The TAPS Checklist as a tool for grassroots development of digital language resources

March 1, 2013

Shannon Bischoff, Indiana University – Purdue University Fort Wayne
Ivy Doak, University of North Texas
Amy Fountain, University of Arizona
John Ivens, University of Arizona
Audra Vincent, University of British Columbia, Coeur d’Alene Tribe
Overview

• Background
  – Research on Coeur d’Alene
  – The original CAOLR (2009)
  – The CAOLR to the COLRC – using TAPS (Chang 2010)

• The TAPS Checklist and the COLRC
  – Target
  – Access
  – Preservation
  – Sustainability

• Some Conclusions
Background

- Coeur d’Alene is a Southern Interior Salishan language (USA/Idaho) not widely learned by children in the home.
- During the first part of the 20th Century a great deal of linguistic and anthropological field work was conducted, resulting in a number of published works.
  - Early collection of texts (Teit 1917)
  - Later publications from this work (Lyon & Green-Wood 2007; Nicodemus et al. 2000)
- And at least 1,200 pages of unpublished field notes from Reichard, Nicodemus, Miyal, Antelope and Nicodemus.
Background

James Teit

Gladys Reichard


photo: http://anthropology.usf.edu/women/reichard/reichard.html
• By the late 1950s, Coeur d’Alene was one of the best documented of the Salish languages.
• However, published documentation was targeted at the scholarly community, and was not readily accessible.
• Unpublished field notes were not, to our knowledge, accessible to the community at all.
• In 2000, these field notes were made available to linguist Shannon Bischoff, for use on a Master’s Thesis at the University of Montana.
• Bischoff went on to produce a dissertation on Coeur d’Alene in 2007.
• In 2009, Bischoff worked with then Coeur d’Alene tribal Language Program Project Coordinator Raymond Brinkman to provide community access to all of these materials.
• The Coeur d’Alene Archive and Online Language Resources (CAOLR) is born.
• The CAOLR was built by Bischoff and a student, Musa Yassin Fort, in a single six-week period.
• Neither had any experience in web development,
• They used only publicly accessible web development training resources.
The Coeur d'Alene Online Language Resource Center

The Original CAOLR

The archive contains the Coeur d'Alene Root Dictionary (Lyon and Greene-Wood 2007) with nearly 1,400 roots and about 7,690 "word" forms, Reichard's 1939 Stem List with some 1,300 stems, a list of roughly 200 affixes, and over 1,200 pages of Coeur d'Alene texts (copies of Reichard's 1927-1929 field notes and typed manuscripts) and English translations from Reichard 1947. The archive also includes a link to Ivy Doak's grammatical sketch of Coeur d'Alene. A list of the Coeur d'Alene phonemes, a comparison of orthographies (Nicodemus, Reichard, and Salishan), and a guide to pronunciation are also included.

The Coeur d'Alene archive was created, and is maintained, by Shannon Bischoff and Musa Yassin Fort. The website was created without any prior web design/building knowledge using w3schools.com free tutorials over a period of six weeks in the summer of 2009. Following the simple HTML, Java script, and CSS lessons at the site, anyone can create a similar archive. The search mechanisms were a bit more complicated and require PHP, which can also be learned at the w3 schools website. If you would like to help or have questions, comments or suggestions, contact us at crd [dot] archive [at] gmail [dot] com.

Other websites with Coeur d'Alene material include the following: The official Coeur d'Alene website; Ivy Doak's Coeur d'Alene website; the 1917 publication of Teit's Coeur d'Alene "tales" which include different versions of a small number of the Coeur d'Alene texts in the archive; and Reichard's 1938 grammar published in The Handbook of American Indian Languages, v. 3. Reichard's 1947 An Analysis of Coeur d'Alene Indian Myths is available at the University of Michigan Digital Library General Collection.

All material in the archive is copyrighted and may not be reproduced in any form without the written permission of the archive management or Coeur d'Alene tribe.

http://academic.uprm.edu/~sbischoff/crd_archive/start1.html
The Original CAOLR

Root Dictionary

About

The root dictionary presented here was compiled by John Lyon and Rebecca Greene-Wood and contains nearly 1,400 roots and about 7690 "word" forms. The data come from Lawrence Nicodemus's Coeur d'Alene dictionary. The original work was published by UMOPL and can be found at Lyon and Greene-Wood 2007. All material is copyrighted by the Coeur d'Alene Tribe and may not be copied in any format without written permission from the Coeur d'Alene Tribe. For reasons of search ability and clarity of presentation, the organization of the dictionary has been altered from its original form.

Notes on organization of entries

Within a root header, the entries are organized beginning with the least complex and move towards more complex forms. Each entry is separated by a new line and numbered. The entries first appear in the Salishan orthography, then the Nicodemus, and finally an English translation. Nicodemus sometimes identifies the simplest forms as (stem), but not in all cases. Intransitive and simple nominalized forms directly follow than reduplicated forms.

http://academic.uprm.edu/~sbischoff/crd_archive/start2.html
The Original CAOLR

Root Dictionary

http://academic.uprm.edu/~sbischoff/crd_archive/start2.html
The Original CAOLR

**Stem List**

http://academic.uprm.edu/~sbischoff/crd_archive/stmstrt.html
The Original CAOLR

Affix List

http://academic.uprm.edu/~sbischoff/crd_archive/afstrt.html
The Original CAOLR

Affix List

http://academic.uprm.edu/~sbischoff/crd_archive/afstrt.html
## The texts:

Reichard's 1947 English translations and analysis of the following texts is available online at the [University of Michigan Digital Library General Collection](http://academic.uprm.edu/~sbischoff/crd_archive/start1.html).

<table>
<thead>
<tr>
<th>Text Title</th>
<th>Source Type</th>
<th>Language</th>
<th>PDF Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Chief Child of the root (Transformer)</td>
<td>typed manuscripts</td>
<td>English</td>
<td><a href="http://academic.uprm.edu/~sbischoff/crd_archive/start1.html">PDF</a></td>
</tr>
<tr>
<td>B. Coyote Cycle (2-21)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. (2) Origin of Indian tribes (From Parts of Monster)</td>
<td>typed manuscripts</td>
<td>English</td>
<td><a href="http://academic.uprm.edu/~sbischoff/crd_archive/start1.html">PDF</a></td>
</tr>
<tr>
<td>3. (3) Coyote overpowers Sun (Securing Sun Disk)</td>
<td>field notes</td>
<td>both</td>
<td><a href="http://academic.uprm.edu/~sbischoff/crd_archive/start1.html">PDF</a></td>
</tr>
<tr>
<td>4. (4) Coyote Steals his daughter-in-law</td>
<td>field notes</td>
<td>both</td>
<td><a href="http://academic.uprm.edu/~sbischoff/crd_archive/start1.html">PDF</a></td>
</tr>
<tr>
<td>5. (6) Little Beaver</td>
<td>field notes</td>
<td>both</td>
<td><a href="http://academic.uprm.edu/~sbischoff/crd_archive/start1.html">PDF</a></td>
</tr>
</tbody>
</table>
CAOLR to COLRC

• CAOLR contains a variety of resources, made accessible to the community for the first time.
  – It was built in a very compressed timeframe.
  – It was built with very limited knowledge of web development and web-based language archiving best practices.

• Question: Could a site like the CAOLR be brought into line with ‘best practices’ for online language resource development,
  – On a model still as consistent as possible with ‘grass-roots’ style web development?
The emerging COLRC

http://academic.uprm.edu/~sbischoff/crd_test/ (password required while in ‘test’)
The TAPS Checklist and CAOLR to COLRC

• The COLRC takes the CAOLR as a starting point;
• Attempts to bring the site into alignment with best practices
  – Bird and Simons (2003), best practices for portability of language documentation and description; and
  – Chang (2010), TAPS Checklist for responsible online archiving of digital language resources.
Scoring a site with TAPS

- Four major areas of review.
- Sixteen Numbered Items associated with these areas
  - Each has a main question
  - Most are accompanied by a number of in-depth questions probing various aspects of the item.
- Each item can be scored:
  - Yes = 3 points
  - ? = 2 points
  - No = 1 point
- Ideal Archive: All items answered ‘yes’.
- Used to identify points of strength and weakness, and as a way of guiding development.

Chang 2010
TAPS Areas

- Target
- Access
- Preservation
- Sustainability

Chang 2010
• Target refers to the ‘fit’ of the archive with regard to the data to be deposited and the needs of the identified designated communities.

Chang 2010: 82
Item 1: Mission Statement

- **TAPS Item**: Does the archive have a mission statement that reflects a commitment to the long-term preservation of digital information?
- **CAOLR**: No. No such mission statement had been developed.
- **COLRC**: Yes. A mission statement addressing these topics is a key resource for the site.
Item 1: Mission Statement

Mission Statement

The Coeur d'Alene Online Language Resource Center (COLRC) aims to provide digital Coeur d'Alene language resources to the Coeur d'Alene community and others following best practices for digital language resources (e.g. Chang 2010). The site is located at a server provided by the University of Puerto Rico at Mayaguez. The COLRC team is committed to long term stability of the website and resources (see Long Term Planning below). The site is maintained by a team of volunteers which include students, linguists, and community members who have been funded in part by NSF BSC-1160627 and NSF BSC-1160394. Comments on the site are welcome: please use the 'Contact Us' link on the left menu to provide your feedback.

Designated Community: The primary target for resources within the COLRC is the Coeur d'Alene community. In addition, linguists and the wider scholarly community will find resources in the COLRC of interest.

Ongoing Relationship: Since its inception, the COLRC team has worked closely with the Coeur d'Alene Tribal Linguist and members of the Coeur d'Alene Language Programs in the development of the COLRC. Members of the community have actively participated in the construction of the COLRC and continue to inform its content, function, and design as members of the COLRC team. The COLRC team also includes, and has worked to include, the participation of leading scholars of the Coeur d'Alene language in the development of the COLRC and will continue to do so.

Target http://academic.uprm.edu/~sbischoff/crd_test/ (password required while in ‘test’)

Item 2: Submission Criteria

- **TAPS Item:** Does the material that I want to submit fall within the scope of the archive’s collection policy in terms of content and type?

- **CAOLR:** Scope was defined, in part, by the materials in hand; in consultation with the tribal Language Program Director.

- **COLRC:** Yes. Scope is similar to CAOLR, but mission statement addresses scope for future development; and a procedure for submission of new materials.

- See in-depth questions for more detail…

*Target* Chang 2010: 83
Item 2: Submission Criteria

- In-depth questions: What is the content of material I want to deposit?
- Content includes
  - published resources, no longer held under copyright;
    - here permissions were secured from digital archives already housing these resources.
  - published resources, held under copyright;
    - here permissions were secured from copyright holders.
    - we also checked for expired copyright, and then contacted publishers, even if copyright had expired.
  - unpublished resources, collected by Reichard in collaboration with Nicodemus, Miyal, Antelope, and Nicodemus.
    - Here the concern was to ensure that these materials were not deemed sensitive by the community; and
    - permissions were secured from archives already housing these resources.
Item 2: Submission Criteria

• In-depth questions: What formats are they in?
  – Published print resources were available in .pdf form from internet archives of various kinds.
  – Unpublished print resources were in hard copy, and were scanned to .pdf by Bischoff and Yassin Fort.
  – All pdfs created by Bischoff and Yassin Fort were also saved as .png (image) files, to ensure accessibility for users not able to access pdf.
  – Sound recordings were available in .wav format, and were converted to .mp3 to improve web accessibility.

Target Chang 2010: 84
Item 2: Submission Criteria

- In-depth questions: Does my content fall within the collection policy of the archive?
- The content that was targeted was that which was identified by Bischoff in consultation with Brinkman.
- Audra Vincent, community member and scholar, joined the COLRC team and identified resources for inclusion, and provided leadership on the project.
- Linguist Ivy Doak, who has conducted fieldwork and contributed much to the available literature on Coeur d’Alene language joined the COLRC team, and also identified content for collection based on her work.
Item 2: Submission Criteria

- In-depth questions:
- Do my materials fall within the preferred submission formats of the archive?
- Where materials were not already in the preferred formats for the COLRC they were converted to those preferred formats:
  - .pdf, .png, .mp3.

Target Chang 2010: 84
Item 3: Designated Communities

- Is my desired audience (specify: ) a good match for the groups of users the archive targets (e.g., language community, academic community, etc.)?
- CAOLR: ? This resource was developed primarily for the Coeur d’Alene tribal Language Programs.
- COLRC: Yes. This resource also seeks to serve the scholarly community, the resource has a number of features designed to meet the needs of these audiences.
Item 3: Designated Communities

- In-depth questions:
- What are the desired user communities for the data I want to deposit?
- This project’s goal is to support the Couer d’Alene tribal language programs, as well as the wider scholarly community.

Target  Chang 2010: 85
Item 3: Designated Communities

- In-depth questions:
  - Do these fall within the designated communities of the archive?
  - Yes.

**Designated Community:** The primary target for resources within the COLRC is the Coeur d’Alene community. In addition, linguists and the wider scholarly community will find resources in the COLRC of interest.

**Ongoing Relationship:** Since its inception, the COLRC team has worked closely with the Coeur d’Alene Tribal Linguist and members of the Coeur d’Alene Language Programs in the development of the COLRC. Members of the community have actively participated in the construction of the COLRC and continue to inform its content, function, and design as members of the COLRC team. The COLRC team also includes, and has worked to include, the participation of leading scholars of the Coeur d’Alene language in the development of the COLRC and will continue to do so.

Target: Designated community and ongoing relationships statements from the COLRC Mission Statement.  
Chang 2010: 85
Item 3: Designated Communities

- **In-depth questions:**
  - Particularly if the language community is an important user of the deposited materials, does the archive cater to that user community?
- **The site attempts to do so by:**
  - Ensuring that all resources are in readily accessible formats; and
  - Providing clear, stable, and useful navigation; and
  - Allowing users to view resources in relevant contexts. This is evident in the ‘texts’ area; and
  - Providing a robust and flexible ‘search’ feature on the site.

Target Chang 2010: 85
Item 3: Designated Communities

Texts

Please note: the 'search' functionality on this site is currently under revision, and so is not yet available for use.

The unpublished field notes and typed manuscripts of Coeur d'Alene myths and tales presented in this archive were recorded in 1927 and 1929 by Gladys Reichard. The texts cover what Reichard classified into "myths and tales" and "tales with historical elements". The "myths and tales" are further divided into "Coyote cycle" and "myths not in Coyote cycle."

The numbering of each text follows numbering of Reichard 1947 with the numbering of the typed manuscripts in parentheses "( )" directly preceding the title, the numbers corresponding to the field notes can be found in the field note files. In addition, the titles reflect Reichard 1947 and may differ from titles on field notes and typed manuscripts. Complete English translations of each text can also be found and come from An Analysis of Coeur d'Alene Indian Myths (see Reichard 1947 in the left menu).

Analysis of some texts along with other excellent Coeur d'Alene resources can be found at Ivy Doak's Coeur d'Alene website. The 1917 publication of Teit's Coeur d'Alene "tales" which include different versions of a small number of the Coeur d'Alene texts in the archive can also be found online (see Teit 1917 on the left menu).

Target  
http://academic.uprm.edu/~sbischoff/crd_test/  (password required while in ‘test’)
Item 3: Designated Communities

- Texts are presented in a variety of formats;
  - audio files (where available)
  - images or pdf
  - hand-written or typed field notes.
  - or both.

- ‘both’ allows users to see the mark-up used in the conversion of raw texts to published versions.

Target: http://academic.uprm.edu/~sbischoff/crd_test/ (password required while in ‘test’)
Item 3: Designated Communities

Target: http://academic.uprm.edu/~sbischoff/crd_test/ (password required while in ‘test’)

[Handwritten text image]

‘both’
Item 3: Designated Communities

- Resources on the site presented in up to three different orthographies, to serve multiple communities of users:
  - ‘Reichard’; the Boasian transcription used by Gladys Reichard in the original works;
  - ‘Salishan’; used by Salishan linguists today;
  - ‘Nicodemus’; developed by Lawrence Nicodemus for use by the Coeur d’Alene community, and the orthography most often used today within that community.
Item 3: Designated Communities

Coeur d’Alene Online Language Resource Center

In the archive
- about
- spelling and pronunciation guide
- root dictionary
- stem list
- affix list
- texts
- audio files
- methodological notes
- bibliography
- advanced search
- source files
- contact us

On the web

Spelling and Pronouncing Coeur d’Alene

Please note: the ‘search’ functionality on this site is currently under revision, and so is not yet available for use.

Coeur d’Alene has been spelled using at least three different systems – the “Reichard Orthography”, the “Nicodemus Orthography” and the “Salish Orthography”. The Reichard Orthography was used by many linguists and anthropologists in the first half of the 20th century, and the works of Gladys Reichard use this system. Because it was developed as a transcription system for scholars, it is not necessarily easy for non-linguists to use. The Nicodemus Orthography was developed for use by the Coeur d’Alene community. The Salish Orthography is used by many linguists and anthropologists today to transcribe Coeur d’Alene. You will find it used in contemporary scholarly work about the language. This site is intended to support users who want to use any of these systems.

The list below shows how the letters used in each of the three spelling systems (the “Reichard Orthography”, the “Nicodemus Orthography” and the “Salish Orthography”) that have been used to write Coeur d’Alene correspond to each other. Letters not listed on this table are used in all three systems as they would typically be used in English.

<table>
<thead>
<tr>
<th>Reichard</th>
<th>Nicodemus</th>
<th>Salish</th>
</tr>
</thead>
<tbody>
<tr>
<td>t’, t’w</td>
<td>(w, ’w</td>
<td>5’, 5’</td>
</tr>
</tbody>
</table>

Target

http://academic.uprm.edu/~sbischoff/crd_test/ (password required while in ‘test’)
Item 3: Designated Communities

Salishan orthography

Nicodemus orthography

Target: http://academic.uprm.edu/~sbischoff/crd_test/ (password required while in ‘test’)
Item 3: Designated Communities

Target: http://academic.uprm.edu/~sbischoff/crd_test/ (password required while in ‘test’)

http://academic.uprm.edu/~sbischoff/crd_test/
## Item 3: Designated Communities

<table>
<thead>
<tr>
<th>Reichard</th>
<th>Doak (root)</th>
<th>Salishan</th>
<th>Nicodemus</th>
<th>comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>atsqā̱́</td>
<td>?acqe?</td>
<td>acqa?</td>
<td>atsqa'</td>
<td>go out, singular and plural</td>
</tr>
<tr>
<td>ats'x</td>
<td>?ac'x</td>
<td>ac'x</td>
<td>ats'qh</td>
<td>look at</td>
</tr>
<tr>
<td>ayxʷ</td>
<td>?ayxʷ</td>
<td>ayxʷ(-t)</td>
<td>ayqhw(-t)</td>
<td>be tired</td>
</tr>
<tr>
<td>axil</td>
<td>?axil</td>
<td>axil</td>
<td>aqhil</td>
<td>do thus</td>
</tr>
<tr>
<td>arʷ</td>
<td>?aⁿʷ</td>
<td>aⁿʷ</td>
<td>a(w)</td>
<td>do thus, be much, many</td>
</tr>
</tbody>
</table>

[Target](http://academic.uprm.edu/~sbischoff/crd_test/) (stem list with comments)
### Item 3: Designated Communities

**Salishan**

<table>
<thead>
<tr>
<th>Directional Prefixes</th>
<th>Nicodemus</th>
</tr>
</thead>
<tbody>
<tr>
<td>ci-</td>
<td>tsi-</td>
</tr>
<tr>
<td>či-</td>
<td>chits-</td>
</tr>
<tr>
<td>te-</td>
<td>te-</td>
</tr>
<tr>
<td>tkʷʷi-</td>
<td>tkweE-</td>
</tr>
<tr>
<td>tcp-</td>
<td>tep-</td>
</tr>
<tr>
<td>tus-</td>
<td>tus-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Locative Prefixes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>cen-</td>
<td>tsen-</td>
</tr>
<tr>
<td>č-</td>
<td>ch-</td>
</tr>
</tbody>
</table>

Links to original publication (Reichard 1938)

**Target**

[http://academic.uprm.edu/~sbischoff/crd_test/](http://academic.uprm.edu/~sbischoff/crd_test/) (password required while in ‘test’)

---

**Coeur d'Alene Online Language Resource Center**

<table>
<thead>
<tr>
<th>browse</th>
<th>Affix List</th>
</tr>
</thead>
<tbody>
<tr>
<td>directional prefixes</td>
<td>Nicodemus</td>
</tr>
<tr>
<td>locative prefixes</td>
<td>Salish</td>
</tr>
<tr>
<td>lexical prefixes - lexical suffixes</td>
<td>English Gloses</td>
</tr>
</tbody>
</table>

[Links to original publication](http://academic.uprm.edu/~sbischoff/crd_test/) (password required while in ‘test’)

---

**Links to original publication**

(Reichard 1938)
Item 4: Ongoing Relationship

• Does the archive accept the responsibility to interface with the language community as a provider community?
  – (This could involve revenue sharing and interaction with the language community as owners of their own language development efforts.)

• CAOLR: ? . This is not addressed.

• COLRC: Yes. This is outlined in the mission statement.
Designated Community: The primary target for resources within the COLRC is the Coeur d'Alene community. In addition, linguists and the wider scholarly community will find resources in the COLRC of interest.

Ongoing Relationship: Since its inception, the COLRC team has worked closely with the Coeur d'Alene Tribal Linguist and members of the Coeur d'Alene Language Programs in the development of the COLRC. Members of the community have actively participated in the construction of the COLRC and continue to inform its content, function, and design as members of the COLRC team. The COLRC team also includes, and has worked to include, the participation of leading scholars of the Coeur d'Alene language in the development of the COLRC and will continue to do so.

Designated community and ongoing relationships statements from the COLRC Mission Statement.
Item 4: Ongoing Relationship

- In-depth questions:
- What types of interaction do I anticipate needing to take place between the language community and the archive?
- Ongoing discussions about any possible sensitivity of materials;
- Ongoing discussions about accessibility of materials.

Target Chang 2010: 87
Item 4: Ongoing Relationship

• In-depth questions:
• Will the archive support these?
• The Mission Statement outlines a method for addressing these issues;
• The ‘contact us’ link has been improved in the COLRC from the CAOLR ‘contact’ feature.

Submission To contribute to the COLRC please use the 'Contact Us' link on the left. We will consider all requests related directly to the Coeur d'Alene language and culture. At this time we are unable to consider any other resources.

Submissions statement from the COLRC Mission Statement.
Item 4: Ongoing Relationship

• In-depth questions:
• Is potential revenue sharing an issue for my deposit?
• No. The archive is not predicted to generate revenue.
• Access refers to the accessibility and usage of the data and corresponding metadata once materials are deposited.
Item 5: Discoverability

• Are the metadata for materials deposited at the archive searchable online?
  – I.e. posted on the web or aggregated through participation in a service such as OLAC so that they are discoverable through Internet search engines (e.g., Google, Yahoo!, Bing, etc.)?

• CAOLR: No.

• COLRC: Yes. Metadata for all resources, following the Dublin Core standards, is available on the site.
Item 5: Discoverability

Discoverability statement from the COLRC Mission Statement.

Access
Item 5: Discoverability

- In-depth questions:
- Does the archive have the necessary guidelines and standards to help me, as the depositor, provide quality descriptive metadata?
- Yes, the Dublin Core data elements can be accessed in a variety of ways; existing metadata on the site is clearly identified and can be used as a model.

Access Chang 2010: 88
Item 5: Discoverability

• In-depth questions:
  • Is the metadata posted on the Internet?
  • Yes.
Item 5: Discoverability

A. Chief Child of the root (Transformer)
1. (1) Chief Child of the root (Transformer)
   Typed Manuscript: image files or PDF | metadata
   Published English Translation: image files or PDF

B. Coyote Cycle (2–21)
2. (2) Origin of Indian tribes (From Parts of Monster)
   Typed Manuscript: image files or PDF | metadata
   Published English Translation: image files or PDF

3. (3) Coyote overpowers Sun
   Hand-written Field Notes and Typed Manuscript: both
   Hand-written Field Notes: image files or PDF | metadata
   Typed Manuscript: image files or PDF | metadata
   Published English Translation: image files or PDF

**Access**
http://academic.uprm.edu/~sbischoff/crd_test/ (password required while in ‘test’)
Item 5: Discoverability

• In-depth questions:
• If so, are these records easily found on the Internet?
• Yes. Metadata records are available from the top menu of each area, and from the index area of the texts pages on the site.
Item 5: Discoverability

- In-depth questions:
  - Is the metadata aggregated through a service such as OLAC?
  - Not yet, but this is within the scope of COLRC development.
Item 5: Discoverability

• In-depth questions:
  • Does my desired designated community have adequate resource discovery opportunities through the archive’s approach to descriptive metadata?
  • Yes. Metadata records are complete and consistent.
Item 6: Fixed Identifiers

• Does the archive assign a persistent identifier to each item among its digital holdings so that it can be referenced and located in perpetuity?
  
• CAOLR: No.
• COLRC: Yes.

Fixed Identifiers statement from the COLRC Mission Statement.
Item 7: Reach

- Will the audience that I wish to reach (specify:) be able to access the materials once they are deposited in the archive?
  - CAOLR: Yes.
  - COLRC: Yes.

Reach statement from the COLRC Mission Statement.

Access
Item 7: Reach

- In-depth questions:
- Will members of the designated communities be expected to have access to the Internet?
- Yes, at least via the tribal Language Programs offices.
- The tribe has its own internet called Red-Spectrum, which is available for a monthly fee.
- All files can also be provided on stable digital media (such as a thumb drive) to community members lacking internet connectivity.*

* Note: site navigation and search mechanisms utilize PHP, which must be executed on a server, however.
Item 7: Reach

- In-depth questions:
  - Will members of the designated communities need to maintain an e-mail address?
  - No.
Item 7: Reach

• In-depth questions:
  • Will the metadata be available in English only, or will it be available in another language that is more accessible to the designated community?
  • The designated communities are primarily English speaking (they may be multilingual).
Item 7: Reach

- In-depth questions:
  - Will the archive charge fees for copies of data on media that are usable by members of the designated communities?
  - No.
Item 8: Access and Use Restriction

• Does the archive have policies and procedures to ensure that any restrictions I or the provider community place on access to the materials will be honored?
• CAOLR: ? No such policies were published on the site.
• COLRC: Yes.

Access and Use Restrictions statement from the COLRC Mission Statement. 
Chang 2010: 92
Item 8: Access and Use Restriction

- In-depth questions:
  - How does the archive deal with copyright?
  - Does the archive require transfer of ownership?

*Copyright and Transfer of Ownership* All materials in the COLRC are in the public domain or used by permission. All the material in the COLRC is available via a number of other resources either for a fee or for free (e.g., libraries, traditional archives, the Internet Archive). Ownership of items in the COLRC remains with the owner of the physical documents or physical resources (e.g., recordings) deposited in the COLRC. See metadata for individual items regarding copyright, information regarding ownership, and permanent locator information (e.g., ISBN, DOI, physical archive locations, or permanent URLs). Use of resources in the COLRC is restricted to the copyright of individual resources.

Copyright and Transfer of Ownership statement from the COLRC Mission Statement.
Item 8: Access and Use Restriction

- In-depth questions:
- Does the archive allow materials to be deposited with restrictions on access?
- If so, what restrictions are possible and how are requests for access handled?
- Does the archive allow materials that are closed to access to be deposited?
- How long will periods of closed access be?

Access Restrictions statement from the COLRC Mission Statement.

Access

Chang 2010: 92
Item 8: Access and Use Restriction

• In-depth questions:
  • What are the archive’s conditions of use policies?

Conditions of Use statement from the COLRC Mission Statement.

Access

Chang 2010: 92
Preservation

Preservation refers to the overall system and technical structures of the archive that ensure materials will be managed in ways that make them available and usable, with their authenticity and integrity intact, far into the future.

Chang 2010: 96
Item 9: Evidence of Long-Term Planning

• Does the archive adhere to written policies and procedures for the long-term preservation of digital materials
  – (e.g., the archive has written standards for implementation and is engaged in formal, periodic review and assessment that responds to technological developments and evolving requirements)?

• CAOLR: ?

• COLRC: Yes.
Long Term Planning The COLRC team endeavor to ensure the longevity of the COLRC as an online resource and the long-term maintenance of the digital resources accessible via the COLRC. The COLRC team is committed to yearly evaluations of the COLRC with regards to technological developments, evaluation of changes in the needs and concerns of the Coeur d’Alene community, and necessary changes to the COLRC and its management. To this end, each June until 2018 the COLRC will be evaluated in terms of its resources, community needs, and management by the current COLRC team. In June of 2018 the Mission Statement and Preservation strategies will be re-evaluated and updated as needed. In addition, at that time a committee will be established composed of community members, scholars, and others to manage the COLRC for the longer term. It is also the goal of the COLRC team to turn management of the COLRC to the Coeur d’Alene Language Programs in 2018 or shortly thereafter.

Preservation statement from the COLRC Mission Statement.
Item 10: Preservation Strategies

- Will the archive refresh and update digital materials as needed to counter obsolescence of hardware and software over time?
  - CAOLR: ?
  - COLRC: Yes.
Item 10: Preservation Strategies

Preservation Strategies In June 2012 all web pages developed in 2009 were updated to HTML4. Obsolescent formatting (e.g. frames) were replaced with appropriate, and standard, formats. This process of updating hardware, software, and markup will continue as needed. The annual June evaluation of the COLRC will include recommendations regarding the refreshing of data and data migration. All original files will be preserved in the format that they were first deposited into the COLRC to demonstrate integrity and authenticity. In addition, the materials accessible via the COLRC will continue to be backed up each June as new storage technology is developed and older storage technology (e.g. DVD, USB) becomes obsolete. Resources and data will be updated, altered, or otherwise changed for preservation on a case-by-case basis.

Preservation strategies statement from the COLRC Mission Statement.
**Item 11: Integrity**

- Does the archive use fixity\(-dixity\)* metadata to ensure that copies of digital materials will be complete and unchanged – (e.g., a checksum, or digital signature, etc.)?
- CAOLR: ?
- COLRC: ? This is feasible, but would require additional development on the COLRC site.

*‘fixity metadata’ sounded both silly and daunting to us. -dixity was added by Bischoff, and we believe it de-dauntifies the expression.*
Item 12: Authenticity

- Does the archive ensure that digital materials contain what they claim to contain
  - (e.g., by verifying that digital objects are what the metadata say they are, by permanently associating adequate metadata, and by faithfully maintaining provenance metadata to document any changes to the digital objects that occur while they are in the care of the archive)?
- CAOLR: ?
- COLRC: Yes.

Preservation  Chang 2010: 102
Item 12: Authenticity

The COLRC team strive to ensure metadata match the resources in the COLRC by regularly (each June) evaluating metadata and the match with COLRC resources. In addition, by documenting origins and all changes to materials and metadata, the depositors are ultimately responsible for vouching for the authenticity of the resources at the COLRC.

Authenticity statement from the COLRC Mission Statement.
Sustainability

• Sustainability refers to the demonstrated organizational robustness of the archive, lending long-range viability to the functions that it performs.

Chang 2010: 104
Item 13: Adequate Infrastructure

- Does the archive appear to be adequately staffed (in terms of numbers of staff and skill sets of the staff) and have the technical infrastructure to ensure continuing maintenance and security of materials—(e.g., quality media, environmentally-controlled storage, access-controlled storage area)?
  - CAOLR: ?
  - COLRC: ? There remain long-term infrastructure questions with the COLRC.
Infrastructure statement from the COLRC Mission Statement.

Sustainability
Item 14: Financial Sustainability

- Does the archive appear to have secured sources of long-term funding?
- CAOLR: No. No funding was secured.
- COLRC: ? Current funding is short-term.
Item 14: Financial Sustainability

Financial Sustainability statement from the COLRC Mission Statement.

Sustainability
Item 15: Disaster Preparedness

• Is the archive engaged in responsible backup practices and prepared to recover its digital holdings in case of disaster (e.g., disaster recovery plan, offsite storage of backups)?

• CAOLR: ?

• COLRC: Yes.
Disaster Preparedness statement from the COLRC Mission Statement.

Sustainability
Item 16: Succession Plan

- Does the archive have a reasonable succession plan to ensure that materials will be accessible and preserved elsewhere if the archive ceases to exist?
  - CAOLR: No.
  - COLRC: ? There is a short-term succession plan.

Sustainability Chang 2010: 107
Item 16: Succession Plan

Succession Plan statement from the COLRC Mission Statement.

Sustainability
TAPS and the CAOLR

• Overall Score:
  – CAOLR: 28/48
  – COLRC: 44/48
• TAPS items on which COLRC Scores 3:
  – all ‘target’, and ‘access’ items (1 through 8)
  – in ‘preservation’, items 9, 10, 12.
  – in ‘sustainability’, items 13, 14, 16.
• TAPS items on which COLRC Scores 2
  – 11, Integrity (fixity metadata is needed);
  – 13, Infrastructure (long term infrastructure is needed);
  – 14, Funding (long term funding is needed);
  – 16, Succession plan (long term succession plan is needed).
• There are no TAPS items on which COLRC scores 1.
The TAPS Checklist has driven the development of the COLRC, and pointed out key areas of need; particularly in ‘target’ and ‘access’.

However, the TAPS Checklist highlights several endemic challenges to responsible web-based language resource development.

– Long-term funding and infrastructure concerns are particularly challenging.
Lessons

• The CAOLR was developed on a very short, intensive timeline (6 weeks).
• In developing the CAOLR into the COLRC, we are able to address key shortcomings of the original data plan; and
• We are able to isolate key concerns to address in future development.
The Team

- Audra Vincent, University of British Columbia, Coeur d’Alene Tribe
- Shannon Bischoff, Indiana University – Purdue University Fort Wayne
- Ivy Doak, University of North Texas
- Amy Fountain, University of Arizona
- John Ivens, University of Arizona
Acknowledgements

• Project supported by the National Science Foundation awards BCS-1160627 and BCS-1160394
• Thanks also to Philipp Brandenburg, Raymond Brinkman, Debbie Cole, Colleen Fitzgerald, Musa Yasin Fort, Terry Langendoen, Susan Penfield, Gary Simons, Ken Rehg, Nick Thieberger and Mizuki Miyashita for useful discussion and development of the ideas presented in this project.
• Thanks to undergraduate students Christopher Roberts, Meagan Messina, Anna Loukianova, and Kayla Mayers.
• All errors are the responsibility of the COLRC team.
References


References


