PBCore 103:
The PBCore XML Schema

The PBCore Project at the
Carl and Ruth Shapiro Family National Center
for Accessible Media at WGBH (NCAM)
http:www.pbcore.org
http://ncam.wgbh.org/pbcore

PBCore
Public Broadcasting
Metadata Dictionary Project

Carl and Ruth Shapiro Family
NCAM
National Center for Accessible Media

What to expect
from this presentation

- In this session you will learn...
  - What an XML Schema is
  - What a namespace is and what it does
  - How the PBCore XSD can be used to move metadata from one information system to another
  - Why PBCore compliance is important

What is the PBCore project at NCAM?

- Project objectives:
  - Define, promote and advocate for the use of PBCore
  - Develop effective training materials and creating software tools to support the use of PBCore
  - Identify strategies for PBCore sustainability
- Project timeline: 3/2005-9/2007

3

What is PBCore?

- ...a metadata dictionary, used to describe many different types of media items
- ...useful for locating, sharing and exchanging media items
- ...available for <u>free</u> to anyone, not just public broadcasters

4

The PBCore Web Site

- Your one-stop shop for all things PBCore:
  http://www.pbcore.org
  - user guide
  - list of elements in various representations
    - full documentation, quick index, alphabetical, cheat sheet, etc.
  - listserv sign-up, training info
  - training schedules and materials, archived sessions
  - many other resources (metadata primer, XML schema info, etc.)

5

Part I
What's an XML Schema?

6

What's an XML Schema?

- Defines the legal building blocks of an XML document
- Ensures that applications communicate using the same language, the same grammar and the same structuring of data
- Allows machines to carry out rules made by people
- Analogous to a (very strict) recipe:
  - what ingredients are allowable, and in what order
  - ensures a consistent outcome every time

7

What's an XML Schema?

- A schema defines...
  - the sequence in which elements appear in an XML document
  - interrelationships between different elements
  - types of data used to express elements and attributes (text string, number, date, timestamp, etc.)

8

What's an XML Schema?

- A schema defines...
  - what elements can appear in a document:
    - attributes, elements
    - number of occurrences
    - the order and number of child elements
    - default or fixed values for elements and attributes
- XML Schema is a W3C recommendation (http://www.w3.org/XML/Schema), published 2001.
- An XML schema is also called an XSD (XML Schema Definition).

What's a namespace?

- Some schemas use elements with identical names but different definitions.
- A namespace is used to define and distinguish identical element names used in different schemas:
  - HTML "title" vs PBCore "title"
  - the PBCore element "contributor" indicates someone who has contributed intellectual content to a media item
  - to the general public-broadcasting world, "contributor" is used to indicate a group or person who has contributed money

What's a namespace?

- Declaring namespaces in a document avoids collisions by and confusion over identically named elements.

- By declaring namespaces, an XSD can accommodate identically named elements, each validating from rules taken from multiple specifications.
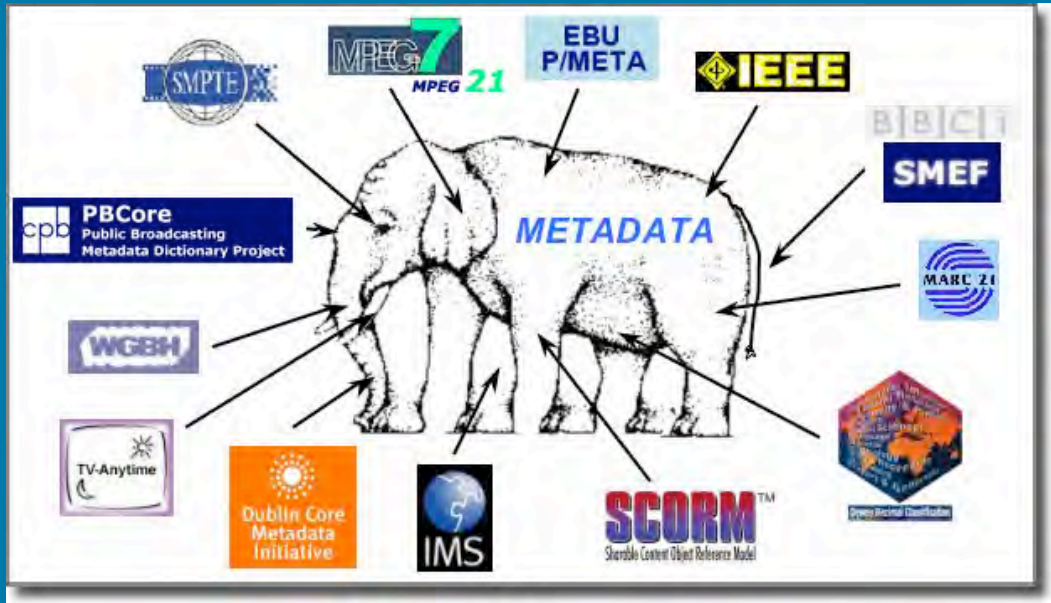
11

What's a namespace?

- PBCore contains unique elements as well as elements borrowed from the Dublin Core.

- All are formalized under the PBCore namespace.
  - Context for the PBCore Namespace
    http://www.pbcore.org/PBCore/PBCoreNamespaceContext.html
  - Access the PBCore Namespace
    http://www.pbcore.org/PBCore/PBCoreNamespace.html

12

Namespaces

13

Namespaces

14

Part II
The PBCore XML XSD

What is the PBCore XML XSD?

- PBCore's XSD defines the framework of the PBCore elements and their inter-relationships and interdependencies.
- Download the PBCore XSD at http://www.pbcore.org/PBCore/PBCoreXMLSchema.html.

```
<!-- the pbcore genre - this element may occur as many times as
        desired, however if it does occur, then the genre tag inside is
        required.  that genre tag has a controlled vocabulary. -->
<xsd:element maxOccurs="unbounded" minOccurs="0" name="pbcoreGenre">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element maxOccurs="1" minOccurs="0" name="genre"
          type="pbcore.string.type.base">
          <xsd:annotation>
            <xsd:documentation xml:lang="eng">"The descriptor genre describes the
              manner in which the intellectual content of a media item is presented,
              viewed or heard by a user. It indicates the structure of the
              presentation, as well as the topical nature of the content in a
              generalized form."</xsd:documentation>
            <xsd:documentation xml:lang="eng">"Picklist at
              http://www.utah.edu/cpbmetadata/PBCore/picklists/picklist_genre.html"</
xsd:documentation>

          </xsd:annotation>
        </xsd:element>
```

"genre" at PBCore

05.01 genre

- Describe this Element
- Show Me Examples
- Other Attributes of this Element

Describe this Element

NAME Info	genre
DEFINITION Info	The descriptor **genre** describes the manner in which the intellectual content of a media item is presented, viewed or heard by a user. It indicates the structure of the presentation, as well as the topical nature of the content in a generalized form.
ELEMENT INTERDEPENDENCIES Info	⟶ This element, along with its sibling element **genreAuthorityUsed**, is hierarchically bound to the container **pbcoreGenre**
REFINEMENTS & ENCODING SCHEMES Info	PBCore Aggregate Picklist COPY & PASTE PICKLIST View PBCore Genre List ⌄ View Tribune Sports Tags ⌄

17

"genre" at PBCore

|	
GUIDELINES FOR USAGE Info	The descriptor genre is somewhat of a catch-all container field that includes descriptions about the manner in which intellectual content for a media item is presented, as well as its structure and topical nature. For example, the genre description for a program might include: • Documentary. Environment. • Tutorial. Oceanography. • Comedy. Political satire. The industry has had many debates about the difficulties in separating the structure of the presentation against it topical nature. In practice, digital asset management systems and media guides, both paper-based and electronic, have stopped making the distinction. A compelling argument for merging the descriptors is based on the habits of end users as they enter keywords and other descriptors into search engines. Too much granularity has also confounded those who are cataloging media items. The genre tags used by PBCore are aligned with those in use by TiVo, the Tribune Media tags, and the PBS Program Offer Data Service (PODS). Select as many descriptors as is appropriate to fully capture the flavor of a media item's intellectual content and the creative form in which it is presented.
OBLIGATION TO USE Info	Recommended
REPEATABLE ELEMENT Info	Apply once within its container, **pbcoreGenre** (which itself can be applied multiple times for a media item)
TYPE OF DATA ENTRY Info	Text String

18

How is the PBCore XSD structured?

- The PBCore data model shows hierarchically how programmers might structure the database of elements:

➡ PBCore description document or master container (1)

↳ Content classes (4)

↳ Element containers/sub-containers (18)

↳ Elements (53)
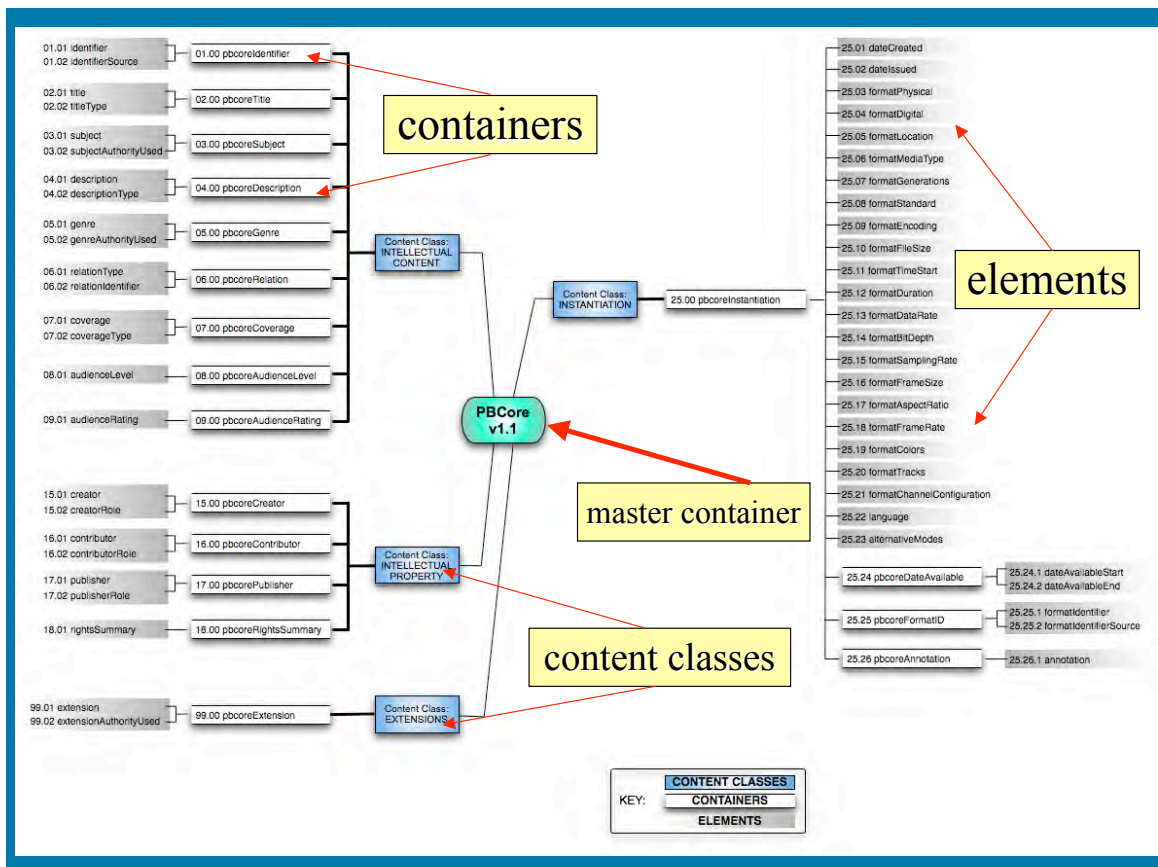
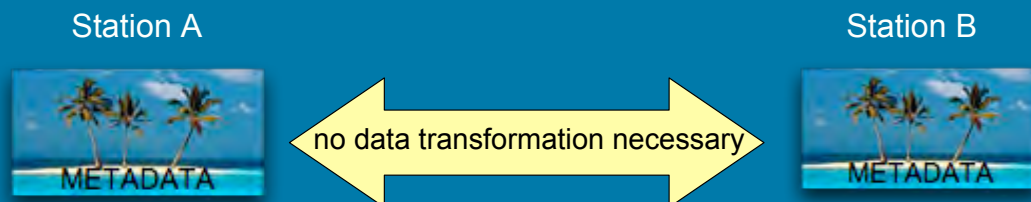# Two scenarios using PBCore: no transformation necessary

1. <u>Station A</u> uses an in-house tool to create and save records of its media assets. This tool can export the records as XML compliant with the PBCore XSD.

    <u>Station B</u> uses the PBCore Project's cataloging tool to create and save records of its media assets. This tool can also export the records as XML compliant with the PBCore XSD.

Two scenarios using PBCore: no transformation necessary

Because both stations use tools that export records compliant with the PBCore XSD, they can exchange records without first transforming them.

Station A   Station B

no data transformation necessary

METADATA   METADATA

Two scenarios using PBCore: transformation is necessary

2. <u>Station X</u> purchased a tool (System X) to create and save records of its media assets. This tool exports the records as XML (in Format X) that is not compliant with the PBCore XSD.

   <u>Station Y</u> purchased a different tool (System Y) to create and save records of its media assets. This tool exports the records as XML (in Format Y) that is not compliant with either the PBCore XSD <u>or</u> Format X.

   Because the stations use tools that export records in different formats, they cannot exchange records without first transforming them into a new format that can be understood by both tools.
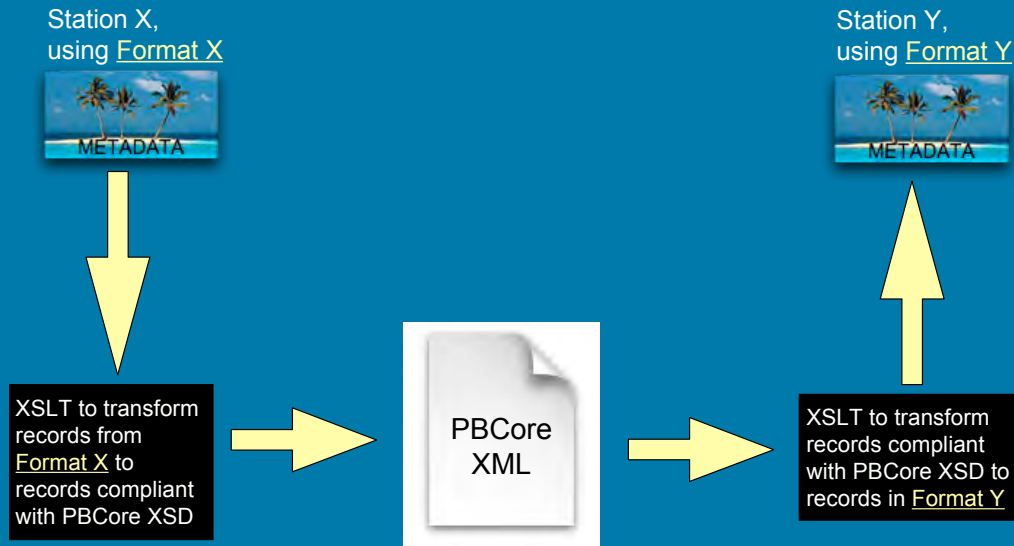
23

Two scenarios using PBCore: transformation is necessary

- The PBCore XSD can be used to move data between the two incompatible systems:

  – Station X uses a transformation method (such as XSLT, a language for transforming XML documents) to translate XML records <u>from</u> Format X into PBCore XSD-compliant data.

  – Station Y transforms the PBCore XSD-compliant data <u>into</u> Format Y before importing the records into System Y.
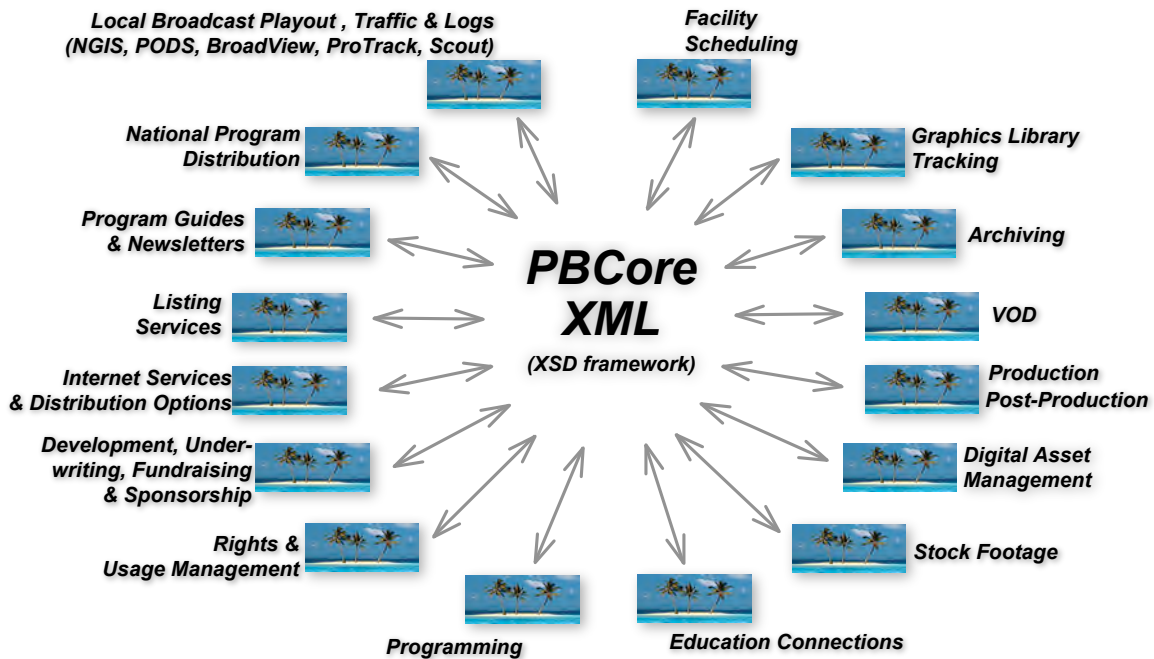
24

Two scenarios using PBCore: transformation is necessary

Station X,
using Format X

METADATA

XSLT to transform records from Format X to records compliant with PBCore XSD

PBCore XML

XSLT to transform records compliant with PBCore XSD to records in Format Y

Station Y,
using Format Y

METADATA

25

Using PBCore within a single broadcast facility

- Many stations have departments which maintain their own "data islands:" metadata that supports departmental functions and needs.

- PBCore-compliant XML documents could act as the data-transformation framework that allows these different departments/systems to share data.

26

Local Broadcast Playout , Traffic & Logs
(NGIS, PODS, BroadView, ProTrack, Scout)

Facility Scheduling

National Program Distribution

Graphics Library Tracking

Program Guides & Newsletters

Archiving

PBCore XML
(XSD framework)

Listing Services

VOD

Internet Services & Distribution Options

Production Post-Production

Development, Under-writing, Fundraising & Sponsorship

Digital Asset Management

Rights & Usage Management

Stock Footage

Programming

Education Connections

Transforming data using the PBCore XSD

- Rather than individual stations creating customized export/import transformations (e.g., with XSLT) between different systems, it makes sense to have each system communicate with one standard: the PBCore XSD.

- The soon-to-be-released PBCore cataloging tool will offer the ability to describe assets using PBCore-compliant metadata and then export that metadata according to the PBCore XSD structure.

What does it mean to be PBCore compliant?

- Follow PBCore dictionary elements and rules:
  - ensure the underlying elements have matching meanings, even if their visible labels vary
  - follow the rules for expressing data
  - apply the elements and attributes consistently
  - document variances in implementation

29

What does it mean to be PBCore compliant?

- If you ignore the XSD, then data cannot be properly exported or imported between information systems.
- If you're using PBCore to tag media items that you intend to share with the outside world, you <u>must</u> follow the dictionary rules and element obligations.
- If you're using PBCore for internal purposes only and don't plan to share your resources with the outside world, you can apply the dictionary rules as you see fit.

30

Part III
Resources

PBCore listserv

- Be part of a community of practice where PBCore adopters share information and resources.
- To **join :**
  - Go to the PBCore Web site (http://www.pbcore.org) and click on the "Join Listserv" button.

Upcoming PBCore training sessions

No.	Topic	Description	Live WebEx Time	On-Demand Availability
I	*PBCore 101: An Overview*	An introduction to the nature of metadata and the PBCore.	Wednesday 2007-01-31/2:00pm EST Repeated Thursday 2007-02-01/2:00pm EST	**PowerPoint-zip** **PDF** **QuickTime Video-zip** **QuickTime/iPod Video-zip** **MP3 audio-zip**
II	*PBCore 102: The Elements*	A more detailed look at the structure of PBCore, what each of the elements are, and how to use them to describe media objects of all kinds.	Wednesday 2007-02-28/2:00pm EST Repeated Thursday 2007-03-01/2:00pm EST	**PowerPoint-zip** **PDF** **QuickTime Video-zip** **QuickTime/iPod Video-zip** **MP3 audio-zip**
III	*The PBCore XML Schema*	A general discussion of what an XML Schema is, and a look at how the PBCore XML Schema can be used to implement the PBCore metadata dictionary within information and cataloging systems. Also explains how important the XML Schema is for sharing metadata and interoperability.	Wednesday 2007-08-01/2:00pm EST Repeated Thursday Live 2007-08-02/2:00pm EST	PowerPoint-zip PDF QuickTime Video-zip QuickTime/iPod Video-zip MP3 audio-zip
IV	*Cataloging Tools for PBCore*	A detailed discussion of tools that use PBCore, specifically, the PBCore FileMaker Pro cataloging tool we've developed. The tool facilitates the markup of descriptions for media objects and has XML data export capabilities for sharing and interoperability.	Wednesday 2007-09-19/2:00pm EST Repeated Thursday Live 2007-09-20/2:00pm EST	PowerPoint-zip PDF QuickTime Video-zip QuickTime/iPod Video-zip MP3 audio-zip

- **Session IV** will present the PBCore project's cataloging tool and illustrate its use. To apply to be a beta tester, please send a note to Marcia Brooks: marcia_brooks@wgbh.org.

Resources

General

- PBCore.org (http://www.pbcore.org/)
- The PBCore listserv (http://www.pbcore.org)
- The PBCore project at NCAM (http://ncam.wgbh.org/pbcore)
- The PBCore v1.1 XSD (http://www.pbcore.org/PBCore/PBCoreXMLSchema.html)

Resources

Metadata

- PBCore metadata resources (http://www.pbcore.org; click on "Resources and Links" button)
- "The Case for Shared Metadata Standards" (http://www.current.org/tech/tech0209metadata.html)
- A Primer for Understanding PBCore (http://www.utah.edu/cpbmetadata/PBCore/PBCorePrimer.html)

35

Resources

Metadata search tools

- Wisconsin Public TV's Media Library Online (http://wptmedialibrary.wpt.org/), a PBCore implementation through CWIS (http://scout.wisc.edu/Projects/CWIS/)
- NETA/Utah Education Network Media Exchange http://www.uen.org/media_exchange
- RELOAD (http://www.reload.ac.uk/)

36

Contact information

Marcia Brooks

Project Director, PBCore Project

marcia_brooks@wgbh.org