# Activity 4: What you see and what you get: implementing metadata

You are creating a dataset to track incidents involving police response to dissent in China.

Here are some of the metadata elements defined for this dataset:

* Title of incident
* Type of incident
* Description of incident
* Cause of incident
* Response to incident
* Keywords
* Sources used

## Task #1: Metadata creation

Read the three source documents in the Lhasa 2008 packet.

Use that data to create metadata about the incident in the table below.

| **Element** | **Value** |
| --- | --- |
| Title of incident |  |
| Type of incident |  |
| Description of incident |  |
| Cause of incident |  |
| Response to incident |  |
| Keywords |  |
| Sources used |  |

## Task #2: Identifying interpretive diversity in metadata implementation

In small groups, share the metadata that you generated.

What are some of the ways that your implementation of each element differed?

The following table lists some common dimensions of difference. Which appeared for which elements? Did other types of difference also occur?

| **Dimension of difference** | **Scope** | **Examples** |
| --- | --- | --- |
| Syntax | Includes differences in capitalization, punctuation, parts of speech, word order, verb tense, and plurality. | * Dalai Lama vs. Dalai lama * Tibet vs. Tibetan vs. Tibetans * “Demonstrators were taken into custody” vs. “Demonstrators are in custody” |
| Specificity | Includes differences in level of generality and differences in using kinds or instances as descriptors. | * China vs Lhasa * religion vs. Tibetan Buddhism * newspaper vs. *New York Times* * social media vs. Facebook * color vs. purple |
| Exhaustivity | Includes differences in level of detail. | * A one-sentence description vs. a paragraph description * One keyword vs. six keywords * All sources listed (B’s Twitter account, Y’s Facebook page, Z’s YouTube channel) vs. a term that indicates the kind of sources used (“social media”) |
| Synonymy | Includes the use different terms to indicate the same or very similar concepts. | * Turquoise vs. teal * Plaza vs. square |
| Perspective | Includes differences in focus, emphasis, and interpretation. | * “Protest” vs. “demonstration” vs. “riot” |
| Structure | Includes differences in format. | * 06/13/2016 vs. June 13, 2016 * Lhasa, Tibet, China vs. China (Lhasa) |

## Task 3: Developing controlled vocabularies

One way to reduce interpretive diversity in metadata implementation is to use controlled vocabularies for metadata values.

When developing a controlled vocabulary, here are some basic design guidelines:

* Values should be jointly exhaustive: the included values should describe any possible item.
* Values should be mutually exclusive: the included values shouldn’t overlap.
* Values should be at the same level of abstraction: some values shouldn’t be very general while others are specific; some values shouldn’t be kinds while others are instances.
* Value labels should be clear and easy to understand, using the simplest possible language.
* Value labels should be grammatically parallel (for example, all adjectives, all nouns).

As we know from our foray into online dating vocabularies, these guidelines may not be possible in all situations.

In this task, you’ll develop a controlled vocabulary for incident types.

Your vocabulary should encompass the following examples of incidents:

* Participating in a protest march.
* Writing on a blog.
* Writing letters to government officials.
* Meeting with foreign journalists.
* Going on a hunger strike.
* Initiating a lawsuit.
* Putting flyers on buses.
* Organizing an economic boycott.
* Contributing posts to Sina Weibo.
* Defending an accused person in legal proceedings.
* Taking pictures of someone being arrested.
* Throwing rocks at police.
* Spray-painting graffitti in a public place.
* Petitioning a local government.
* Selling foreign books.
* Practicing falun gong exercises in a park.
* Signing a manifesto.
* Organizing a work stoppage.
* Attending a clandestine prayer service.
* Making a speech.
* Accusing an official of corruption.
* Lying in front of a bulldozer.

Write your controlled vocabulary, and any instructions for its use, on a blank sheet of paper.

Should you find this easy, create another controlled vocabulary for source types.

Your vocabulary should incorporate the following types of sources:

* An article in the *New York Times.*
* A comment to a story posted on the *New York Times.*
* A video linked from a Sina Weibo post.
* A post on Tencent Weibo.
* ACLED (Armed Conflict Location and Event Database).
* A WeChat conversation.
* A corporate press release.
* An official government press release.
* A statement given by a government official.
* A report issued by Xinhua news agency.
* A photo posted on Qzone.
* A personal Web site.
* A corporate Web site.
* A government Web site.
* An article in a scholarly journal.
* A television news report from Taiwan.
* The Twitter account of an American journalist who covers China.
* A story in a regional newspaper in China.
* A podcast.
* An article in *Le Monde.*
* An article in *Caixin*, a Chinese news magazine.

## Task 4: Apply another group’s controlled vocabulary

In this task, you’ll take the controlled vocabulary for incident types created by another group and use it to create metadata for a new incident.

Read the source document in the Haikou 2016 packet.

Individually, use the Incident Type controlled vocabulary that you received from the other group to assign a type to the incident described in the source document.

As a group, compile a list of all the incident types you assigned, and write down your individual metadata decisions. Return the list of your decisions to the group whose vocabulary that you used.

If you have extra time, read the 2 source documents in the Beijing 2016 packet and use the Incident Type vocabulary to assign types to the incident documented in those two documents.

Did your group all understand and use the new vocabulary in the same way?

Did the other group use your vocabulary as you expected?