Archiving Nuuchahnulth language data[*]
Adam Werle
2011–2015, for Huu-ay-aht First Nations

1. Introduction

This document is for adults who are learning or recording the Nuuchahnulth language. It introduces some of the standards and software that are available for preserving language data, and walks the reader step by step through how to use them. It is intended to serve both as a general introduction to these concepts, and as a guide that users can refer back to as needed.

This guide is written especially for advanced high school and university-level courses on Nuuchahnulth language and research. For such courses, the *tips* in the text can also serve as academic exercises, either at home or in the classroom.

In order to make the guide useful to users of diverse backgrounds, it is written as much as possible in an accessible and non-technical style. We assume, for example, that the reader might have little or no background in information technology or linguistics. Where possible, we also discuss computing with reference both to Windows and Macintosh machines, noting important differences between them.

However, we do assume a certain amount of background on the part of the reader. For example, we assume that you know how to use a word processor, and are familiar with the basics of the Nuuchahnulth language, such as the alphabet, pronunciation, and some common words and phrases. Last, we assume that you know how to download and install software, or have someone in your household who does.

Regardless of their technical background, many language researchers will have little time to read this guide. For these users, this guide begins with a small number of quick tips on which formats and software to use (see Section 2.1).

This is not a guide to equipment. For suggestions on selecting audio equipment, try starting at the FirstVoices page *www.firstvoices.com/Audio_Recording_Buyers_Guide.htm*.

This guide was last revised during summer 2014, and all Internet links given here were still good at that time. If you find any links or other information which is out of date, please contact Adam Werle (*werle–at–uvic.ca*).

2.   Archival standards

2.1. Overview

To *archive* information means to preserve it for future use. Usually, we do not know who might eventually use our data, or how far in the future it might be needed. Nevertheless, there are steps we can take to make our data as easy to use as possible, for as long as possible. In Section 2, we will discuss some of these steps. But if you are pressed for time, here is a quick list of which formats to use:

(1)   **Summary: recommended data formats**

Data type	Archival formats	Sharing formats
text	XML (letters in Unicode)	PDF, TXT, HTML
audio	WAV (44,100 Hertz, mono)	MP3
images	TIFF, PNG, JPEG	PNG, JPEG
video	AVI, MPEG-2	MP4, FLV

We can illustrate the importance of good archiving with some examples. Since we are most concerned here with language data, think about audio (sound) recordings. Many important Nuuchahnulth recordings are on old reel-to-reel tapes. But reel-to-reels were replaced by cassette tapes in the 1970s, then CDs in the 1980s. Today, you can listen to reel-to-reel recordings only if you have an antique reel-to-reel player.

Now consider *word processing,* which refers to using computers to create printed words. Although word processing technology has been available since the 1970s, anything written with word such technology before about 1990—even if it was saved on a disk— is now inaccessible and useless.

In 2000, floppy disks were still in common use, and were standardly included in desktop computers. Today, this technology has been replaced by CDs, DVDs, and flash memory, and it is very difficult to find a computer with a floppy drive. If you have any old data on floppy disks, it is now almost inaccessible.

In the rest of Section 2, we will discuss and recommend several standard file formats both for archiving, and for sharing immediately.

2.2. Basic archiving principles

Although we cannot predict how people will store their data in the future, archival experts have learned some lessons about how to preserve information for future use. To begin with, good archiving depends on *standards*. A standard is a way of doing something that a significant number of people agree on.

Several standards are important for our purposes. One is the Nuuchahnulth alphabet, which is the standard way to write the language. In addition, when we record electronic information about Nuuchahnulth, whether it is text, audio, or video, we must choose a standard *format* in which to save our data.

The E-MELD project (*www.emeld.org*) was created specifically to help language researchers across the world choose the right formats for their data. E-MELD offers the acronym *LOTS* to describe good archival file formats:

(2)   **L**   lossless
       O   open standard
       T   transparent
       S   supported

A *lossless* format is one that preserves as much information as possible. Its opposite is *lossy*. Lossy formats destroy information through file compression. While compression makes smaller file sizes, it also involves a loss of information. For example, MP3 is a popular audio format because it takes up so little space, but it does so because it is lossy. Therefore, MP3 is good for sharing, but bad for archiving.

An *open standard* is one that is not owned by any person or company, but can be used freely by everyone. If a format is owned, then the owner can choose at any time to restrict the format's use, or make people pay to use it.

A *transparent* standard is a more technical notion, but means essentially that the data can be accessed directly, without passwords or decryption. This makes it easier to create and share new software that uses the standard.

Last, a *supported* standard is widely used in software like browsers, word processors, and search engines. This ensures that others will be able to read it.

2.3. Archiving versus sharing

While archiving preserves data through time, we also want to use our data today. We must therefore distinguish *archiving* from *sharing*. Unlike archival data, data intended for immediate sharing can be low-quality, can use temporary technology, and should be quick to send and read. Therefore, good sharing formats (also called *presentation formats*) are often different from good archival formats.

For example, AVI is a good archival format for video, because it satisfies all four LOTS criteria. However, because AVI is lossless, its files can be quite large. For sharing, it is often better to use one of the compressed formats MP4, FLV (Flash Player), or WMV (Windows Video), which are smaller and faster to download.

Whenever you create or update data, save it *first* in a lossless, archival format. This will be the archival version. Next, if you like, you can create sharing versions in lossy formats. This order is important, because one can always create a compressed version from an uncompressed one, but not the other way around.

2.4. Metadata

A last important aspect of good archiving is *metadata*. Metadata accompany and describe archival data, providing extra information like where the data come from, who created it, and other details that a person using the data might need to know.

To see the importance of metadata, consider the handwritten notes about Nuuchahnulth produced in the early 1900s by scholars like Edward Sapir, Alex Thomas, Morris Swadesh, and Mary Haas. Today, most of these are preserved at the American Philosophical Society in Philadelphia. In some cases, either archivists or the scholars themselves labelled the materials, indicating who provided the information, who wrote it down, and when. Where there are no metadata, we must guess.

To take another example, years from now someone may want to use photographs or audio recordings that you have made. Will they be able to tell who appears in these photographs, or whose voice is recorded? If you include clear metadata with the materials that you create, then chances are that they will.

Metadata can either accompany the core data, or be embedded within it. The first kind can consist of labels, cataloguing information, or electronic text files. The second, embedded kind consists of elements called *tags*. Metadata tags identify things like *title*, *author*, *date*, and *keywords*—but in principle, tags can identify any kind of information. XML documents, for example, consist entirely of tags and their contents.

An important effort to standardize metadata is the *Dublin Core Metadata Initiative* (*dublincore.org*), which proposes principles for tagging electronic data. For our purposes, however, the quality of metadata often comes down to how you label your data—and in particular, how you name your files.

2.5. Naming files

One of the simplest ways to create good metadata is by naming files wisely. When a person looks for useful information, they might start with a search engine, catalogue, or file system, but eventually they will find themselves looking at files, trying to decide which ones to ignore, and which to investigate. The more transparently a file is named, the faster a user will know whether the file contains what he or she is looking for.

Generally speaking, a file name should say a little about the file's contents, and maybe a little about where it comes from. On the other hand, file names should also be brief, easy to read, and easy to type. Therefore, deciding what to name a file consists partly in deciding what sets it apart from other, similar files.

Keep in mind that files are typically kept together in some kind of container—which might be a cardboard file box, a compact disc, or a computer folder—and that the container itself will provide some information about the files it contains. What file names need to do, then, is say how the files in a container differ from each other.

Therefore, when you name your electronic files, ask yourself how you can best use file names to organize the files that share a folder. These, for example, are some of the important kinds of information that can distinguish one file from another:

(3) • what the file contains
    • the date the file was created
    • the date the file was last revised
    • the person who created the file
    • the person who provided the information

Remember that operating systems and browsers usually list files in alphabetical order. If you put more than one kind of information into a file name, then the file will be sorted by the information that comes first:

(4)   interview_20110501.xml
       interview_20110805.xml
       report_20110314.xml
       report_20110909.xml

If you name files by the date when they were created, consider writing the date in the order year-month-day. This way, when an operating system sorts the files in alphabetical order, they will also be in the proper chronological order.

When you name recordings of interviews, it is sometimes wise to use the filename to indicate who is speaking, what they are speaking about, or both. In this example, audio files of Nuuchahnulth interviews are named according to the date when they were made, the speaker, and the subject:

(5)   `20080930_JA_johnsfamily.wav`
        `20090405_TS_deerstory.wav`
        `20090114_CN_greetingsandthanks.wav`

If you have several documents written by different people, then you might make the first part of each filename the name of its author. The operating system will then automatically sort the files alphabetically by author:

(6)   `powell_1991.pdf`
        `sapir_1924.pdf`
        `sapir_swadesh_1939.pdf`
        `thomas_arima_1970.pdf`

For files that are repeatedly revised, like your notes, it is sometimes a good idea to include the last date of revision in the filename. Whenever you revise, select *File > Save As*, and add the date:

(7)   `notes_tseshaht_20110210.txt`
        `notes_tseshaht_20110401.txt`

Some of the data that we deal with are entirely in the Nuuchahnulth language. In such cases, files can be named either in Nuuchahnulth or in English. Usually English is better for file names, because it is understood by more people, and uses no special symbols. On the other hand, Nuuchahnulth might be better when we need to distinguish Nuuchahnulth words or phrases from each other, because different Nuuchahnulth words can sometimes have the same translation in English:

(8)   `na7aatah_listen.mp3`
        `naacsa_see.mp3`
        `naacsiichitl_see.mp3`

This example also illustrates that when we use Nuuchahnulth in file names, it is best to use *plain text*, because some software might have trouble with Nuuchahnulth symbols. For more on plain text, see SECTION.

There is no single best method for naming your files. Just try to imagine how your work will be used in the future, and use file names accordingly, as a way to help others (and yourself!) to find information quickly and easily.

- *Tip: Since file formats are so important, we need to be able to see file extensions. These are the codes at the ends of file names that indicate their format, like .wav, .txt, and so on. Unfortunately, both Windows and Macintosh typically hide extensions. To make Windows show file extensions, go to* Start > Control Panel > Folder Options *(or, in older versions of Windows,* Start > Settings > Control Panel > Folder Options*). In the* View *tab, uncheck the box labelled "Hide extensions for known file types".*

2.6.   Standard data formats

In this section, we will expand on the brief recommendations given in 2.1. There, we offered quick suggestions regarding which electronic file formats are best (at this time) for archiving and sharing language data. Here, we provide further detail on various standards for text, audio, images, and video.

The following recommendations are based partly on information from E-MELD's *School of Best Practices*, which summarizes the state of the art in electronic archiving. For further information, try starting at E-MELD's *Best Practices in a Nutshell*:

(9) • E-MELD School of Best Practices: *emeld.org/school*
       • E-MELD Best Practices in a Nutshell: *emeld.org/school/bpnutshell.html*

Text Standards. Text intended for archiving should be saved as XML, using Unicode fonts. For sharing, distribute text as PDF, TXT, or HTML.

- **XML** (*Extensible Markup Language*) is a flexible standard that is adaptable to any type of data, including text. An XML file uses *tags* to describe and organize its contents, in a way that is easy for software to decode. XML files have the extension *XML*, or other extensions with an *X* that indicate that they are XML-compatible. For example, XML-based HTML can take the extension *XHTML*. Since 2007, documents created in Microsoft Word and Excel are saved automatically as XML, with the extensions *DOCX*, *XLSX*, respectively. For more on XML, see Section 4.

- **Unicode** is not a file format, but a standard way of encoding the symbols of all the world's languages as numbers. Be sure to use Unicode fonts for any Nuuchahnulth text that you create. For more information on Unicode, see Section 3.1.

- **PDF** (*Portable Document Format*) is the current preferred standard for sharing text. To create PDF files on a Windows machine, download the free *CutePDF Writer* from *www.cutepdf.com*. PDF also has strengths as an archival format, and may improve in the future. Although it is owned by Adobe, it is easily accessible using the free *Adobe Reader*. On the other hand, PDF is more difficult to edit than other text formats, and requires significant technical knowledge to be tagged for archiving. In short, archiving as PDF may be the right choice for some technical users, but XML and TXT are better archiving formats for non-technical users.

- **TXT** (plain text) is the simplest text format, is universally supported, and is even acceptable for archiving, as long as your data require no special formatting. Both Windows and Macintosh include free text editing software, called *Notepad* and *TextEdit*, respectively. If you include Nuuchahnulth symbols in a TXT file, then save it with UTF-8, or another Unicode-compliant encoding:

(10)
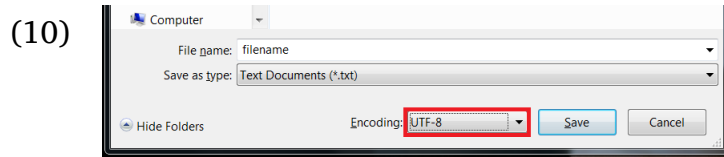
- **HTML** (*Hypertext Markup Language*) is currently the standard format for web pages. It is a good sharing format, and easily viewed with any web browser.

Audio Standards. Archive audio recordings as WAV files. Share audio as WAV or MP3.

- **WAV** (*Waveform Audio File Format*) is currently one of the best formats for archiving audio—even though it is proprietary—because it is lossless and widely supported. Record and save WAV files at a quality of 44100 Hertz, and using one channel (mono). For more on recording audio, see SECTION.

- **MP3** (short for *MPEG-2 Audio Layer III*, and not to be confused with MPEG-3) is the most popular format for sharing audio. It is compressed, universally supported by media playing software, and is the default format for downloaded music.

Image Standards. Archive high-quality images as TIFF. However, where quality is less important, as with scans of old text documents, consider using PNG or JPEG, which are lossy but create much smaller files. Share images as PNG or JPEG.

- **TIFF** (*Tagged Image File Format*) is a lossless format that can preserve images at high quality, and also allows embedding of information about the image, such as its source, its contents, and so on.

- **PNG** is a lossy image format. It is currently one of the most popular formats for sharing images on the web, because it is compressed and loads quickly, and is better than JPEG for sharp images that contain boxes or lines.

- **JPEG** (or *JPG*) is another lossy image format. It is sometimes better than PNG for soft images like photographs.

- **Tip**. *The first step in preserving old documents, whether handwritten or typed, is to scan them. While many scanners scan directly to PDF, this is bad for archiving, because PDF documents are difficult to edit and repurpose (see above). Instead, scan photographs to TIFF, and text to PNG or JPEG. Try scanning as black and white, at 150 dpi (dots per inch). While low quality, this is often sufficient to make the scans readable, and makes the size of the scanned files smaller. Later, the scans can be combined as a single PDF for sharing, or entered into a computer as electronic text to make them searchable.*

Video Standards. For archiving video, we recommend the AVI and MPEG-2 formats. For sharing, we recommend MP4 and FLV.

However, standards for digital video are changing quickly, especially with the increase in use of smartphones and tablet computers. The next important milestone in digital video may be the HTML5 standard, which is still under development, but is intended to establish new standards for sharing video and other content on the web.

A particular consideration with video is file size, which is a more serious issue than it is with audio, images, or text. Video files in lossless formats can be too large to store, much less share. Depending on your situation, and on what you are using video for, you may decide that a lossy format with smaller file sizes is right for your needs.

- **AVI** (*Audio Video Interleave*) is a lossless and widely supported video format. Although recommended for archiving, its files are too large for easy sharing.

- **MPEG-2** is a lossy but high-quality format, and yields smaller file sizes than AVI. E-MELD suggests MPEG-2 as a minimally acceptable archival format.

- **MP4** (*MPEG-4 part 14*) is a non-proprietary video format, the preferred format on Youtube (*www.youtube.com*), and widely supported by hardware and software.

- **FLV** (*Flash Video*) may be the currently most popular format for sharing video on the Web. It is proprietary, and requires the free *Adobe Flash Player*.

- **MOV** (*QuickTime*) is a popular and widely supported sharing format for video. It is better looking than other lossy formats, but has bigger file sizes.

- **WMV** (*Windows Media Video*) is a widely supported, proprietary format. It looks less good than MOV, but produces smaller file sizes.

3.   Symbols, fonts, and keyboards

3.1. Background

If you use the Nuuchahnulth language in computer documents, then you will need a way to type the Nuuchahnulth alphabet. As recently as ten years ago, this was a significant problem for word processing, web pages, and other computer applications. Now these tasks have become much easier, thanks to a standard called *Unicode*.

Section 3 of this guide offers some background to these issues, followed by specific instructions on how to type Nuuchahnulth symbols by computer. In this overview, we begin by explaining the important concepts of alphabets, fonts, and Unicode.

There are several reasons why computing in Nuuchahnulth is not straightforward. The first has to do with the origins of the alphabet. The English and Nuuchahnulth alphabets are based on the Roman, or Latin alphabet. But both English and Nuuchahnulth use more sounds than Latin, making this alphabet inadequate for their needs. Compare the numbers of sounds versus letters in the three languages:

(11)

	Latin[1]	English	Nuuchahnulth[2]
consonants	15	24	35
vowels	10	13-15	8-10
total sounds	25	37-39	43-45
total letters	23	26	45

The English and Nuuchahnulth alphabets address this problem in different ways. Whereas English uses a lot of ambiguous and double letters—like *th, ng, ei, aw*—the Nuuchahnulth alphabet adds enough symbols to provide a letter for every sound. This reflects its careful design, based on many lessons learned about alphabets:

(12)
a	aa	c	c̓	č	č̓	e	ee	h
ḥ	i	ii	k	kʷ	k̓	k̓ʷ	ł	ƛ
ƛ̓	m	m̓	n	n̓	o	oo	p	p̓
q	qʷ	s	š	t	t̓	u	uu	w
w̓	x	xʷ	x̣	x̣ʷ	y	y̓	ʔ	ʕ

The other difficulty with computing in Nuuchahnulth is that computers are designed primarily for English and other major languages. As a result, until just a few years ago, people who wanted to type Nuuchahnulth symbols by computer needed special fonts.

A *font* is a particular way of displaying letters, numbers, and other symbols. Since computers store letters as numbers, a computer font is a mapping from numbers to symbols. When a computer displays letters on a screen, or prints them on paper, a font translates numbers into visual symbols.

Before Unicode, computers in North America and Europe used only the numbers zero to 255 for symbols. This allowed only 256 different symbols, including capital and lowercase letters, digits, punctuation, and so on. If you wanted a symbol not in the usual 256, you needed a *legacy font* that mapped numbers to unusual symbols.

Take, for example, the legacy fonts Nuu2 and "Nuuchahnulth", which were created specifically for Nuuchahnulth. These fonts matched Nuuchahnulth symbols to numbers that were usually reserved for symbols like the slash (/) and caret (^). This becomes clear when text written in these fonts is displayed in a different font:

(13)

Original font	**Example**		**Aboriginal Sans (Unicode)**
Nuu2	hišukʔiš c̓awaak	→	hiSuk/iS Cawaak
Nuuchahnulth	hisˆuk%isˆ c`awaak	→	hisˆuk%isˆ c`awaak
Doulos SIL	hišukʔiš c̓awaak	→	hišukʔiš c̓awaak
Aboriginal Serif	hišukʔiš c̓awaak	→	hišukʔiš c̓awaak

---

[1] The classical Latin alphabet was *A B C D E F G H I K L M N O P Q R S T V X Y Z.*
[2] Some letters in the Nuuchahnulth alphabet are rare, or used only in some dialects, including *e, o, x̣,* and *x̣ʷ.* Short *e* is used only in Barkley dialects, and short *o* only in Kyuquot-Checleseht.

As this example demonstrates, the problem with legacy fonts is that each one uses its own private code. If it is displayed in a different font, the result is gibberish. Now compare the Unicode fonts Doulos SIL, Aboriginal Serif, and Aboriginal Sans. When these are transformed, the font changes but the symbols do not.

Unicode is a new standard for encoding symbols. It specifies a number for every symbol in most of the world's written languages, which is about 110,000 symbols. For Nuuchahnulth, only eight special symbols are needed. These are their Unicode numbers:[3]

(14)	**symbol**	**short name**	**code**	**symbol**	**short name**	**code**
	ł	barred el	0142	ʷ	raised double-u	02B7
	ƛ	barred lambda	019B	̌	wedge (caron)	030C
	ʔ	glottal stop	0294	̓	comma above	0313
	ʕ	pharyngeal stop	0295	̣	dot below	0323

Four of these symbols are stand-alone letters (ł ƛ ʔ ʕ). The others are *diacritics* (ʷ ̌ ̓ ̣), which are typed after a stand-alone letter, and are displayed over, under, or next to it.

Most new software is now Unicode-compliant, including operating systems, word processors, and web browsers. As long as symbols are encoded according to the Unicode standard, the software will figure out a way to display them correctly.

Where problems remain is in *typing* Nuuchahnulth text in the first place. Usually, we enter text using keyboards, which have no keys for Nuuchahnulth symbols. The rest of Section 3 is concerned with how to enter Nuuchahnulth symbols into a computer, in ways that are compliant with the Unicode standard.

3.2 Typing phonetic symbols

This section offers several strategies for typing Nuuchahnulth symbols by computer. There are several ways to do this, depending on whether you need a quick or long-term solution, and on whether you can install software on the computer you are using.

First, if you are word processing, make sure you are using Unicode fonts. Both Windows and Macintosh include Unicode fonts, called *Lucida Sans Unicode* and *Lucida Grande*, respectively. See Section 3.4 on how to download and install other fonts.

Second, the best long-term solution to typing Nuuchahnulth symbols is to install an *electronic keyboard*. This is software that maps keystrokes made on a physical keyboard to unusual symbols. See Section 3.5 on how to download and install keyboards.

However, there are other, one-time solutions that are appropriate for some situations. In this section, we describe *Insert Symbol* and keyboard shortcuts in Microsoft Word. Section 3.3 discusses plain text alternatives to the Nuuchahnulth alphabet.

---

[3] The reason why Unicode numbers sometimes contain letters is that these are hexadecimal, or base-16 numbers, which use sixteen digits instead of ten. The six extra digits are written as *a b c d e f*.

If you are entering Nuuchahnulth text in Microsoft Word, one option is to use the *Insert Symbol* function. Clicking on *Insert > Symbol* calls up the *Insert Symbol* window (see image). This window allows you to search through a grid containing all of the symbols available in the current font. To type a symbol, you can either click on it, or enter its Unicode number in the box labelled "Character code".

(15)
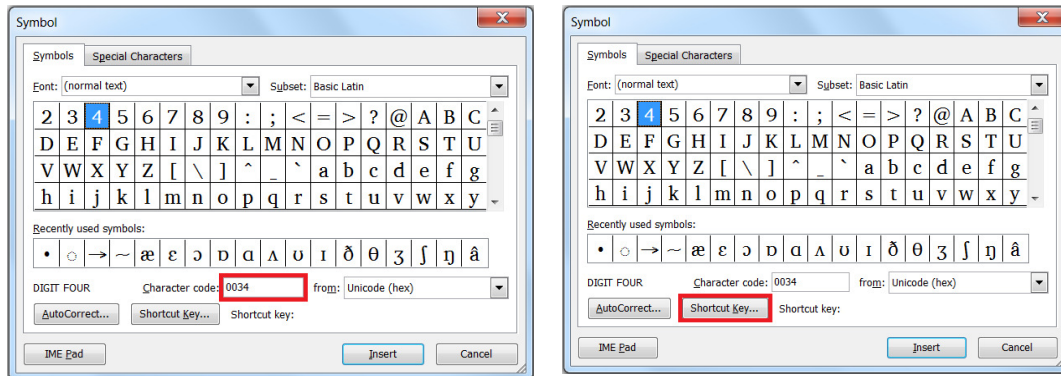
In addition, the *Insert Symbol* window allows you to set up a keyboard shortcut for any symbol. Highlight a symbol in the grid, then click on the *Shortcut Key* button. This will open another window, where you can enter a key combination that will automatically type the desired symbol. You can use any shortcut, but if you like, these are some recommended shortcuts that will not interfere with other keyboard functions:

(16) ł   Alt-;   ʔ   Alt-/   ʷ   Alt-v   ̓   Alt-'
      ƛ   Alt-[ʕ   Alt-?   ̆   Alt-=   ̣   Alt-.

Some drawbacks to these methods are that they work only in Microsoft Word, and only on the computer where you set up the shortcut. Nevertheless, in some cases they may be the best solution, such as on a network computer at a school or office, where you might not be able to install a keyboard, but can still program keyboard shortcuts.

Yet another solution to the difficulty of typing Nuuchahnulth symbols is to write the language in plain text. This strategy is discussed in the following section.

3.3. Alternatives to the Nuuchahnulth alphabet

While the Nuuchahnulth alphabet represents a standard way of writing the language, other alphabets are used as well. Some, like the International Phonetic Alphabet, are technical linguistic alphabets. Others represent Nuuchahnulth with plain text. In this section, we offer a brief description of some of these other alphabets. We address when these are used, and their compatibility with capitalization.

One alternative writing system is the International Phonetic Alphabet (IPA), which is an international standard for representing human languages phonetically. You may choose (or be required) to use the IPA if you write about Nuuchahnulth on Wikipedia (*en.wikipedia.org*), or for another technical or international audience.

Another important alternative to the Nuuchahnulth alphabet is *plain text*. Plain text consists of symbols found on a normal keyboard, which are encoded the same both in older software, and in newer Unicode-compliant software. Some cases where plain text is appropriate are in emails or text messages, in file names (see Section 2.5), and in writing for people who do not know the Nuuchahnulth alphabet.

One plain text alphabet is the one proposed for Nuuchahnulth by Thomas and Arima in 1970, and used in volumes 9 to 12 of *Nootka Texts*. Another system is *Easy Read*, as used by the Huu-ay-aht Na-naniiqsu Cultural Society.[4]

In the following table, we compare the Nuuchahnulth alphabet to the IPA, Thomas-Arima, and Easy Read. In a fifth row, we list additional plain text versions of Nuuchahnulth letters that one sometimes encounters.[5]

(17)

	a	aa	c	ċ	č	č̓	e	ee	h	ḥ
Nuuchahnulth alphabet	a	aa	c	ċ	č	č̓	e	ee	h	ḥ
International Phonetic Alphabet	ʌ	aː	ts	t's	tʃ	t'ʃ	ɛ	ɛː	h	ħ
Thomas-Arima	a	aː	ts	ts'	ch	ch'	e	eː	h	ḥ
Easy Read	a	aa	ts	ts'	ch	ch'	e	ee	h	hh
plain text	a	aa	c	c	ch	ch	e	ee	h	h

i	ii	k	kʷ	k̓	k̓ʷ	ł	ƛ	ƛ̓	m	m̓	n	n̓	o	oo	p	ṗ	q	qʷ	s	š
ɪ	iː	k	kʷ	k'	k'ʷ	ɬ	tɬ	t'ɬ	m	ˀm	n	ˀn	ɔ	ɔː	p	p'	q	qʷ	s	ʃ
i	iː	k	kw	k'	kw'	l	tl	tl'	m	m'	n	n'	o̱	o̱ː	p	p'	q	qw	s	sh
i	ii	k	kw	k'	kw'	l	tl	tl'	m	m'	n	n'	o	oo	p	p'	q	qw	s	sh
i	ii	k	kw	k	kw	lh	tl	tl	m	m	n	n	o	oo	p	p	q	qw	s	sh

t	t̓	u	uu	w	w̓	x	xʷ	x̱	x̱ʷ	y	ẏ	ʔ	ʕ
t	t'	ʊ	uː	w	ˀw	x	xʷ	χ	χʷ	j	ˀj	ʔ	ʕ
t	t'	o	uː	w	w'	x	xw	x̱	x̱w	y	y'	ʔ	ʔ̱
t	t'	u	uu	w	w'	x	xw	xx	xxw	y	y'	–	–
t	t	u	uu	w	w	x	xw	x	xw	y	y	7	9

---

[4] Volumes 9-12 of *Nootka Texts* were recorded and prepared by numerous authors and editors, and published from 2000 to 2009 by the Canadian Museum of Civilization. Other references:

    Thomas, A. and E.Y. Arima. 1970. *t'aːt'aːqsapa: A Practical Orthography for Nootka*. National Museum of Man Publications in Ethnology, No. 1. Ottawa: National Museums of Canada.

    Arima, E.Y., Denis St. Claire, Louis Clamhouse, Joshua Edgar, Charles Jones, and John Thomas. 1991. *Between Ports Alberni and Renfrew: Notes on West Coast Peoples*. Canadian Ethnology Service Paper 121. Hull: Canadian Museum of Civilization.

[5] The IPA interpretations of Nuuchahnulth vowels in this table are based on Carlson et al. (2001):

    Carlson, Barry F., John H. Esling, and Katie Fraser. 2001. *Nuuchahnulth*. Journal of the International Phonetic Association 31. 275-279.

A convention that this table does not make clear is that, in Easy Read, neither the glottal stop (ʔ) nor the pharyngeal stop (ʕ) is written. Rather, a word edge or hyphen preceding a vowel indicates the presence of one of these consonants. A circumflex over the vowel indicates that the unwritten consonant is the pharyngeal:

(18)   **Nuuchahnulth**   **Easy Read**   **English**

      ʔiisaak   ii-saak   'respect'
      wiiʔuu   wii-uu   'nephew'
      ʕiniiƛ   î-niitl   'dog'
      miʕaat   mi-âat   'sockeye salmon'

Related to one's choice of alphabet is the question of whether to use capital letters in Nuuchahnulth. While most technical writing never capitalizes in Nuuchahnulth, some writers choose to capitalize according to the same rules as in English (that is, capitalizing proper nouns, and the first word of a sentence). In most cases, this is up to each individual writer, but there are certain constraints on capitalization.

The Nuuchahnulth alphabet lacks a standard for capitalization. On one hand, most of the linguistic documentation of the language follows the phonetic tradition of using a single, invariant form for each letter, usually lowercase. On the other hand, Nuuchahnulth is used alongside English, which uses both capital and lowercase letters. Further, efforts to capitalize Nuuchahnulth letters are not entirely straightforward, because some Nuuchahnulth letters do not have capital forms—namely, ƛ ƛ̓ ʔ ʕ.

However, if one wishes to capitalize words written in the Nuuchahnulth alphabet, then this guide offers the following recommendation. The letters ƛ ƛ̓ ʔ ʕ always take the same form, regardless of capitalization. All other letters, including the barred l (Ł ł), can be capitalized as in English, and any diacritics added to their capital forms:

(19)   **English spelling**   **Nuuchahnulth**   **Nuuchahnulth with capitals**

      Tseshaht   čišaaʔatḥ   Čišaaʔatḥ
      Uchucklesaht   ḥuučuqƛisʔatḥ   Ḥuučuqƛisʔatḥ
      Gilbert Island   łiwitis   Łiwitis

      Tla-o-qui-aht   ƛaʔuukʷiʔatḥ   ƛaʔuukʷiʔatḥ
      Ehattesaht   ʔiiḥatisʔatḥ   ʔiiḥatisʔatḥ
      Ahousaht   ʕaaḥuusʔatḥ   ʕaaḥuusʔatḥ

Writing in the IPA is never capitalized, because every IPA symbol has a single, standard form. By contrast, Nuuchahnulth written in plain text is not governed by a standard. Since plain text uses the same symbols as standard English, letters can be capitalized or not, according to the author's wishes.
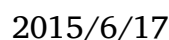
3.4. Installing and using fonts

We have recommended that you use Unicode fonts to write Nuuchahnulth. Here, we provide specific instructions on how to download, install, and use fonts. You might also like to try the instructions at *www.languagegeek.com/faq/InstallingFonts.html*. If you are not technically minded, please ask a techier person for assistance.

The best sources of Unicode fonts for Aboriginal languages are the Summer Institute of Linguistics (*www.sil.org*) and Languagegeek (*www.languagegeek.com*). We recommend the fonts Doulos SIL, Charis SIL, Aboriginal Serif, and Aboriginal Sans.

Begin by following these steps to download a one or more font files that are appropriate for your operating system:

- Find SIL fonts at *www.sil.org/computing/catalog/index.asp#fonts*.
- Find Languagegeek fonts at *www.languagegeek.com/font/fontdownload.html*.
- SIL offers "compact" versions of some of its fonts. Do not download these.
- If your computer asks "Would you like to save this file?", click *Save*.
- After downloading, the files should be on the Desktop, or in the Downloads folder.
- If font files are in a zipped folder (extension *ZIP*), then you will need to extract them. If you have unzipping software installed, simply click on the ZIP file to open a folder showing the font files, which have the extension *TTF*.

Installing fonts in Windows. If the font file has the extension *EXE*, simply click it, and it will install itself. Otherwise, go to *Start > Control Panel > Fonts*. This opens the font folder. Drag the font files into the font folder, and they will be installed.

(20)

Installing fonts on a Macintosh. Click on the font file. This should display a preview of the font. Click the *Install Font* button, and the font will be installed.

(21)

3.5. Installing and using keyboards

Among the methods described in Section 3.2 for typing in Nuuchahnulth, the best is to use an *electronic keyboard*. Unlike a hardware keyboard, an electronic keyboard is software that tells a computer how to interpret keystrokes on the physical keyboard.

For example, North American keyboards have a key in the lower right for slash (/). Normally, when pressed, this signals the computer to enter a slash symbol (Unicode 002F). By contrast, the Nuuchahnulth *Punctuation Keys* keyboard, if installed, causes the computer to interpret this keystroke as a glottal stop (ʔ, Unicode 0294).

In this section, we offer instructions on downloading and installing two different Nuuchahnulth keyboards. One produces Nuuchahnulth symbols from keystrokes usually associated with punctuation, and the other does so from number keys:

(22)   **Punctuation keyboard**   **Numbers keyboard**

The following table conveys the same information as these diagrams. Given a Nuuchahnulth symbol and keyboard, it tells which key enters that symbol:

(23)

| symbol	punct.	numbers	symbol	punct.	numbers
˚̓	'	'	x̣	`	6
˚̇		1	ḥ	;	7
ʷ	-	2	ł	]	8
š	\	3	ʔ	/	9
č	=	4	ʕ	?	0
ƛ	[	5	ŋ		N

Begin by following these steps to download one or both of these keyboards:

• Find Languagegeek keyboards at *www.languagegeek.com/nwc/nwc_keyboards.html*. Scroll down or follow the link to "Wakashan". Follow the appropriate link either to "Mac download" or to "Windows download".
• If your computer asks "Would you like to save this file?", click *Save*.
• This downloads a single ZIP file either to the Desktop, or to the Downloads folder. It contains keyboards for all Wakashan languages, including Nuuchahnulth.
• On the same page, you can also download the keymaps shown above as a PDF file.

Installing keyboards in Windows*.* Click on the ZIP folder to open it. You will see two folders concerning Nuuchahnulth, called *32NuucaanulNumber* and *32NuucaanulPunc*. These contain the software for the number and punctuation keyboards, respectively. You can install one or both, but they must be installed one at a time. First extract or drag the folders to the Desktop. Then, for each keyboard you want to install, open its folder and click the file *setup.exe*, and the keyboard will be installed. Depending on your version of Windows, you may have to restart your computer.

(24)
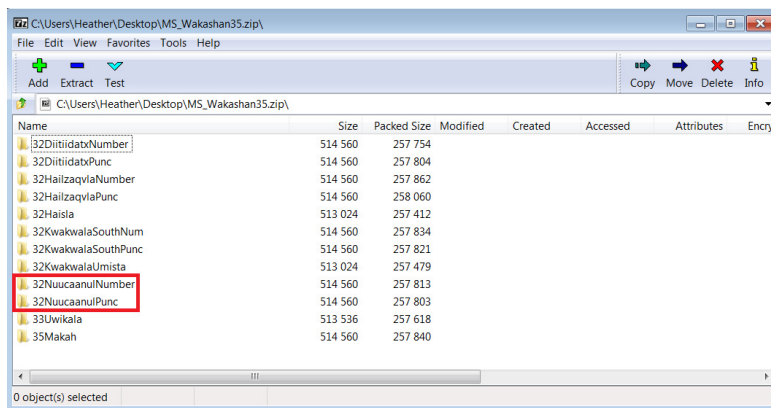
Once you have installed the keyboards, you should see the letters *EN* (for "English") in the System Tray at the lower right corner of the screen. Click on *EN* and make sure that "English (Canada)" is selected. Then click on the keyboard icon to select a keyboard, either "Number Keys" or "Punctuation Keys". This activates the keyboard, and the relevant keys will enter Nuuchahnulth symbols instead of their usual symbols.

(25)
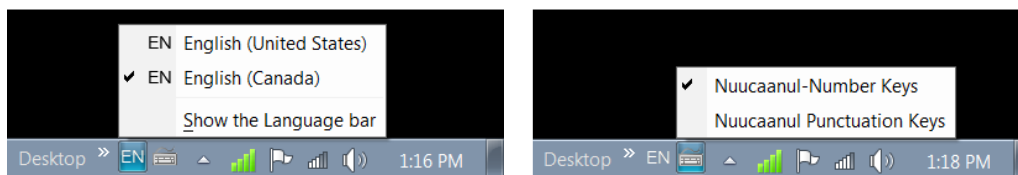
Installing keyboards on a Macintosh*.* The keyboard file that you downloaded is called something like *Mac_Wakashan37.zip*. Click it to unzip it, after which it will be called something like *Wakashan37.bundle*. Now open *Finder > Go > Computer > Macintosh HD > Library > Keyboard Layouts*. Drag the unzipped keyboard file into *Keyboard layouts*. Then log off or restart the computer.

(26)
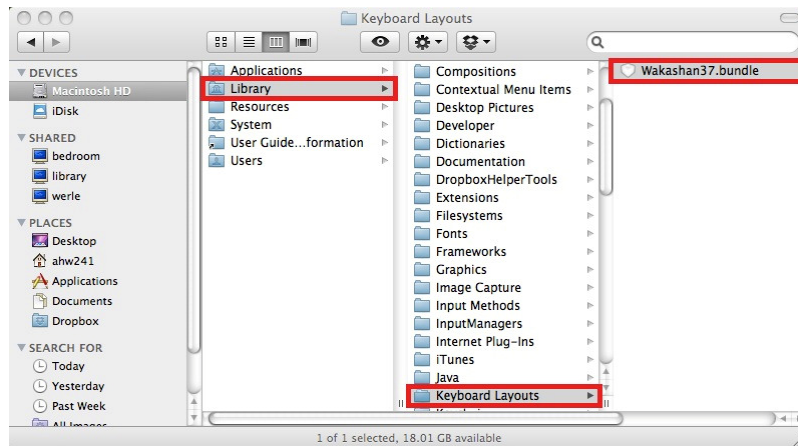
Open *System Preferences* > *Personal* > *Language & Text* > *Input Sources*. Find your Nuuchahnulth keyboard in the list. Put a check in its box, and it is installed.

(27)

To activate the keyboard, click the flag at the top right corner of your screen, and select a keyboard. A Canadian flag means that the English keyboard is active, while a Nuuchahnulth hand means that a Nuuchahnulth keyboard is active, and the relevant keys will enter Nuuchahnulth symbols instead of their usual symbols.

(28)
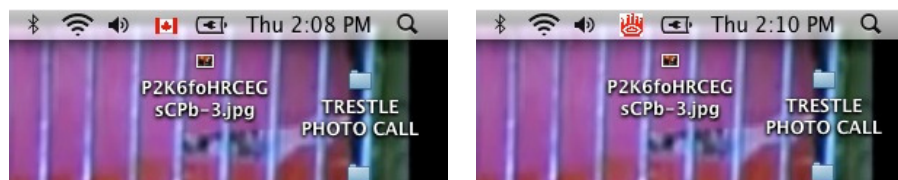
3.6. Converting legacy fonts and alphabets

Not yet written.

4.   Software tools

4.1   Overview

Not yet written.

4.2. Audio with Audacity

These are some of the important steps in recording and archiving language audio:

> (29) *Summary: recording and archiving language audio*
>
> • Check recording settings: mono (one channel), 44100 Hertz (44.1 kHz).
> • Position the speaker, and make him or her comfortable.
> • Check recording levels. The signal should be strong, but without clipping.
> • Record.
> • Optional: edit and amplify the recording.
> • Save audio as WAV for archiving, and as MP3 for sharing.

In the rest of this section, we describe these steps in somewhat more detail. These instructions are oriented towards the audio program Audacity, but are appropriate for most audio software. We recommend Audacity for several reasons:

• Audacity is free.
• Audacity is user-friendly.
• Audacity is powerful, with many different functions for recording and editing. you can use it both to record new language data, and to edit existing recordings.
• Audacity can record for a long time without freezing. This makes it useful for recording long stories, and for digitizing old audio (such as from reel-to-reel tapes, cassette tapes, and so on).

Download Audacity. Download Audacity at *audacity.sourceforge.net*. In addition to the Audacity program, you will need to download a *LAME MP3 encoder* in order to process MP3 files. You will see a link for this when you download the main program.

Check recording settings. If you are planning to record, first set your recording preferences. This can be done before the actual recording session, because it does not require the presence of your speaker. You will find *Preferences* either on the *File* menu, or on the *Edit* menu, depending on your version of Audacity.

• On the *Audio I/O* tab, under *Recording,* set *Device* to the microphone, interface, or sound mixer that you are using to record. Set *Channels* to *1 (Mono)*.

• On the *Quality* tab, make sure that the *Default Sample Rate* is 44100 Hertz.

(30)
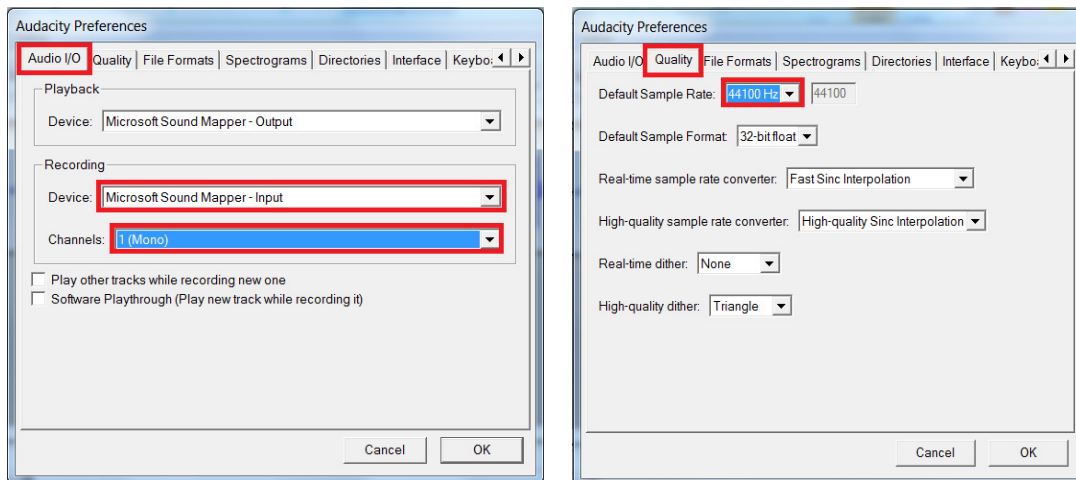
- *Tip*. *For recording language data, it is almost always best to record in mono, or one channel of sound. Stereo, or two-channel audio, is common in professional music and film because the two channels can capture multiple voices and instruments in more complex ways than mono. However, recording in stereo requires at least two microphones, and twice as much electronic storage space as mono. Moreover, most language recording involves single voices, for which stereo is pointless.*

- *Tip*. *The sample rate refers to the detail and quality of a recording. The rate of 44100 Hertz happens to be the standard for audio CDs, and is commonly chosen as a standard for language audio too, because of its high quality.*

Position the speaker. When preparing to record, help your speaker find a comfortable position. An ideal recording position puts the mouth very close to the microphone, but not touching it. Prevent "pop" with a pop filter, or by positioning the microphone so that it is not directly in front of the speaker, but under, above, or to the side.

Check recording levels. Record some test sound to check for pop, and to check recording levels. Audacity represents sound as a *waveform*: the louder the sound, the thicker the line. Set the microphone sensitivity so that the signal is strong (loud and thick), but is not clipped (does not go outside the recording window).

(31)   This signal is strong:
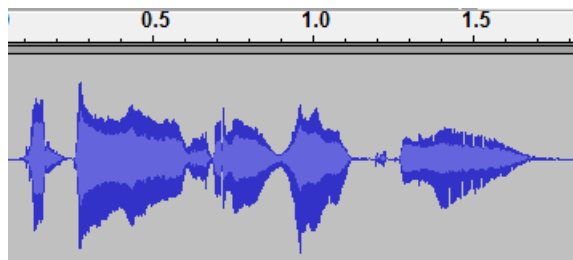
(32)   This signal is clipped:
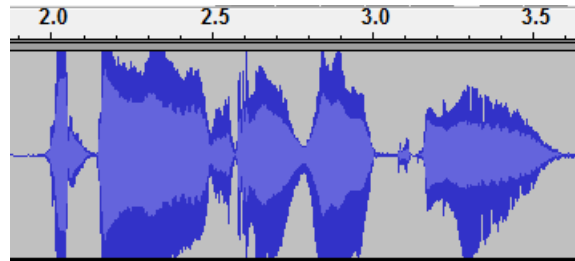
Record. To record, hit the record button. To stop recording, hit the stop button.

Edit. Select the portion of the recording that you want to keep, and trim off the rest with the "Trim outside selection" button. It is often a good idea to amplify your recording, to make it easier to hear on playback. To amplify, select the recording, go to *Effect > Amplify*. Make sure the "Don't allow clipping" box is checked, and hit OK.

• **Tip**. *Always listen to audio after you have edited it, to make sure that it still sounds good. A common mistake is to trim off too much from the beginning or end of a sound, especially if you really just on the visual shape of the waveform, rather than on listening. If you do trim too much, then undo (Edit > Undo).*

• **Tip**. *when playing a recording, you can use spacebar to play or pause the recording.*

Save. It is a good idea when you make a recording to save both an archival copy (usually WAV) and a sharing copy (usually MP3). (See SECTION on naming files.)

• **Tip**. *If you are recording many short segments, like words or phrases, then this tip might save you some time. Instead of recording and saving each segment individually, try recording several at a time, and saving them as one chunk. Later, you can go through these chunks, and edit and save the individual segments. This way, you waste less of your speaker's time during recording sessions, and save the editing for when you are alone.*

• **Tip**. *Sometimes audacity performs unwanted actions while you are trying to type a filename or other text. This can happen when normal keys like A, C, and so on (as opposed to Function, Option, or Control keys) are set as shortcuts. In Windows, go to* File > Preferences > Keyboard. *Click on any commands whose shortcuts are normal keys, and hit Clear.*

4.3. Text with notepad, Word, and Excel

Not yet written.

4.4. Data with Toolbox

Not yet written.

4.5. XML with XMLmind and XLingPaper

Not yet written.

Appendix 1: Online Resources

This is a short overview of some websites that provide software and other resources that are useful for archiving and presenting language data.

(33)

URL	Resources
www.emeld.org	standards, software
www.languagegeek.com	fonts (Aboriginal Serif, Aboriginal Sans), keyboards
www.sil.org	fonts (Doulos SIL, Charis SIL), software (Toolbox)
www.unicode.org	reference to Unicode
www.xmlmind.com	XMLmind program
www.firstvoices.com	FirstVoices archives for Aboriginal languages

Appendix 2: Glossary

archival format: a data format that is suitable for archiving and preserving. Good archival formats are long-lasting, easily accessible, and of high quality. Compare *sharing format*. See 2.2, 2.3.

audio: recorded sound.

browser: software that displays webpages, such as Google Chrome, Mozilla Firefox, Microsoft's Internet Explorer, Opera, and Apple's Safari.

character: a letter or other symbol used in written language.

clipping: an effect in audio recording that happens when the signal is too strong. Fix clipping by reducing microphone sensitivity, either with software or with a mixer.

data: information. The singular form of *data* is *datum*.

diacritic: a symbol that occurs only as part of other symbols, converting them into different symbols. Important diacritics in the Nuuchahnulth alphabet include the dot below (◌̣), the wedge or caron (◌̌), and the hard sign or comma above (◌̓). See Appendix 3.

font: software that displays character codes as letters, and in a particular style.

HTML (*Hypertext Markup Language*): the language in which webpages are written. When you look at a webpage using a browser, you usually do not see the HTML code itself, but rather whatever the HTML of the page tells the browser to show.

images: digital pictures, which might be photographs or artificial pictures.

keyboard: a device for entering letters into a computer. A *hardware keyboard* is a physical device with keys, which are pressed by fingers. A *software keyboard* is software that converts keystrokes on a standard hardware keyboard into non-standard symbols, such as for a minority language.

legacy font: a font that maps numbers to symbols according to some system other than the Unicode standard.

lossless: describes a format that preserves data at a high quality, without compression.

lossy: describes a format that compresses data in order to reduce file size.

metadata: data about data, such as when the data were recorded, who recorded them, keywords, and other information about the data in a file. See *tag*.

mono (*monaural*): having one audio channel. The opposite of stereo.

open standard: a standard that is not owned by any person or company, but can be used freely by everyone. Compare *proprietary*. See 2.2.

pop: an effect in audio recording that results from the puff of breath that accompanies sounds made with the lips, such as *p* and *b*. It can be avoided by positioning the microphone so that it is not directly in front of the subject's lips, or by placing a screen (called a pop filter) in front of the microphone.

presentation format: see *sharing format*.

proprietary: describes a file format, software application, standard, or other tool that is owned by some person or company. Compare *open standard*. See 2.2.

sharing format: a data format that is convenient for sharing with others. Good sharing formats are attractive to look at, have small file sizes, and are easily readable by people or software. Compare *archival format*. See 2.3.

standard: an agreed-upon way of doing things. Following standards makes it easier for people to share and access each other's work.

stereo: having two audio channels. The opposite of mono.

supported: describes standards that are widely used in software.

tag: a kind of metadata that describes the contents of a file, or says where the data are from. For example, MP3 files usually include tags for the artist and title of the track.

text: written language data.

transparent: describes data that are accessible without passwords or decryption.

Unicode: an international standard that provides a unique number for every character of most human written languages. See *unicode.org*.

Unicode font: a font that maps numbers to symbols according to the Unicode standard.

UTF-8: a popular method of encoding Unicode characters while using as little file space as possible. UTF-8 is recommended for storing Nuuchahnulth text.

video: recorded moving images, often with simultaneous audio.

waveform: a way of representing sound in a visual graph. The waveform graphs time on the horizontal axis, and volume on the vertical axis.

XML (*Extensible Markup Language*): an open, flexible file format that is adaptable to any type of data. See 2.6.

Appendix 3: A short reference to Unicode symbols

This is a quick reference to the encoding of some important symbols in Unicode and HTML. Unicode symbols recommended for Nuuchahnulth are marked with a right arrow (→). Symbols that are not recommended are struck through.

		Short name	Unicode character set	Unicode	HTML
	æ	ash	Latin-1	00E6	&aelig
	č	~~simple cee-wedge~~	Latin Extended-A	010D	&ccaron
	ɛ	epsilon (IPA)	IPA Extensions	025B	
	ε	~~epsilon (Greek)~~	Basic Greek	03B5	&epsilon
	ə	schwa	IPA Extensions	0259	&#601
	ħ	barred aitch	Latin Extended-A	0127	&#295
	ɪ	small capital i	IPA Extensions	026A	
→	ł	small barred el	Latin Extended-A	0142	&#322
	ɨ	~~wavy el~~	IPA Extensions	0268	
	ɬ	~~loopy el~~	IPA Extensions	026C	
→	ƛ	barred lambda	Latin Extended-B	019B	
	ɔ	open o	IPA Extensions	0254	
	š	~~simple ess-wedge~~	Latin Extended-A	0161	&scaron
	ʃ	esh	IPA Extensions	0283	
	ʊ	horseshoe	IPA Extensions	028A	
	ʌ	wedge (vowel)	IPA Extensions	028C	
→	ʷ	raised double-u	Spacing Modifier Letters	02B7	
	χ	chi	Basic Greek	03C7	&chi
→	ʔ	glottal stop	IPA Extensions	0294	&#660
→	ʕ	pharyngeal stop	IPA Extensions	0295	&#661
→	̌	wedge (caron)	Combining Diacritical Marks	030C	&#780
→	̓	comma above	Combining Diacritical Marks	0313	
→	̣	dot below	Combining Diacritical Marks	0323	
	ː	double long sign	Spacing Modifier Letters	02D0	
	ˑ	single long sign	Spacing Modifier Letters	02D1	