

---

# **Introduction to Metadata**

## **What is it and why does it matter?**

---

Presentation by Cynthia Heider and Brie Gettleson,  
University of Pennsylvania

---

# What you'll learn today

---

- What metadata is and why it is relevant to you
  - How to create metadata and use it in your own projects with your own materials
  - How to share your metadata for the benefit of others
  - Where to go if you want to learn more about metadata
-

# Session roadmap

---

METADATA, BROADLY

METADATA OPPORTUNITIES  
& LIMITATIONS

PRACTICAL TIPS FOR  
METADATA WORK

METADATA, SPECIFICALLY

APPLYING METADATA  
PROTOCOLS

WHERE CAN YOU  
LEARN MORE?

# Presenters

---



**Brie Gettleson,**  
Latin American  
Studies Librarian at  
the University of  
Pennsylvania



**Cynthia Heider,**  
Public Digital  
Scholarship Librarian  
at the University of  
Pennsylvania

1

# Metadata, Broadly

Defining metadata and related terminology

---

# METADATA, BROADLY

Defining metadata and related terminology

---

## What is metadata?

- "Data about data"
- Metadata is everywhere

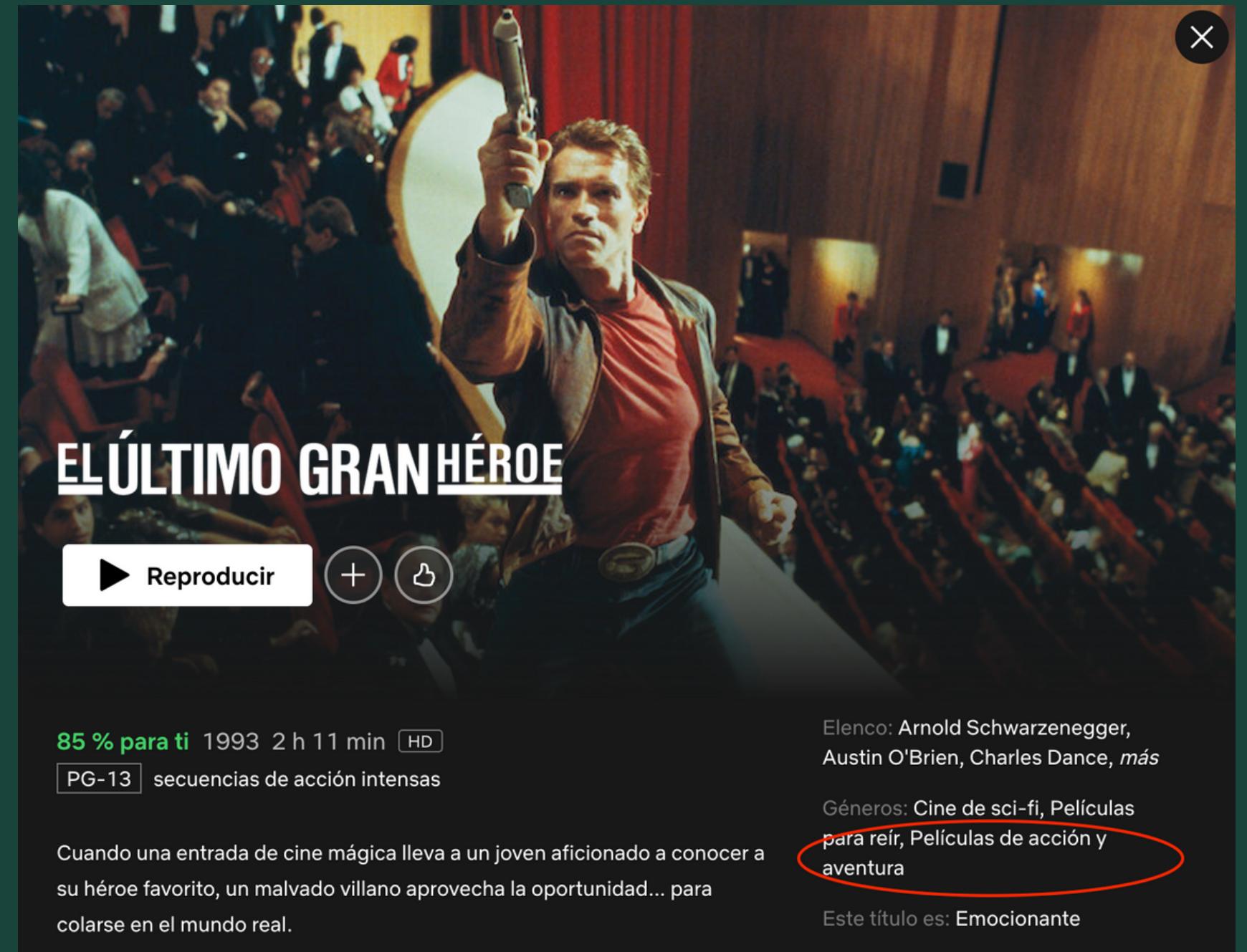


# METADATA, BROADLY

Defining metadata and related terminology

## What is metadata?

- "Data about data"
- Metadata is everywhere
- **Example: Looking for an action film for movie night**

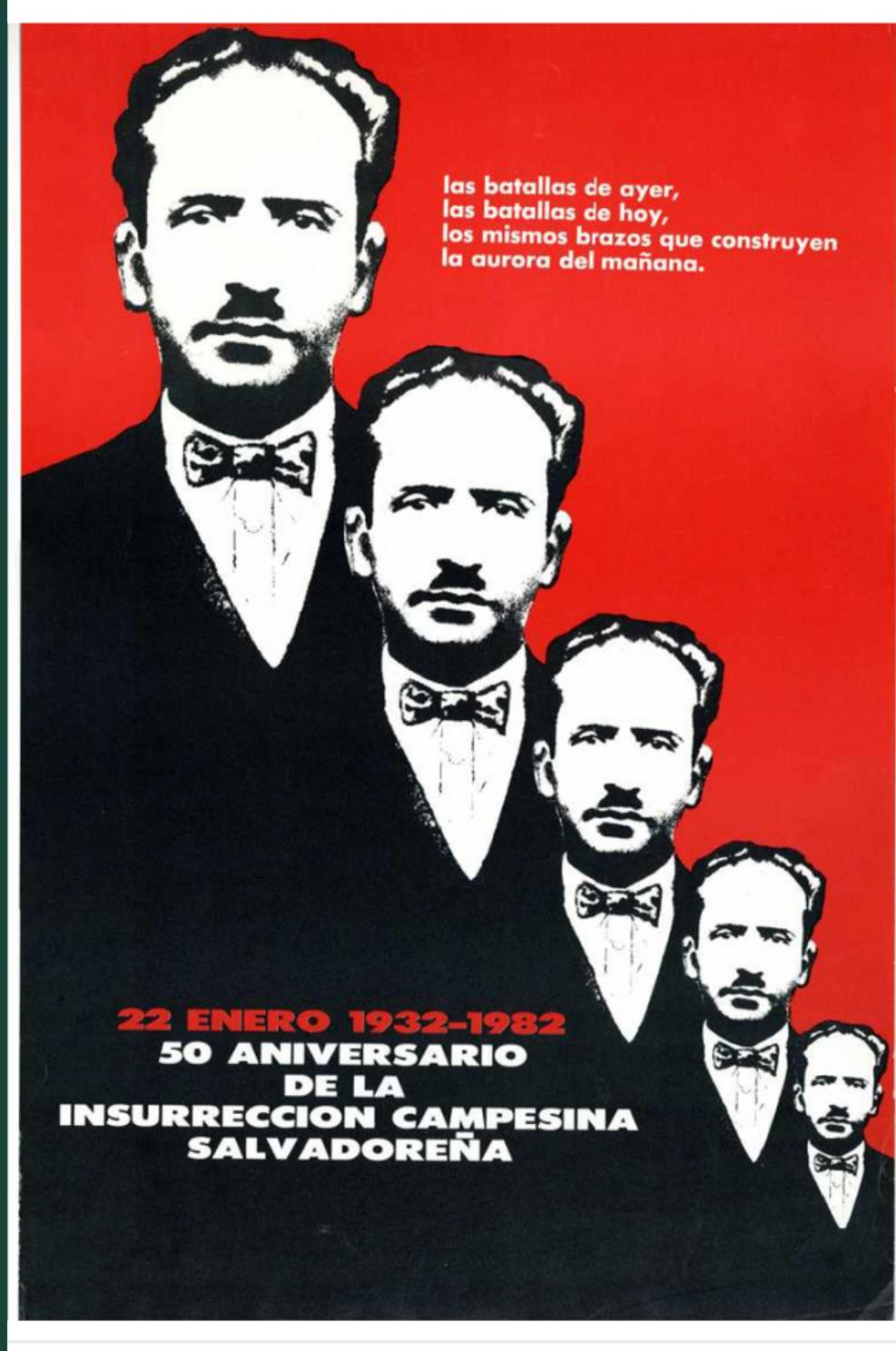


# METADATA, BROADLY

Defining metadata and related terminology

## What is metadata?

- "Data about data"
- Metadata is everywhere
- Example: Looking for an action film for movie night
- **Example: Looking for sources on El Salvador's Civil War**



# METADATA, BROADLY

Defining metadata and related terminology

## What is metadata?

- "Data about data"
- Metadata is everywhere
- Example: Looking for an action film for movie night
- **Example: Looking for sources on El Salvador's Civil War**



**Detalles de los Metadatos**

Título: 50 aniversario de la Insurrección Campesina Salvadoreña

UUID: 340339dc-71eb-42c9-97b1-a3aa768c418b

Fecha de creación: 1982-01-22

Descripción: las batallas de ayer, las batallas de hoy, los mismos brazos que construyen la aurora del mañana.

Identificador digital: mupi\_afiches\_m13\_0091

País de origen: El Salvador

Repositorio físico: Museo de la Palabra y la Imagen

Colección de colaborador: Colección Conflicto Armado. Afiches.

Tamaño: 38.8x58 cm

Notas: Propaganda política producida durante la Guerra Civil Salvadoreña (1980-1992), sobre grupos clandestinos y organizaciones no-gubernamentales, tanto que las fuerzas armadas., Serie MUPI, Carpeta 13

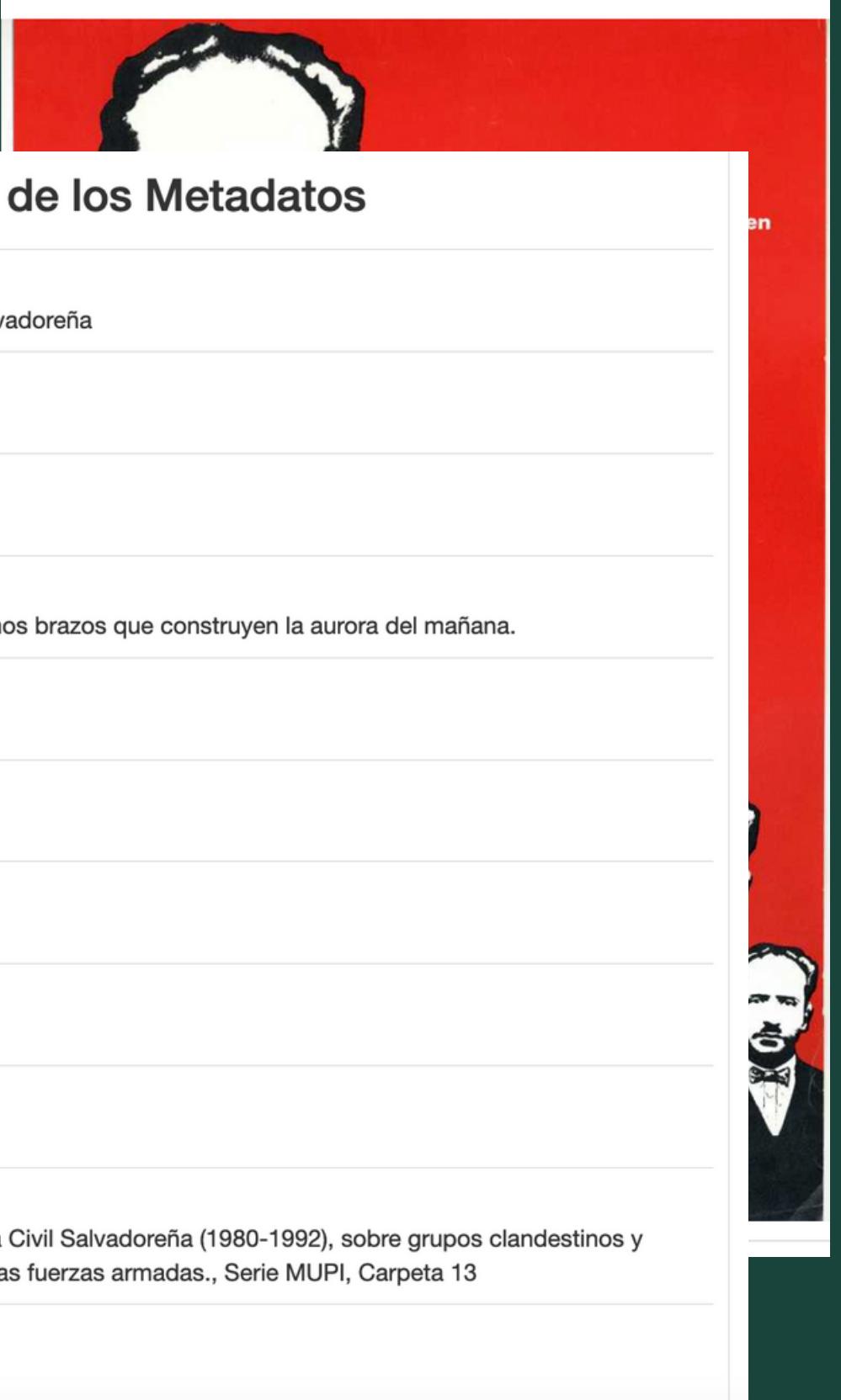
Tipo de recurso: Imágenes fijas

# METADATA, BROADLY

Defining metadata and related terminology

## What is data?

- Data is a representation of information or knowledge.
- It can be a digital or physical object, but sometimes it is more abstract or conceptual.
- **Any of the materials you are using to build a scholarly argument are considered data.**
- Source materials like interviews, artifacts from an archaeological dig, photographs, and diaries



**Detalles de los Metadatos**

Título:  
50 aniversario de la Insurrección Campesina Salvadoreña

UUID:  
340339dc-71eb-42c9-97b1-a3aa768c418b

Fecha de creación:  
1982-01-22

Descripción:  
las batallas de ayer, las batallas de hoy, los mismos brazos que construyen la aurora del mañana.

Identificador digital:  
mupi\_afiches\_m13\_0091

País de origen:  
[El Salvador](#)

Repositorio físico:  
[Museo de la Palabra y la Imagen](#)

Colección de colaborador:  
[Colección Conflicto Armado. Afiches.](#)

Tamaño:  
38.8x58 cm

Notas:  
Propaganda política producida durante la Guerra Civil Salvadoreña (1980-1992), sobre grupos clandestinos y organizaciones no-gubernamentales, tanto que las fuerzas armadas., Serie MUPI, Carpeta 13

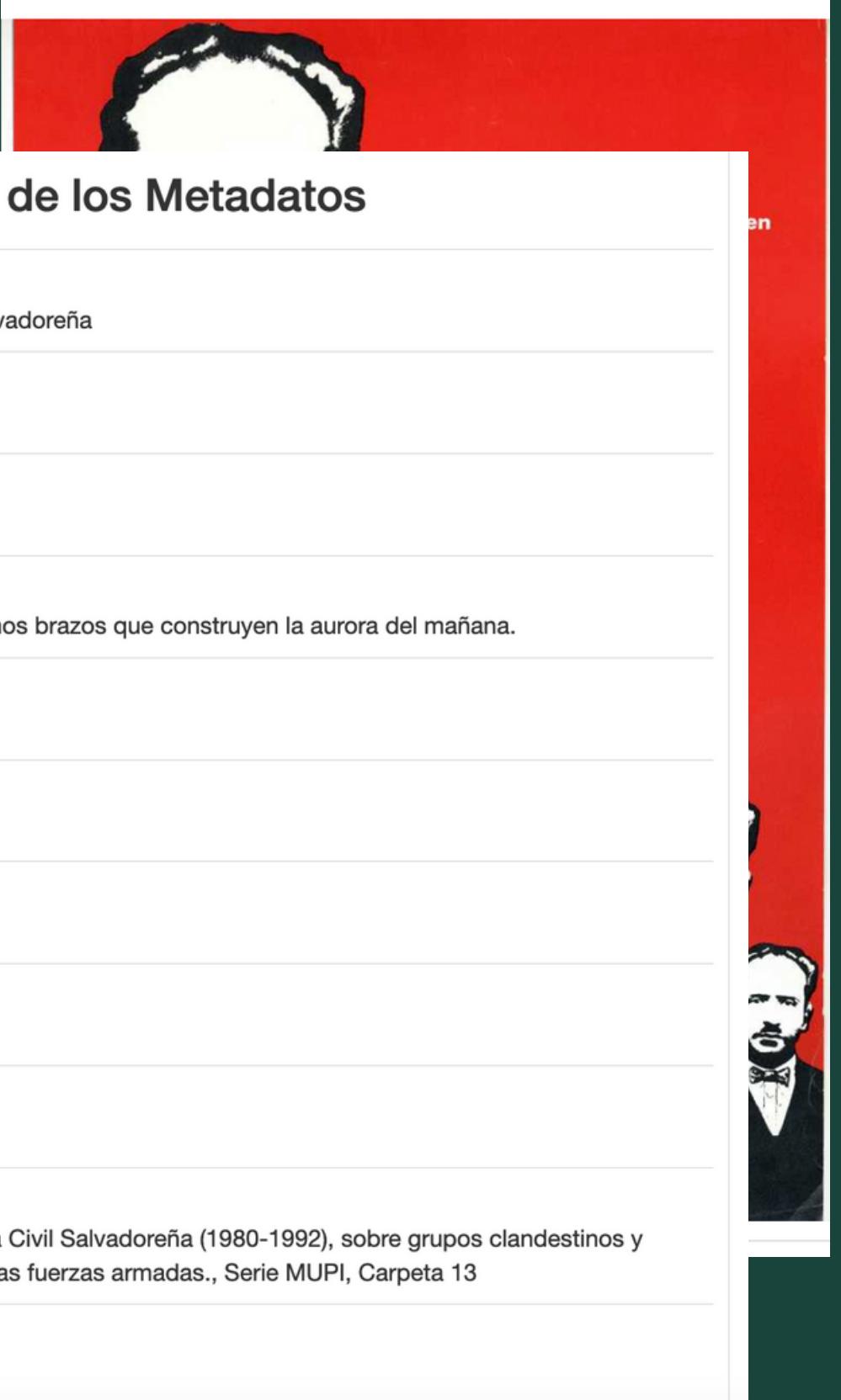
Tipo de recurso:  
[Imágenes fijas](#)

# METADATA, BROADLY

Defining metadata and related terminology

## Types of metadata

- **structural**
  - a page within a book, a document in an archival collection
- **administrative**
  - file size, copyright information
- **descriptive**
  - identifying characteristics like title
- We will focus primarily on creating and using **descriptive metadata** in this session.



**Detalles de los Metadatos**

Título: 50 aniversario de la Insurrección Campesina Salvadoreña

UUID: 340339dc-71eb-42c9-97b1-a3aa768c418b

Fecha de creación: 1982-01-22

Descripción: las batallas de ayer, las batallas de hoy, los mismos brazos que construyen la aurora del mañana.

Identificador digital: mupi\_afiches\_m13\_0091

País de origen: El Salvador

Repositorio físico: Museo de la Palabra y la Imagen

Colección de colaborador: Colección Conflicto Armado. Afiches.

Tamaño: 38.8x58 cm

Notas: Propaganda política producida durante la Guerra Civil Salvadoreña (1980-1992), sobre grupos clandestinos y organizaciones no-gubernamentales, tanto que las fuerzas armadas., Serie MUPI, Carpeta 13

Tipo de recurso: Imágenes fijas

# Activity 1: Creating descriptive metadata

---

In a moment, I will put an image of a document on the screen. (You can also find a link to this image on the document linked to this presentation at <https://bit.ly/45NRXfY>)

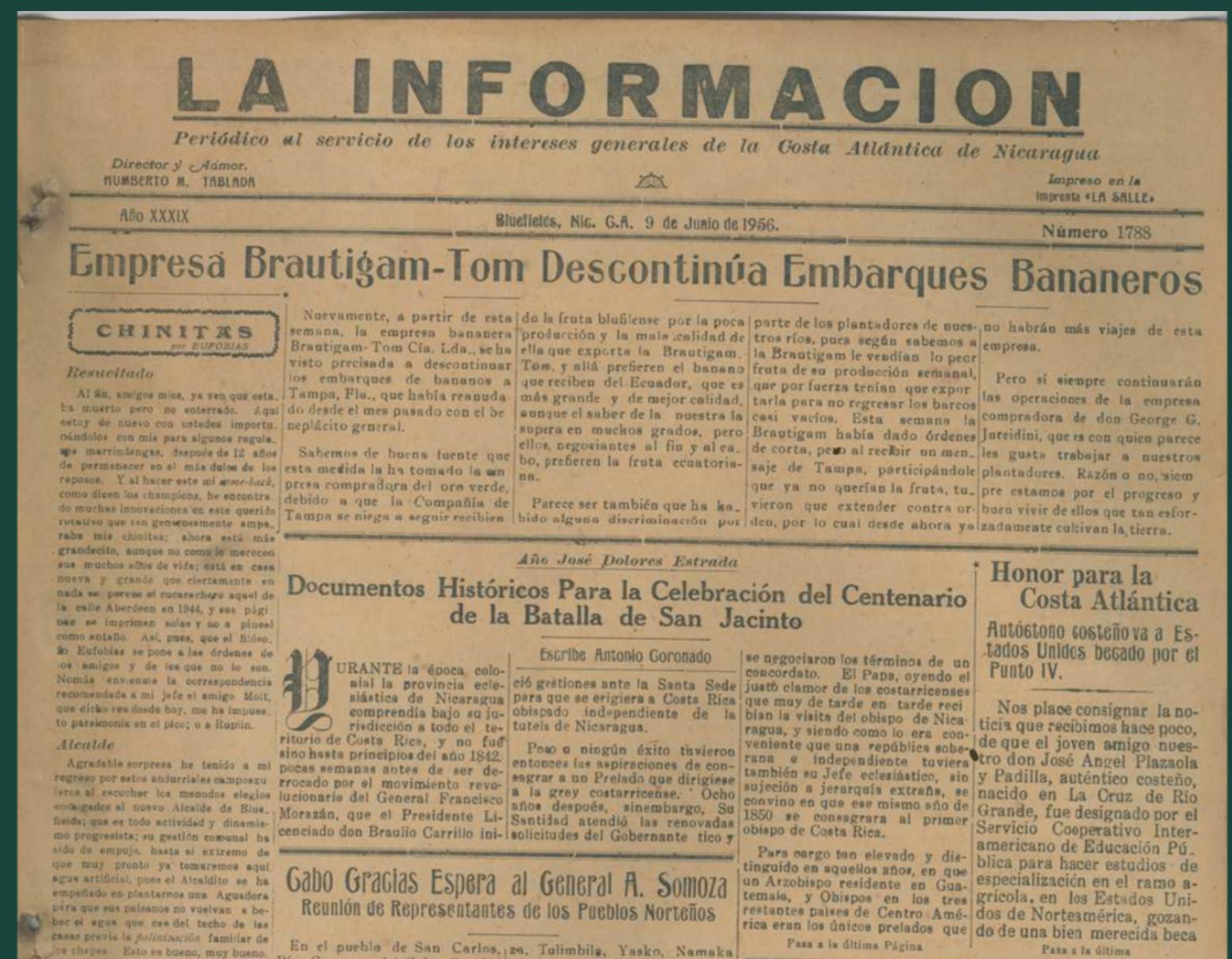
With a small group of other participants, look at the document and come up with some categories that could be used to describe it. What kind of information would you want to know about this object so that you could find it, assess it, or analyze its quality?

We'll take 5 minutes to complete this activity. You'll hear a beep when our time is up.

# Activity 1: Creating descriptive metadata

Possible categories may include:

- A title
- A creator
- Type of object
- Description
- Date
- Language
- If known, possible subjects
- If known, the location where the object is stored



# ② Metadata, Specifically

Metadata in various contexts

---

# METADATA, SPECIFICALLY

Metadata in various contexts

---

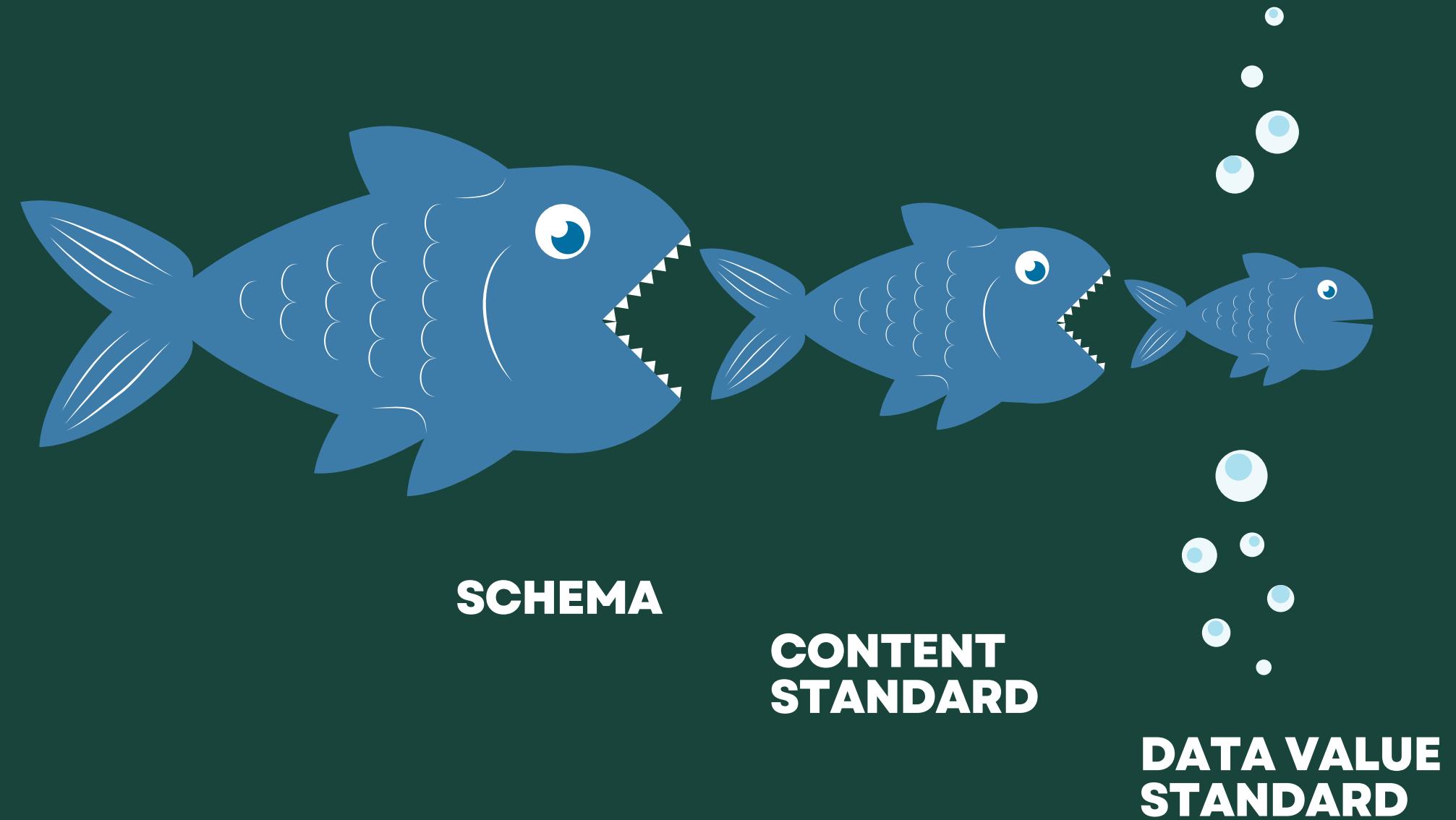
- In order to be useful, metadata must be **structured, consistent, and appropriate to the needs of users**
- Many guidelines have been created to standardize metadata and generally fall into the following categories:
  - **Schemas** provide a framework for descriptive information as metadata
  - **Content standards** provide guidelines to fill in the descriptive metadata fields from the schema in a consistent way
  - **Data value standards** such as **controlled vocabularies** function within content standards to make metadata even more consistent



# METADATA, SPECIFICALLY

Metadata in various contexts

---

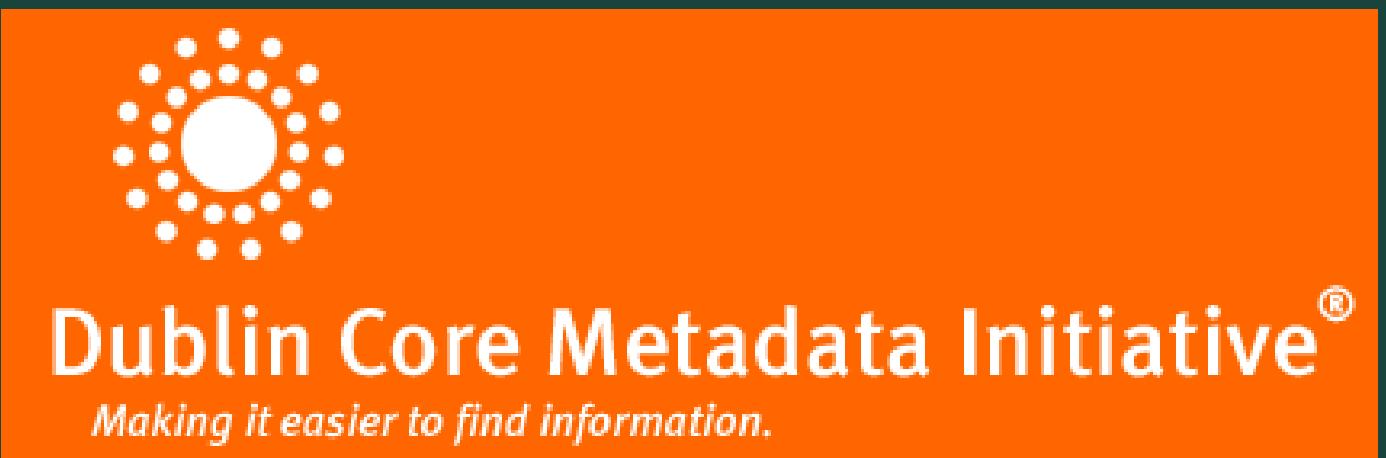


# METADATA, SPECIFICALLY

Metadata in various contexts

## Schemas: Overall structure

- **Schemas** create an order and structure to "elements," or categories required to capture relevant information in the metadata
- The most common metadata schema is called "Dublin Core" or DC
- Dublin Core consists of 15 "core" (required) "elements," but is extendable to work with many more specialized systems



INTRODUCTION TO METADATA: WHAT IS IT AND WHY DOES IT MATTER?  
<https://bit.ly/45NRXfY>

# THE DUBLIN CORE ELEMENTS

Dublin Core Element	Use
Title	A name given to the resource.
Subject	The topic of the resource.
Description	An account of the resource.
Creator	An entity primarily responsible for making the resource.
Publisher	An entity responsible for making the resource available.
Contributor	An entity responsible for making contributions to the resource.
Date	A point or period of time associated with an event in the lifecycle of the resource.

Dublin Core Element	Use
Type	The nature or genre of the resource.
Format	The file format, physical medium, or dimensions of the resource.
Identifier	An unambiguous reference to the resource within a given context.
Source	A related resource from which the described resource is derived.
Language	A language of the resource.
Relation	A related resource.
Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.
Rights	Information about rights held in and over the resource.

# METADATA, SPECIFICALLY

Metadata in various contexts

---

## Content standards: Formatting

- **Content standards** define how to record information in the metadata to meet the needs of a user in a particular context
- They may give guidelines on formatting (like “Last Name, First Name”) or specify whether **data value standards** should be used in certain fields
- DACS (“Describing Archives: A Content Standard”) is often used in archives and is tailored for users of these institutions
  - Example: DACS recommends using “undated” where the date is not known, rather than abbreviations like “n.d.”



# METADATA, SPECIFICALLY

Metadata in various contexts

---

Data value standards: Choosing one value from a list

- **Data value standards** are lists of standardized terms - they can be called thesauri, controlled vocabularies, and authorities
- They restrict the allowed input in a particular field, especially where there can be wide variability in spellings or formatting
- They make it easier to connect related/similar objects by narrowing the list of possible descriptors
- The ISO Codes for the Representation of Names of Languages are international standards that assign a unique identifier code to a language



# Activity 2: Applying metadata standards and guidelines

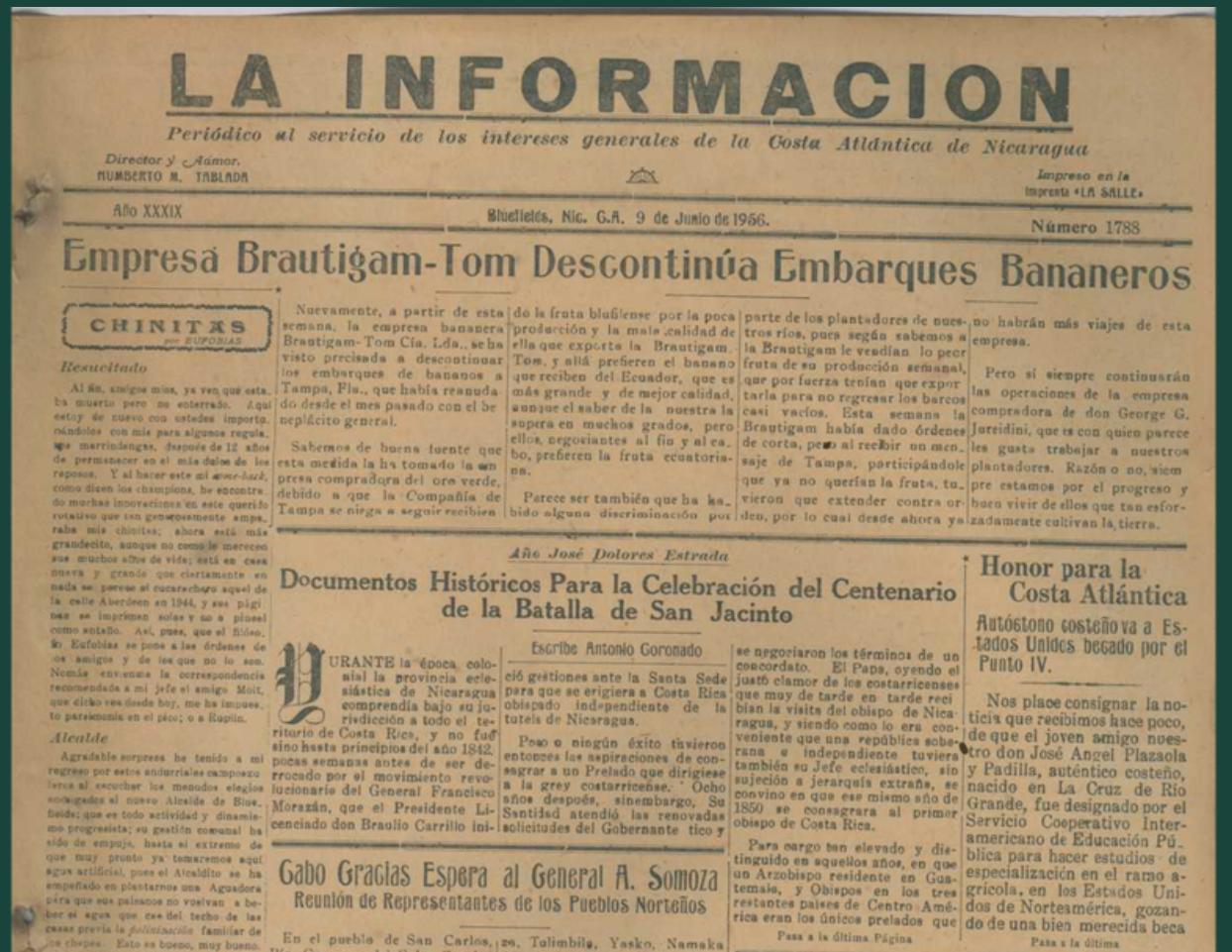
---

First, we will go through an example together. Then you'll have a chance to practice creating metadata according to specific standards in small groups.

Let's return to the document we looked at earlier during the first activity. As a reminder, you can also find a link to this image on the document linked to this presentation at <https://bit.ly/45NRXfY>. In this exercise, we will create metadata for the document according to the following standards:

- We will use the **Dublin Core** metadata schema as a framework.
- We will use the content standard **DACS** for the formatting of the Date field.
- We will use the data value standard **ISO Codes for the Representation of Names of Languages (ISO 639-3)** for the Language field.

# Activity 2: Applying metadata standards and guidelines



# Activity 2: Applying metadata standards and guidelines

---

Now, in a small group, create metadata for the documents below according to the following standards: **Dublin Core** metadata schema, **DACS** for the formatting of the Date field, and **ISO 639-3** for the Language field. You can use a copy of the spreadsheet if it is helpful. You'll have 10 minutes to complete as many fields as you can. You will hear a beep when our time is up.

**Document 1:**

A printed marriage certificate dated 1904 between Felipe Colligüin (nacido 1851) y María Cusecheo Lemunao (nacida 1879). Created by the Registro Civil in Temuko, Chile and written in Castellano and Mapuche.

**Document 2:**

An audio recording of Colombian author Héctor Abad Faciolince reading from his work entitled “Testamento involuntario” in 2015, in Spanish.



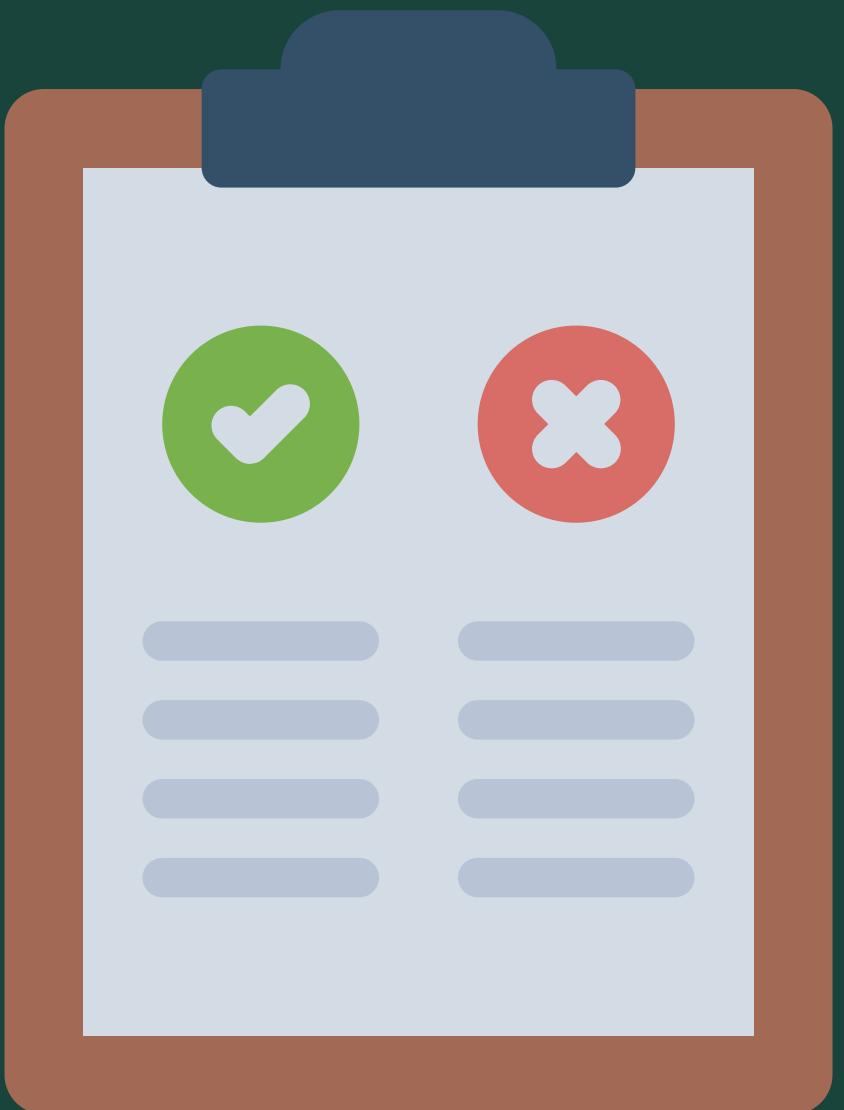
# Metadata's Opportunities & Limitations

---

## METADATA'S OPPORTUNITIES & LIMITATIONS

### Opportunities ("Pros")

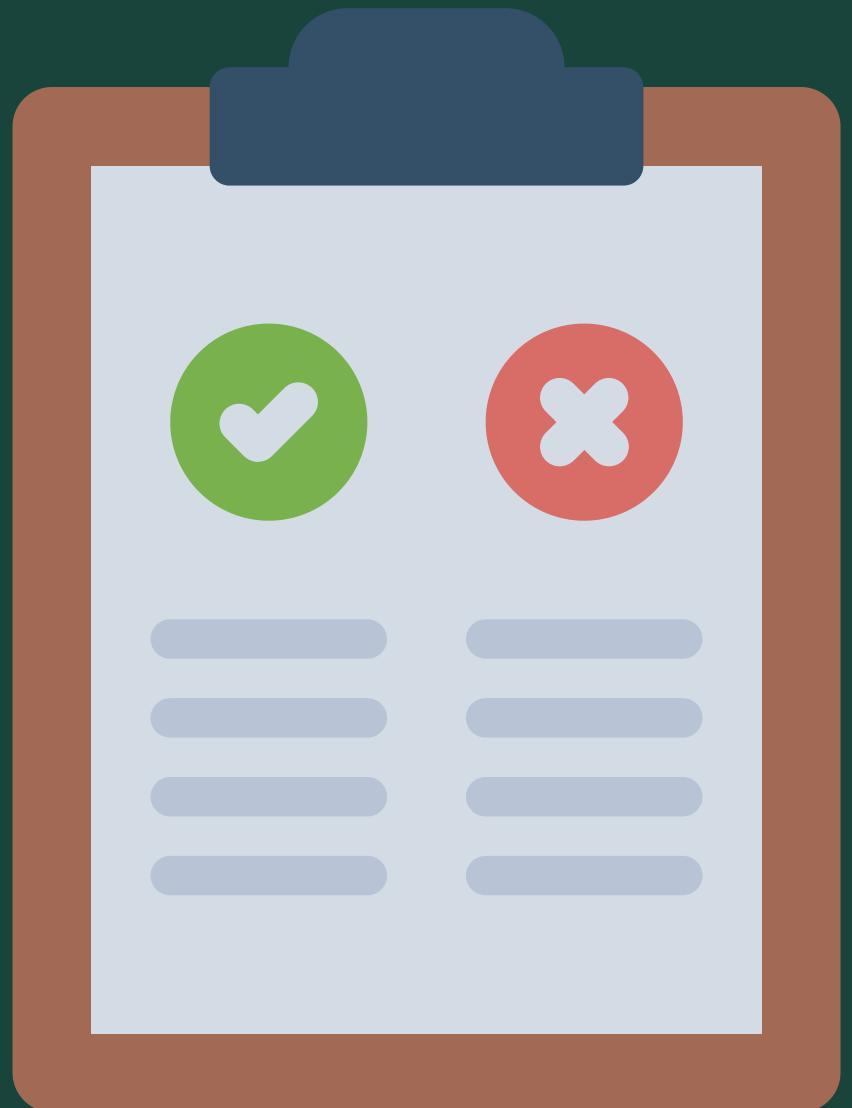
- Preserving data's contextual information makes it more usable
- Ensuring consistency between data makes it easier to find related/similar objects
- Controlling how information should be structured, for example YYYY-MM-DD as a date field, makes the data more legible to both computers and human users



# METADATA'S OPPORTUNITIES & LIMITATIONS

## Limitations ("Cons")

- Systems of classification privilege specific knowledge systems, reducing usability for users who may follow traditional epistemologies such as Indigenous frames of reference
  - Example: TK labels
- Controlled vocabularies can perpetuate the use of outdated, inaccurate, misrepresentative, or harmful language
  - Example: LCSH subject headings
- Local communities of knowledge users and/or creators may have their own preferred categories which are not present or easily incorporated into existing frameworks
  - Example: AREPR



# ④ Applying Metadata Protocols

Deciding what works best for you and your project

---

# APPLYING METADATA PROTOCOLS

Deciding what works best for you and your project

---

Source: [UC Santa Cruz](#)

## Step One

When preparing to describe your resources, there are a number of questions that you will want to consider:

- **What you are describing?**
  - Are you describing a physical object, a digital object, a digital representation of a physical object?
- **What kind of information do you want to record?**
  - Who is my audience? What information is needed to identify the resource? What information is needed to properly contextualize it? How do I want people to find it or interact with it? How do I expect them to search for or discover it? How do I expect to use it? How do I expect others to use it now and in the future? What information is required to communicate who owns it, who can use it and to what extent?

# APPLYING METADATA PROTOCOLS

Deciding what works best for you and your project

---

## Step Two

- As you begin to answer the questions presented in Step One, list out the information that you would like to include as data points, e.g., title, subject, access rights, etc. For example, if you are wanting to overlay images onto a map, you will want to record coordinate data. This is your metadata wish list.

## Step Three

- Consider the descriptive information or metadata that you may already have:
  - Which elements or what kind of information are recorded or represented there? Is information missing about your resources? Is there information that would challenging to find or create?

# APPLYING METADATA PROTOCOLS

Deciding what works best for you and your project

---

## Step Four

- Find your "golden minimum."
  - Determine what information is essential to facilitate discovery, identification, and to give sufficient context, but no more. What exactly is the golden minimum in the space of your project depends on your project goals and available resources.

## Step Five

- Finalize your list of data points. Choose to codify this list as your own metadata schema or map it to an existing schema, such as Dublin Core.

# APPLYING METADATA PROTOCOLS

Deciding what works best for you and your project

---

## Step Six

- Decide whether you want to make use of data value standards (controlled vocabularies, thesauri, encoding or formatting standards). If so, which standards would apply to which fields?
- Alternatively, you can create your own data value standards, such as, a subject vocabulary specific to your topic or collection of resources, or a controlled list of names. Document your decisions as your best practices.

# Homework

---

Apply what you've learned today!

Think about the data and source material you interact with frequently, or which you wish you could interact with more.

- Where is it held? If in an institution, have they applied their own metadata? If so, how robust is it? Is there information missing? Would you be able to find the materials if you didn't already know they existed?
- If it is not held in an institution, how can you help ensure that contextual information is recorded for the material? How can you keep that information consistent while meeting the needs of potential users?





# Where Can You Learn More?