

# Transparent Metadata Capture for Environmental Science

---

Wade Sheldon

*Georgia Coastal Ecosystems LTER*  
*University of Georgia*

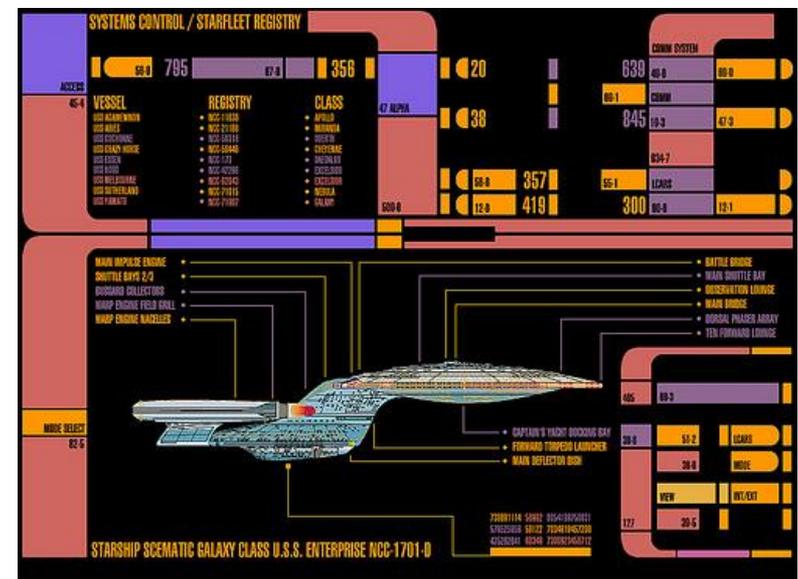
sheldon@uga.edu

ECOLOGICAL RESEARCH FROM A TRULY LONG-TERM PERSPECTIVE:  
WHAT NEEDS TO BE DONE TODAY TO SUPPORT ECOLOGICAL ANALYSIS IN 2100

- If current trends continue, in 2100 we can expect:
  - Environmental information systems will have access to vast stores of data from automated sensor systems
  - Scientists and the public will interact with these systems using natural human interfaces, including touch, speech and 3-dimensional vision
  - Natural language will replace technical query languages and syntax

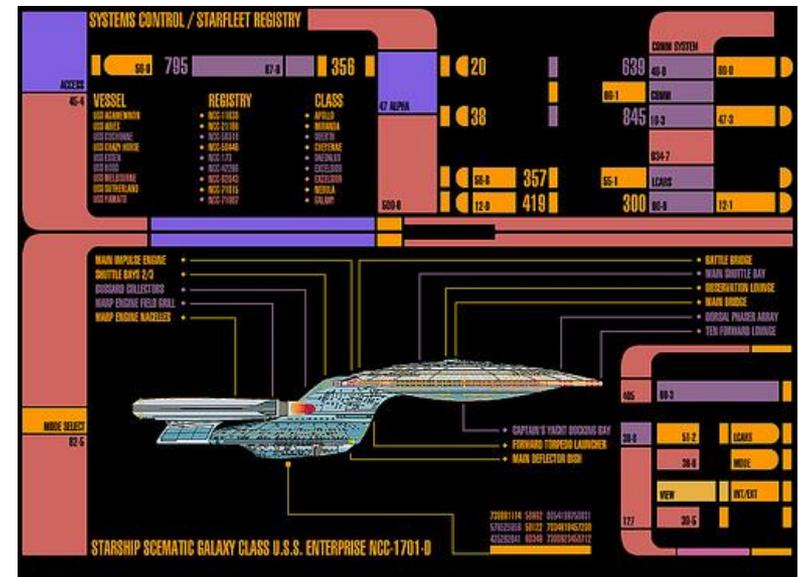
### ECOLOGICAL RESEARCH FROM A TRULY LONG-TERM PERSPECTIVE: WHAT NEEDS TO BE DONE TODAY TO SUPPORT ECOLOGICAL ANALYSIS IN 2100

- If current trends continue, in 2100 we can expect:
  - Environmental information systems will have access to vast stores of data from automated sensor systems
  - Scientists and the public will interact with these systems using natural human interfaces, including touch, speech and 3-dimensional vision
  - Natural language will replace technical query languages and syntax
- In short, we will be approaching the “Star Trek” era of computing where concept replaces syntax and speech replaces text as the means of requesting information



### ECOLOGICAL RESEARCH FROM A TRULY LONG-TERM PERSPECTIVE: WHAT NEEDS TO BE DONE TODAY TO SUPPORT ECOLOGICAL ANALYSIS IN 2100

- If current trends continue, in 2100 we can expect:
  - Environmental information systems will have access to vast stores of data from automated sensor systems
  - Scientists and the public will interact with these systems using natural human interfaces, including touch, speech and 3-dimensional vision
  - Natural language will replace technical query languages and syntax
- In short, we will be approaching the “Star Trek” era of computing where concept replaces syntax and speech replaces text as the means of requesting information
- To get there, we need ubiquitous and sophisticated metadata for **all** environmental data we collect



ECOLOGICAL RESEARCH FROM A TRULY LONG-TERM PERSPECTIVE:  
WHAT NEEDS TO BE DONE TODAY TO SUPPORT ECOLOGICAL ANALYSIS IN 2100

- Today, metadata entry is often the choke point in the data lifecycle, preventing many data from ever reaching archives
- The chief culprits:



ECOLOGICAL RESEARCH FROM A TRULY LONG-TERM PERSPECTIVE:  
WHAT NEEDS TO BE DONE TODAY TO SUPPORT ECOLOGICAL ANALYSIS IN 2100

- Today, metadata entry is often the choke point in the data lifecycle, preventing many data from reaching archives
- The chief culprits:
  - High volumes of sensor data with minimal or unstructured metadata
  - Too much reliance on tedious, manual metadata entry



ECOLOGICAL RESEARCH FROM A TRULY LONG-TERM PERSPECTIVE:  
WHAT NEEDS TO BE DONE TODAY TO SUPPORT ECOLOGICAL ANALYSIS IN 2100

- To break this bottleneck, we need:
  - *Transparent metadata capture* from environmental sensors
  - Analytical software that can read **and** use sensor metadata
  - Data management software that can augment sensor metadata with research context information from multiple sources
- Put another way, metadata must **enable** scientific data archiving, not **inhibit** it
- To solve this problem we need to look outside the box for alternative strategies

### ECOLOGICAL RESEARCH FROM A TRULY LONG-TERM PERSPECTIVE: WHAT NEEDS TO BE DONE TODAY TO SUPPORT ECOLOGICAL ANALYSIS IN 2100

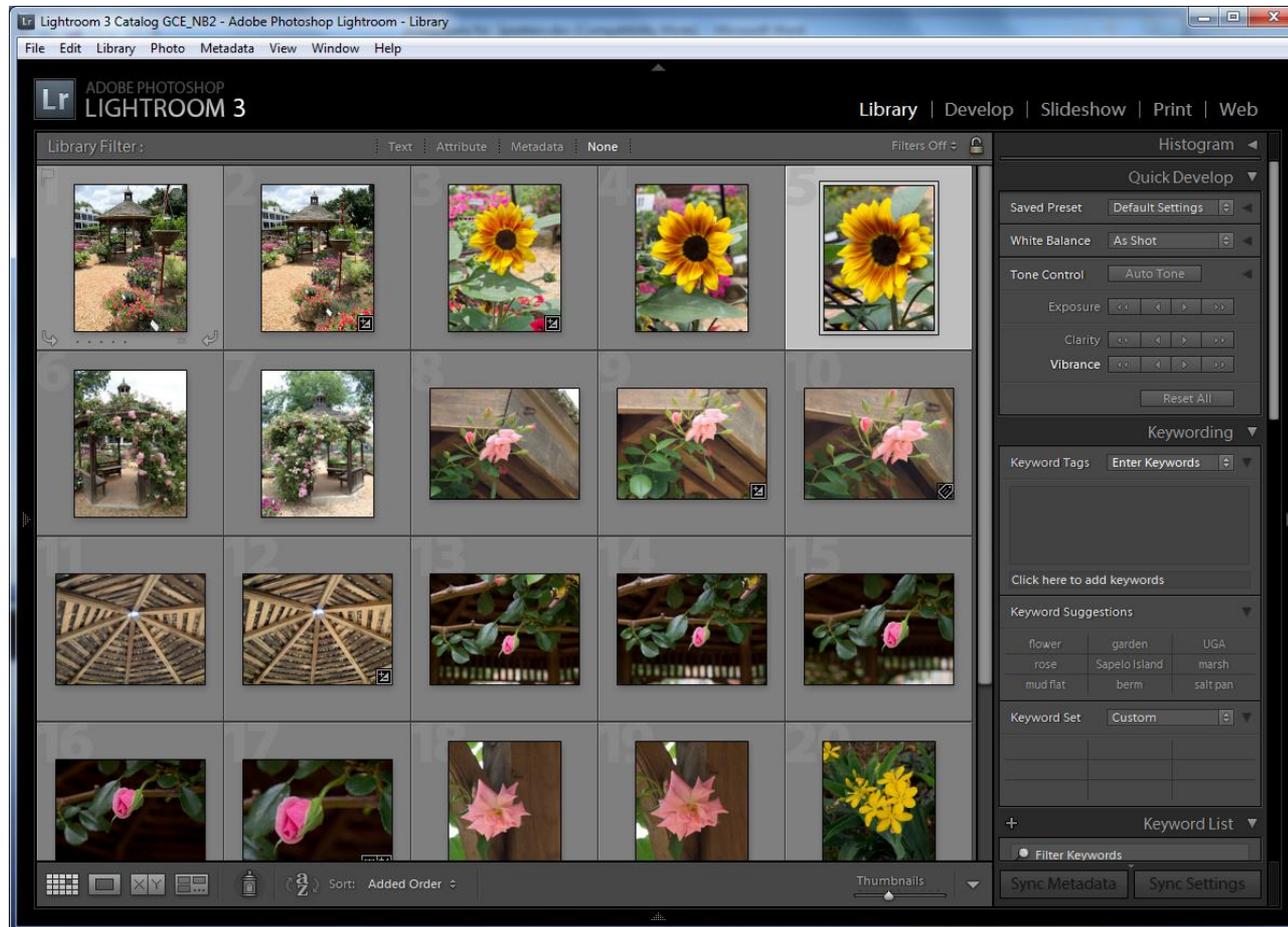
- Digital photography provides a great example of what is possible
- Every click of the shutter is a metadata entry (EXIF, IPTC, GPS)



Tag	Value
Manufacturer	CASIO
Model	QV-4000
Orientation (rotation)	top - left [8 possible values <sup>[14]</sup> ]
Software	Ver1.01
Date and Time	2003:08:11 16:45:32
YCbCr Positioning	centered
Compression	JPEG compression
x-Resolution	72.00
y-Resolution	72.00
Resolution Unit	Inch
Exposure Time	1/659 sec.
FNumber	f/4.0
ExposureProgram	Normal program
Exif Version	Exif Version 2.1
Date and Time (original)	2003:08:11 16:45:32
Date and Time (digitized)	2003:08:11 16:45:32
ComponentsConfiguration	Y Cb Cr -
Compressed Bits per Pixel	4.01
Exposure Bias	0.0
MaxApertureValue	2.00
Metering Mode	Pattern
Flash	Flash did not fire.
Focal Length	20.1 mm
MakerNote	432 bytes unknown data
FlashPixVersion	<a href="#">FlashPix</a> Version 1.0
Color Space	sRGB
PixelXDimension	2240
PixelYDimension	1680
File Source	DSC
InteroperabilityIndex	R98

ECOLOGICAL RESEARCH FROM A TRULY LONG-TERM PERSPECTIVE:  
WHAT NEEDS TO BE DONE TODAY TO SUPPORT ECOLOGICAL ANALYSIS IN 2100

- Editing the photo adds even more metadata (e.g. XMP, ICC)

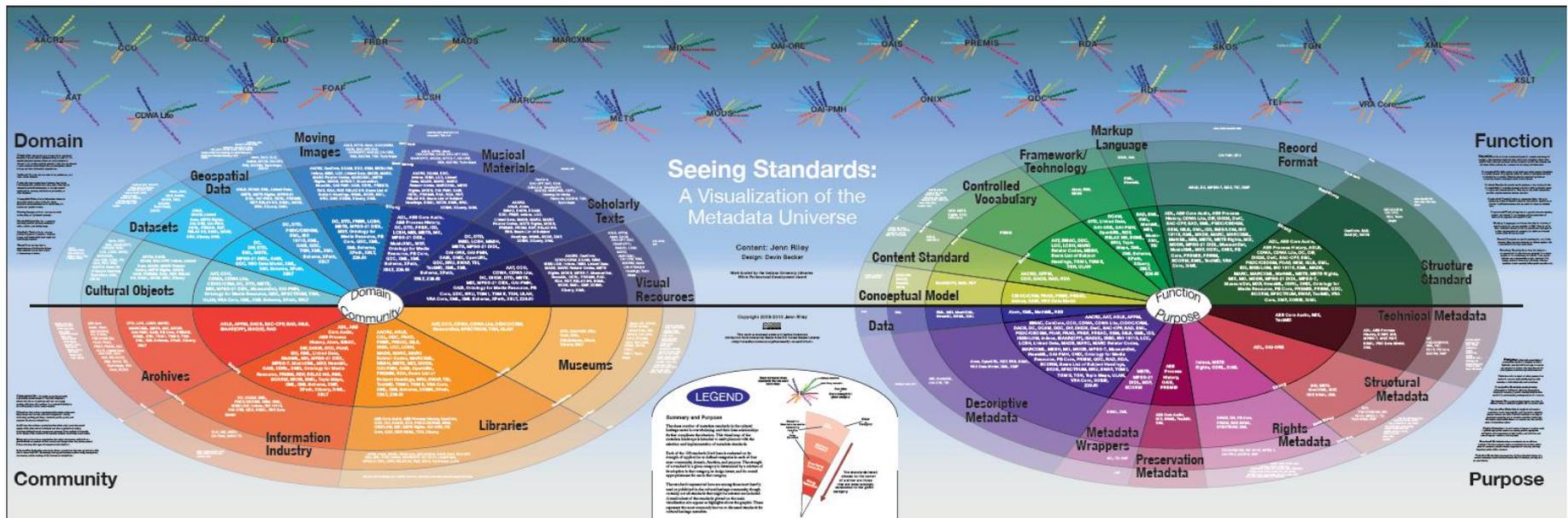


ECOLOGICAL RESEARCH FROM A TRULY LONG-TERM PERSPECTIVE:  
WHAT NEEDS TO BE DONE TODAY TO SUPPORT ECOLOGICAL ANALYSIS IN 2100

- The latest software is adding even more potential for automated metadata creation (e.g. facial recognition, place name lookup)
- This automated metadata capture enables rapid advances in software functionality and user experience with **NO** additional effort by the creator
- To achieve this level of functionality with environmental sensor data, we need to:
  - Lobby sensor and data logger manufacturers to adopt modern standards for embedded metadata
  - Promote development of more metadata-aware software
  - Develop metadata augmentation services to enhance content without additional effort by scientists

## ECOLOGICAL RESEARCH FROM A TRULY LONG-TERM PERSPECTIVE: WHAT NEEDS TO BE DONE TODAY TO SUPPORT ECOLOGICAL ANALYSIS IN 2100

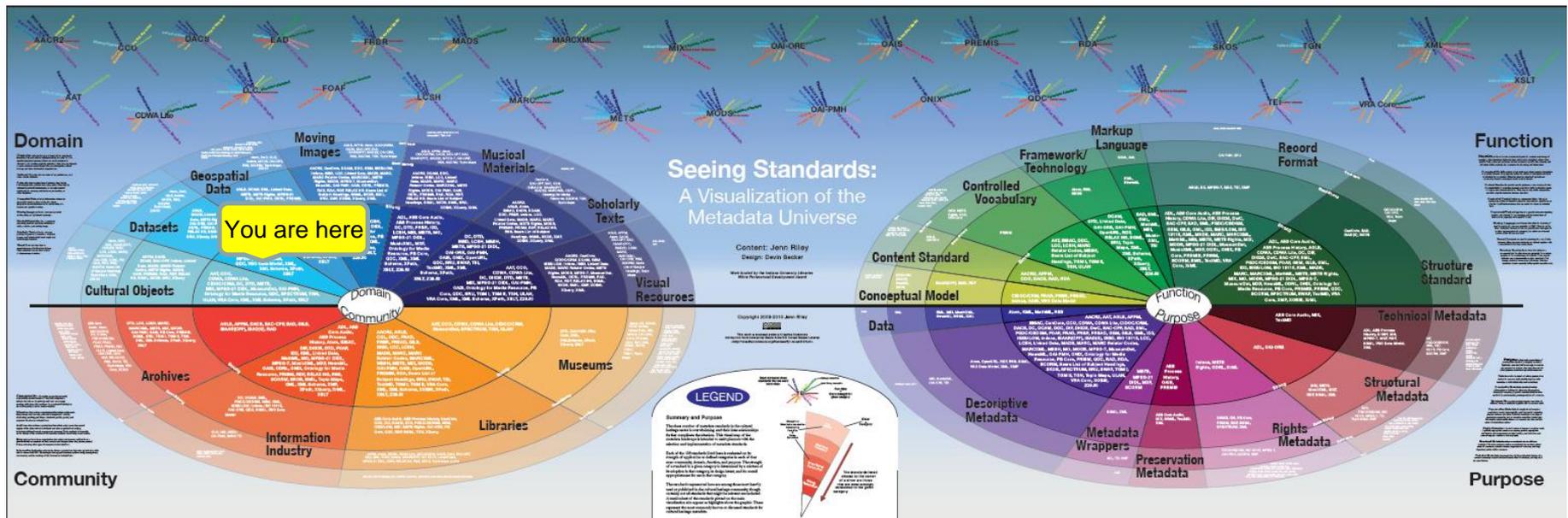
- Main barrier to success:
  - Too many metadata “standards” inhibit rather than accelerate progress
  - We need to remember that “The perfect is the enemy of the good” (Voltaire)



Visual map of metadata standards (J. Riley, Indiana University, 2010)  
<http://www.dlib.indiana.edu/~jenlrile/metadatamap/>

## ECOLOGICAL RESEARCH FROM A TRULY LONG-TERM PERSPECTIVE: WHAT NEEDS TO BE DONE TODAY TO SUPPORT ECOLOGICAL ANALYSIS IN 2100

- Main barrier to success:
  - Too many metadata “standards” inhibit rather than accelerate progress
  - We need to remember that “The perfect is the enemy of the good” (Voltaire)



Visual map of metadata standards (J. Riley, Indiana University, 2010)  
<http://www.dlib.indiana.edu/~jenrile/metadatamap/>

ECOLOGICAL RESEARCH FROM A TRULY LONG-TERM PERSPECTIVE:  
WHAT NEEDS TO BE DONE TODAY TO SUPPORT ECOLOGICAL ANALYSIS IN 2100

- But promising advances are still being made despite these challenges

