Andrew CharltonJones

LIS 545

2/9/24

**Repository Overview**

The repository I have chosen to discuss is [Paradisec](https://www.paradisec.org.au/about-us/) (Pacific And Regional Archive for Digital Sources in Endangered Cultures). Paradisec is a repository of musical works, cultural works, grammar tables, language snippets, fieldwork and other data that helps preserve endangered cultures/languages. Paradisec’s main goal is to preserve works that are in danger of being irreparably damaged. The majority of the data in the repository is anthropological recordings or anthropological field notes.

I chose this repository for two reasons, firstly because its subject matter interests me. I am interested in the ways researchers preserve endangered languages/cultures. I especially wanted to explore the role repositories serve in this work. Secondly, I wanted to investigate how this repository handles issues of Indigenous data sovereignty. Indigenous data sovereignty is an extremely important, but often under-discussed topic in data management. Indigenous data sovereignty is a very complex and nuanced idea. A basic explanation is that it refers to Indigenous people controlling their data and the data created about them. A lack of Indigenous data sovereignty can cause major harm. For example, the Havasupai Tribe vs Arizona State legal case, where the Havasupai tribe provided blood samples to Arizona State University for a study. However, these blood samples were used for additional tests without the Havasupai tribe’s consent or knowledge (Kukutai and Taylor 48-49). This included tests that invalidated Havasupai religious and cultural beliefs like DNA mapping (Kukutai and Taylor 48-49). Sacred/culturally sensitive knowledge is also an important concept to consider with a repository like this. This refers to sacred and/or extremely important Indigenous knowledge that the Indigenous community has decided should not be public knowledge. (Burkhart 17-18). I was curious to explore how the repository would handle these concerns while accomplishing its own goals.

Paradisec is not open to general submissions. Paradisec defines their criteria for submissions as “deposits from linguists, ethnographers, ethnomusicologists and other researchers”. Paradisec presents its role as safely storing data with an eye toward future use. Paradisec also aims to digitize reposited data and offers to digitize received analog data if possible. To deposit data one needs to sign up for an account and then fill out a metadata form. One can then provide Paradisec with the data. Digital data can be directly sent to Paradisec. Physical data requires an explanation/description of the data be emailed to Paradisec which they then review before deciding. An account simply requires an email and password. While the repository is closed it seems likely that exceptions are made. The repository aims to preserve as many works regarding endangered cultures/languages as possible. Therefore, it seems doubtful that rare collections in vulnerable formats would be outright rejected because they were not submitted by ethnomusicologists.

The geographic scope of the repository should be briefly discussed. Paradisec’s geographic scope is hard to determine. Their website primarily mentions Indigenous groups in Oceania. For example, a [photo](https://www.paradisec.org.au/) on the website shows people in Vanuatu reading a dictionary of their language. Paradisec’s process statement indicates that they utilize special criteria to gauge whether they will accept a deposit. These criteria are: the language’s status, the state of the deposited work, and the work’s regionality. However, Paradisec’s website never explicitly states what geographical areas they accept. Furthermore, there is data from/regarding distant geographical locations such as [Italy](https://catalog.paradisec.org.au/collections/AC1/items/000), [Denmark](https://catalog.paradisec.org.au/collections/AC1/items/548), and the [Czech Republic](https://catalog.paradisec.org.au/collections/AC1/items/MISC10). This data is part of a massive collection of anthropological interviews from Arthur Capell, an Australian anthropologist. It is difficult to determine if similar work without the Australian connections would be accepted. However, there is a l[arge collection of Nepali recordings](https://catalog.paradisec.org.au/items/search?collector_id=3747) that have no overt ties to the Pacific. Therefore, currently it seems reasonable to assume that work about Endangered languages/cultures from around the world are donatable. The geographical scope may be more explicitly addressed when actually donating.

The repository is very cognizant of and prepared for data containing culturally sensitive knowledge. The submission page mentions that donors can set whatever viewing guidelines they require. This seems an important safeguard that allows data with sacred knowledge elements to be stored without being widely accessible. Paradisec also provides contact methods for Indigenous communities to use if they discover sacred information.

Viewing and interacting with the material stored in the repository requires users to create an account. As mentioned above, accounts simply require an email address and a password. However, while this will allow you access to most data, not all data is accessible or interactive. Accessible data has the word open in green. The site also mentions that some data stored in the repository is not viewable by users as well. These methods seem designed to safeguard Indigenous data sovereignty and sacred knowledge.

The data is available in a large number of formats and is accessible via several access points as well. For example, this interview about [weaving](https://catalog.paradisec.org.au/collections/AAZ2019/items/CON001) is provided via jpeg, video, audio, and XML. The interview can be downloaded, read on the browser, listened to via the browser, or can be watched via video file in the browser. The files available are tied to the content itself. For example, [this collection of questionnaires](https://catalog.paradisec.org.au/collections/AC2/items/FIJ004) only has two available formats. However, it does still offer a downloadable and a browser-usable access point.

Paradisec’s catalog is run using Nabu, a media management system designed by Paradisec. Nabu is open source and provides a [GitHub repository](https://github.com/nabu-catalog/nabu) with the source code. Paradisec’s catalog uses metadata that confirms to Dublincore and Open Languages Archives Communities Standards. Paradisec also offers a pdf explaining their catalog’s metadata features. Paradisec’s metadata files are provided in XML format.

The DIP provided when you download Paradisec files requires confirmation that the user will not violate copyright, repurpose for commercial use or exploit the data in any way. It is similar to agreements that one gets when downloading info from other the internet other repositories. The DIP is primarily concerned with the various laws or rights involving the data. For example, it explains that most data is still under copyright law.

Works Cited

Burkhart, B.Y.. What Coyote and Thales can teach us: An outline of American Indian epistemology, In *American Indian thought: Philosophical Essays*. (p.15–26). Blackwell Publishing., 2004.

Kukutai, Tahu and John Taylor. *Indigenous Data Sovereignty : Toward an Agenda*. 1st ed., vol. 38, ANU Press, 2016.