Dave Compton
Information Architect, News Platform Capabilities, Thomson Reuters

• Within Thomson Reuters
  – Working with Reuters News & Financial data for ~25 years
  – Reuters News / Media / Trading Systems
  – Co-designer of the Reuters Internet Delivery System / Media Platform
  – Also involved in Content Rights, Metadata, Linked/Open Data

• At the IPTC
  – TR Technical representative
  – NewsML-G2 Technical Lead
  – Contributor to: RightsML, News in JSON, Photo/Video Metadata, Semantic Exchange
  – Board member

• Find me at
  – dave.compton@thomsonreuters.com
  – http://www.linkedin.com/in/davecompton
  – https://twitter.com/dav3c
Protecting Metadata
Agenda

• Context
• Challenges
• Considerations
• Progressing Metadata within TR
• TR Pictures & Social Media
Protecting Metadata
Terminology

• Controlled Vocabulary
  – A set of controlled terms used to categorise content.

• Coding
  – Any value from a Controlled Vocabulary.

• Channel
  – A subset of all content, filtered by a combination of Coding and/or text matching.
Protecting Metadata

Context

As with many large agencies:

• TR has content … *lots of content*

• Fat pipes of data; **broad channels**

  e.g. Sport
  
  – Can be *specialised* by further filtering:
    
    e.g. Indian Cricket / Champions’ League / Ice Hockey

• Controlled Vocabulary terms applied as metadata to categorize the content

• Content **pushed** (by provider) or **pulled** (by consumer)
Protecting Metadata
Context /2

Business View

• B2B: TR -> Large Consumers: E.g. Yahoo, MS
  – Large consumers probably re-channelise the content

• B2C: TR -> Smaller Consumers / Internal Products
  – May not want to, or be able to, re-channelise
  – Need more granular information
Protecting Metadata
Context /3

Functional View

• Internal Use
  – **Broad** channelization
  – Ability to **forward/route** to the appropriate consumers

• External Use
  – ‘**Enough’ metadata** to enable consumers to **filter/search** the content
The Content Challenge
Protecting Metadata
The Content Challenge

• Multiple Media Types
  – Text, Pictures, Graphics, Interactive Graphics, Audio, Video
  – Packages, Events

• Increasing levels of content (items per day)
  – Text 2500
  – Picture 2200
  – Graphic ~100 Static, plus <10 Interactive
  – Video ~140 Online, plus >300 Broadcast

• Diverse content topics
  – Broad: Entertainment, Sport, Finance, Health, …
  – Niche: Reuters America Sports Pictures
Protecting Metadata
The Content Challenge /2

• Identifying Entities
  – **People**: Politicians, Celebrities, …
  – **Places**: Regions, Countries, Cities, …
  – **Organisations**: Public, Private

• Related information
  – ‘**More** like this’
  – Content associated with an **Event** or specific **Data**
  – Content with matching **Entities**
Considerations
Protecting Metadata
Considerations: Questions to Ask

Typical Questions

1. What is the content about?
2. Is there any related content?
3. Can I use the content (for a specific purpose)?
Protecting Metadata Considerations: Business Drivers

Why Use Metadata?

- Must consider both the Provider and the Consumers

**Provider**
- Provide targeted content
- Drive content usage
- Reduce exposure to content misuse

**Consumers**
- Improve content search/discovery
- Find related content
- Understand usage restrictions
Progressing Metadata within TR
Protecting Metadata
Progressing Metadata in TR: Initial Focus

Taking stock

• **Standardise** coding
  – Across Media Types: CAN: Canada? Canoeing?

• **Use identifiers** where possible
  – Entities / Topics / Related data [RIC vs. Org]

• **Synchronise** metadata
  – across binary asset renditions
  – across XML ‘sidecar’ files and binary assets
Protecting Metadata
Progressing Metadata in TR: Current Focus

Further Improvements

• **Use Qualified Coding**
  – Across Media Types: N2:CAN=Canada, PX:CANO=Canoeing

• **News / Events Integration**
  – Event Identifiers in Picture metadata

• **Machine Readable Rights**
  – Policy Identifiers in Picture metadata
Protecting Metadata
Social Media: Twitter

Pictures published to various *curated* Twitter feeds

- **Workflows** vary due to legacy systems

**News**

- [@Reuters](https://twitter.com/Reuters) Top News, Pictures & Video
- [@ReutersWorld](https://twitter.com/ReutersWorld) International News and Analysis
- [@ReutersAgency](https://twitter.com/ReutersAgency) Multimedia news for newsrooms / media orgs

**Photography**: Typically 20-40 Tweets / week

- [@ReutersPictures](https://twitter.com/ReutersPictures) News Photography: Global
Protecting Metadata
Social Media: Twitter @ReutersParisPix

Twitter: Feeds

> Select a specific Post
Protecting Metadata
Social Media: Twitter

Twitter: Specific Post

> Select the Picture Link …
Protecting Metadata
Social Media: Twitter

Reuters Pictures Site: Picture Gallery

> Select the Image …

RTSEXH8
18 May 2010
New York, UNITED STATES
An image of the Snapchat logo created with Post-it notes is seen in the windows of Havas Worldwide at 260 Hudson Street in Lower Manhattan, New York, May 18, 2016, where advertising agencies and other companies have started what is being called a "Post-it note war" with employees creating colorful images in their windows with Post-it notes. REUTERS/Mike Segar

RTSEXH4
18 May 2010
An employee creates a rainbow image on a window with Post-it notes.

RTSEXH5
18 May 2010
Images created with Post-it notes are seen in the windows.

RTSEXH7
18 May 2010
Images created with Post-it notes are seen in the windows.

RTSEXGH
18 May 2010
A message created with Post-it notes is seen in the windows.

RTSEXH3
18 May 2010
Employees create a U.S. flag image on a window with Post-it notes.
Protecting Metadata
Social Media: Twitter

Reuters Pictures Site: Selected Image

> Protected View: Add to cart … or Search again
Social Media Content Strategy
Protecting Metadata
Content Strategy

Free Content? Yes, and No …

• **Some content** is published via Social Media
  – Curated by Editors
  – Remaining content is paid for

• The free content contains *some of the metadata*
  – Free content may omit entity identifiers, links etc.
  – Paid content contains all the richer metadata and renditions

This applies to Pictures, and other Media Types
Summary
Protecting Metadata

Summary

Metadata is important to the **Provider** and the **Consumer**

- **Protection**
  - Do not give away all the content, or all the metadata

- **Visibility**
  - Utilise Social Media to drive traffic to pay sites using ‘free’ content

- **Discovery**
  - Metadata is key to search and filtering for Thomson Reuters and our Consumers

- **Monetisation**
  - Leverage quality content, related content and richer metadata
Protecting Metadata

Aim

Provide relevant content with the rich metadata required to support the Consumers’ needs

Broad Channels and Narrow Filtering

REUTERS/Edgar Su

REUTERS/Pichi Chuang
Any Questions?

Further Resources:
IPTC PhotoMetadata
Reuters News Pictures
@ReutersPictures
@ReutersParisPix

dave.compton@thomsonreuters.com
http://www.linkedin.com/in/davecompton
https://twitter.com/dav3c