IPTC G2-Standards & Video within Thomson Reuters
5th June 2014 | Photo Metadata Conference | CEPIC | Berlin
Dave Compton, Information Architect
News Platform Capabilities, Thomson Reuters
dave.compton@thomsonreuters.com
Dave Compton
Information Architect, News Platform Capabilities, Thomson Reuters

• Within Thomson Reuters
  – Working with Reuters News & Financial data for ~20 years
  – News / Media / Trading Systems
  – All aspects of feed & format handling and transformation
  – Co-designer of the Reuters Internet Delivery System / Media Platform
  – IPTC G2S Technical owner within TR
  – Also involved in Rights, Metadata, Linked/Open Data

• At the IPTC
  – TR Technical representative
  – NewsML-G2 Technical Lead
  – Contributor to: RightsML, News in JSON, Semantic Exchange

• Find me at
  – dave.compton@thomsonreuters.com
  – http://www.linkedin.com/in/davecompton
  – https://twitter.com/dav3c
“Speed is nothing without Accuracy” – Reuters
“Speed is nothing without Accuracy” – Reuters

... this includes Video Metadata
G2 Video in TR

Agenda

• TR & the G2-Standards (G2)
• TR Video
• TR Video Metadata
• Identifying the Pieces
• A side(car) helping of Metadata
• Show me the Data
• Note ...

Salient Information

Geek Alert!
G2 Video in TR
Why did TR choose G2?

• G2 ticks all the Requirements
  – Cross-domain: News, Events ... Rights, Linked Data
  – Multimedia: Text, Picture, Graphic, Audio, Video
  – Content Representation: Inline and/or referenced
  – Consistent Metadata: Administrative, Descriptive, Rights
  – Clarity: Qualified coding
  – Flexibility: Items, Packages, Events, Vocabularies
  – Extensibility: Other namespaces, Extension Points
  – Management: Identification, Versioning, Linking
  – Rights: Copyright, Usage ... Integrate with MRR
G2 Video in TR
Is TR actually using G2?

• Yes: Currently
  – Internally
    • New Text head-end generates G2 *natively*
    • Combination of *native generation* and *transformation*
    • Used across *News* and *Events*
    • Used for the TR *subject coding* scheme definition
  – Externally
    • All ‘simple’ *Multimedia*
    • All *Packages, including Video* and Multimedia

• Yes: In the future
  – Internally
    • New *Picture/Graphic* and *Video* head-ends will generate G2 *natively*
G2 Video in TR Considerations

• Metadata field **semantics** need clearly defining [Source]
• Need a standard **PLUS Guidelines** [Spec + Usage + ?]
• Can a **minimal metadata property set** be defined?
• **Confidence, Relevance and Significance** all help
• Different consumers have **different requirements**
  – Metadata, THEN choose media: [Video?]
  – Media WITH metadata: [Images]
• Don’t forget the **limitations of existing systems’** entry
• Don’t assume **legacy formats** will go away soon
• **Assume nothing ...**
G2 Video in TR
Editorial – Video Stats

• 350 Camera crews globally
• 120 Bureaux globally
• 250 – 350 stories filed per day averaging 2 minutes each
• 60%-80% of the content shot by Reuters
• 20%-40% of content up-picks (3rd party content)
• Mixture of Breaking News & Scheduled events
G2 Video in TR
Editorial – Video Workflow

Reuters Camera Crew with Local News Gathering (LNG) laptop

Video sent over satellite Or file over the Internet

Producer in local bureau edits video and writes script

Edited video and script contributed to WNE over Internet

Script and Video sent to clients via satellite or Internet
G2 Video in TR
Editorial – Video Systems

• WNE: World News Express
  – BCV: Broadcast Video

• TVDAMS: TV Digital Asset Management System
  – OLV: Online Video
G2 Video in TR
WNE vs. TVDAMS

- BCV and OLV Processing

Video Flow
- v0.0.03, 06/10/11, D J Compton, NMG
- Master in: REF-DIAGRAMS
G2 Video in TR
Broadcast vs. Online

- BCV and OLV G2 Representation

BCV vs. OLV
- v0.0.02, 06/10/11, D J Compton, NMG
- Master in: REF-DIAGRAMS
G2 Video in TR
Metadata – The Usual Suspects

• Rights / Usage Metadata
  – Who should be attributed?
  – Do any restrictions apply?

• Item Metadata
  – Is the Item usable?
  – When was this item version created?

• Administrative / Descriptive Metadata
  – When was the content created?
  – Which parties contributed to the content?
  – What languages are used in the video?
  – What is the video about?
G2 Video in TR
Metadata – Value Added

• Associations
  – What **Items** are in this Package?

• Relationships
  – Was this Item / Package created from another resource?

• Per-Part Metadata
  – What metadata is specific to a ‘Shot’ within a ‘Clip’?
G2 Video in TR
Metadata – Video-specific

• Basic
  – Rendition
  – Content Type
  – Width / Height
  – Duration

• Audio
  – Bit Rate / Sample Rate / Variable Bit Rate Flag / Codec

• Video
  – Average Bit Rate / Scan / Sampling / Variable Bit Rate Flag / Aspect Ratio / Codec
Identifying the Pieces
G2 Video in TR

Text was easy ... sort of

- Every Item* has a TR-UNIQUE Identifier
  
  Each set of updates for a given Item has the SAME Identifier
  
  - Item A

- Each set of updates for a given Item has a GLOBALLY UNIQUE Identifier

  Every Item has a UNIQUE Version
  
  - Reuters.com: Item A, Version 1 ... Item A, Version 2 ...

* Item or Package
G2 Video in TR
Text was easy ... sort of 2

• Every Item has a Slugline
  – Item A, Version 1, Slugline 1

• Each Item has a UNIQUE TR Filename
  – Based on: Item A, Version 1, Slugline 1

But we need more for Multimedia ...
G2 Video in TR
Extending for Multimedia

• To include
  – Picture, Graphic, Audio, Video

• Each Content Rendition has a UNIQUE Rendition Name
  – Item A, Version 1, Rendition R1

• Each Content Rendition has a UNIQUE TR URI
  – reuters.com, Item A, Version 1, Rendition R1

• Each Content Rendition has a UNIQUE TR Download URL
  – Based on: reuters.com, Item A, Version 1, Rendition R1
G2 Video in TR Multimedia Example

- Each Item has a TR-UNIQUE Identifier

Each set of updates for a given Item has the SAME Identifier

<altId type="idType:SID" rtr:isOriginal="1">LOVEA4C17HW5A</altId>

- Each set of updates for a given Item has a GLOBALLY UNIQUE Identifier

Every Item has a UNIQUE Version

<newsItem guid="tag:reuters.com,2014:newsml_LOVEA4C17HW5A" version="1" ...>
G2 Video in TR
Multimedia Example 2

• Each Item has a Slugline
  <slugline xml:lang="en">ANTARCTICA-USA-ICE LOSS-TRACKED (O)</slugline>

• Each Item has a UNIQUE TR Filename
  <fileName>2014-05-13T153933Z_1_LOVEA4C17HW5A_RTRMADV_0_ANTARCTICA-USA-ICE-LOSS-TRACKED-O.XML</fileName>
G2 Video in TR
Multimedia Example 3

• Each Content rendition has a UNIQUE Rendition Name
  <remoteContent rendition="rend:stream:6756:16x9:mpg" …

• Each Content rendition has a UNIQUE TR Filename
  <altId type="idType:fileBIN">2014-05-13T153933Z_1_LOVEA4C17HW5A_RTRMADV_STREAM-6756-16X9-MPG_ANTARCTICA-USA-ICE-LOSS-TRACKED-O.MPG</altId>

• Each Content rendition has a UNIQUE TR URI
  <remoteContent residref="tag:reuters.com,0000:binary_LOVEA4C17HW5A-STREAM:6756:16X9:MPG" …

• Each Content rendition has a UNIQUE TR Download URL
G2 Video in TR
Multimedia Example - Summary

• Identification

Content ID: LOVEA4C17HW5A
Item ID: tag:reuters.com,2014:newsml_LOVEA4C17HW5A

Slugline: ANTARCTICA-USA-ICE LOSS-TRACKED (O)

Item Filename: 2014-05-13T153933Z_1_LOVEA4C17HW5A_RTRMADV_0_ANTARCTICA-USA-ICE-LOSS-TRACKED-O.XML

Rendition: rend:stream:6756:16x9:mpg

A side(car) helping of Metadata
G2 Video in TR
Together, separately

• Create a G2 Item to describe each multimedia component
  – Text: Script
  – Video: Video Rendition(s)
  – Picture: Keyframe Rendition(s)
  – Audio: Audio Track Rendition(s)

• Create a G2 Package to group the multimedia components
  – Associated the multimedia components

• Create any related multimedia content rendition(s)
  – Referenced by the multimedia components
G2 Video in TR
Delivering benefit

BC Video Package

Text Item (Script)

Video Item (BCV)

OL Video Package

Text Item (Script)

Video Item (OLV)

Audio Item

Picture Item (Script)

MPG, MP4
3GP, 3G2
ASF
FLV

MP3

JPG

OLV Items
- v0.0.02, 03/06/14, D J Compton, PlatCap
- Master in: REF-DIAGRAMS
G2 Video in TR
Accessing Multimedia Content

• Reuters Media Connect Platform – Architecture

Feed Processing
- v0.0.02, 24/10/08, D J Compton, Reuters Media
- Master in: CS / NFR006
G2 Video in TR
Accessing Multimedia Content

• Supports
  – Push: FTP
  – Pull: Content Downloader
  – Hosted: MediaExpress
  – API: Web Services

• Querying by
  – Video package
  – Item
  – Channel
  – Keyword
  – More ...
G2 Video in TR

Summary

• Leveraged G2 across all Multimedia
  – For ease of adoption

• Standardised Metadata across Media Types / Renditions
  – For a more consistent content experience

• Defined and included relationships
  – To facilitate content discovery & association

• Provided multiple content access methods
  – To maximise content retrieval and usage
G2 Video in TR

Next Steps

• G2 Metadata
  – Ongoing refresh of head-ends: Picture, Graphic, Audio, Video
  – Enhance per-shot properties. Esp. Rights
  – Enhance Item::Item and Item::Resource data
  – Enhance Video metadata: better, deeper
  – Expose keywording

• Discoverability
  – Enhance/link: People, Places, Locations ...
  – Link: News :: Events :: Data

• Binary Metadata
  – Adding (more) metadata to binary media
Any Questions?

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