are preserved in the image files.</td>
<td><img src="https://example.com" alt="Exif" /> <img src="https://example.com" alt="IPTC" /> <img src="https://example.com" alt="IPTC" /></td>
<td><img src="https://example.com" alt="Exif" /> <img src="https://example.com" alt="IPTC" /> <img src="https://example.com" alt="IPTC" /></td>
<td><img src="https://example.com" alt="Exif" /> <img src="https://example.com" alt="IPTC" /> <img src="https://example.com" alt="IPTC" /></td>
<td><img src="https://example.com" alt="Exif" /> <img src="https://example.com" alt="IPTC" /> <img src="https://example.com" alt="IPTC" /></td>
</tr>
<tr>
<td>Facebook - <a href="http://www.facebook.com">www.facebook.com</a></td>
<td>Metadata not shown anymore, all embedded metadata stripped-off from image files.</td>
<td><img src="https://example.com" alt="Exif" /> <img src="https://example.com" alt="IPTC" /> <img src="https://example.com" alt="IPTC" /></td>
<td><img src="https://example.com" alt="Exif" /> <img src="https://example.com" alt="IPTC" /> <img src="https://example.com" alt="IPTC" /></td>
<td><img src="https://example.com" alt="Exif" /> <img src="https://example.com" alt="IPTC" /> <img src="https://example.com" alt="IPTC" /></td>
<td><img src="https://example.com" alt="Exif" /> <img src="https://example.com" alt="IPTC" /> <img src="https://example.com" alt="IPTC" /></td>
</tr>
<tr>
<td>Flickr FREE account - <a href="http://www.flickr.com">www.flickr.com</a></td>
<td>A few metadata fields shown, 'by' was overridden, for any downscaled rendition all embedded metadata are stripped-off from image files, only the Original rendition keeps all metadata.</td>
<td><img src="https://example.com" alt="Exif" /> <img src="https://example.com" alt="IPTC" /> <img src="https://example.com" alt="IPTC" /></td>
<td><img src="https://example.com" alt="Exif" /> <img src="https://example.com" alt="IPTC" /> <img src="https://example.com" alt="IPTC" /></td>
<td><img src="https://example.com" alt="Exif" /> <img src="https://example.com" alt="IPTC" /> <img src="https://example.com" alt="IPTC" /></td>
<td><img src="https://example.com" alt="Exif" /> <img src="https://example.com" alt="IPTC" /> <img src="https://example.com" alt="IPTC" /></td>
</tr>
<tr>
<td>Google+ - plus.google.com</td>
<td>Primarily Exif metadata shown, all embedded fields are preserved.</td>
<td><img src="https://example.com" alt="Exif" /> <img src="https://example.com" alt="IPTC" /> <img src="https://example.com" alt="IPTC" /></td>
<td><img src="https://example.com" alt="Exif" /> <img src="https://example.com" alt="IPTC" /> <img src="https://example.com" alt="IPTC" /></td>
<td><img src="https://example.com" alt="Exif" /> <img src="https://example.com" alt="IPTC" /> <img src="https://example.com" alt="IPTC" /></td>
<td><img src="https://example.com" alt="Exif" /> <img src="https://example.com" alt="IPTC" /> <img src="https://example.com" alt="IPTC" /></td>
</tr>
</tbody>
</table>
Comparision
A Collaboration on Moving Roadkill Data out of A Silo
一雙無法睜上的深邃眼珠，你看見了嗎～～
請和我們一同努力，讓悲劇不再發生！

四處爬爬走(路殺社, Reptile Road Mortality)
謝中和
2013/10/03 南蛇-幼蛇
台東縣池上鄉台20甲線約3.05k
緯度: N 23°7.873' (23°7'52.4")
經度: E 121°11.482' (121°11'28.9")

林志欣 #南蛇# "Ptyas mucosus"
10月7日 1:17 · 已編輯 · 誇

Water Chen 標本號RN1295 採集號ZH20131003 慶
11月18日 10:37 · 誇
台灣野生動物路死觀察網 (Taiwan Roadkill Observation Network)

標記之物種：
Ptyas mucosus
標記之俗名：
南蛇

臉書：

圖片作者：謝中和

2013/10/03 南蛇-幼蛇

目測路殺發生時約20公尺長約3.05m

緯度：N 23°7.873' (23°7'52.4")
經度：E 121°11'14.82" (121°11'28.9")

資料來源：
臉書社 (FB)

紀錄地點：
鳳林鎮, 恆春鎮
Taiwan
You Have Got App!
"Drupal In the Middle" App

• Requirements for a (road)killer App:
  – Written arrangement for all participants
  – We like images raw! (EXIF et. al.)
  – Provenance is paramount

• The App uploads raw images et. al. to Drupal
• Drupal pushes digests to the Facebook group
  – sanitized sensitive/personal data
  – link back to the originals in Drupal

• Work in Progress!
路殺社APP隱私權政策

適用範圍

以下的隱私權政策，適用於您在路殺社APP及其延伸服務網站活動時，所涉及有關個人資料的收集、處理及利用說明。

使用者於閱讀路殺社APP使用說明及隱私權政策之後，點選同意並繼續使用本網站之行為，視為已同意本政策所述之資料收集、處理、及利用方式。

路殺社APP如何收集使用者資料？

在使用路殺社APP的過程中，用戶所使用之瀏覽器將會記錄其所在地位點之經緯度，並在上傳照片時自動將相關資訊一併傳送至伺服器端。

若使用者未設置資訊屏蔽機制，則路殺社APP之伺服器，將為您保留用戶連線設備IP位址、登入時間，及其使用瀏覽器等相關資訊。

Privacy Policy  Public Licensing  Anonymous Posting
In Drupal

In Facebook
Openness
Openness?

- Libre
  - can be used by people

- Digital
  - can be used by machines and put online

- Raw
  - can be modified and repurposed

- Common (format & vocabulary)
  - can be exchanged and interlinked

- Transparent
  - (the process) can be fixed; meta-level
Openness as Best Practice

- Libre
- Digital
- Raw
- Common
- Transparent

Open Content, Open Data, Open Source, Open Collaboration, Open Science, ...
Fin