



# Digitizing Kashmiri Literary Works into कॉशुर

कॉशिर्य अदबी कामि (देवनागरी मंज़)



ॐ नमः  
शिवाय



# Overview

The current online presence of Kashmiri content is minimal, with an even smaller amount available in Devanagari Kashmiri. Moreover, the limited content that does exist is not compliant with modern Unicode standards. Given the global dispersion of the Kashmiri Pandit community, the preservation and enjoyment of our literature now largely depend on digital means. Making Kashmiri content, particularly in Devanagari Kashmiri (कॉशुर), widely accessible online will not only help introduce the next generation of Kashmiri Pandits to the Kashmiri language but also create a foundational data set that could enable major tech companies like Google and Microsoft to develop language tools such as translators, spell checkers, and more.

If we do not take ACTION NOW, we will surely lose this vast treasure and our language.





# Why are we doing it

1

**Lack of digital Devanagari Kashmiri content on the web.** Today, there is next to nothing standard Devanagari Kashmiri content on the web.

2

**Modernize existing digital content.** The Devanagari Kashmiri content that is available is not modern Unicode compliant and hence not consumable by Google/Microsoft search bots.

3

**Build a foundation** for modern day machine learning tools to learn from this content and allow कश्मूर to be available in translation softwares, dictionaries, spell-checkers, dictation systems

4

Provide a **web searchable digital library** for Kashmiri Pandit community, especially our future generations, to get access to Kashmiri literature and enjoy the literature and other audio-video content that can be built on top.



Project strategy

1. Keep the goals simple and achievable.  
Slow progress is ok. (KISS principle)
2. Focus on actual individual contribution. Less talk, more work.



Project Goals

1. Gather as many sources of literary works as possible.
2. Digitize and Securely Store the content on cloud.



Project Non Goals - WHAT WE WILL **NOT** DO (in the first phase)

1. Build any tools or dictionaries or spell checkers on top of the digital library.
2. Replace any existing content with new content



# Literary Sources

1

**Manuscripts** - Handwritten, that never made it to a publisher. Estimated at 90% Nastaliq, 10% Devanagari.

About ~100 books

2

**Physical books** - Non Standard Devanagari Kashmiri.

About ~100 books

3

**Digital Books/Documents** - Non Unicode Devanagari Kashmiri.

About ~100 books

4

**Physical books** - Nastaliq

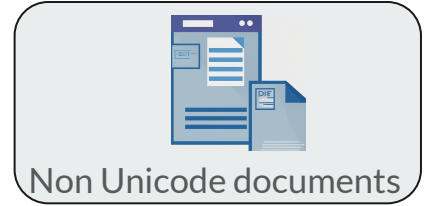
About ~100 books



# High Level Process

# काँशुर खज़ाकक कलथकँन्य बनावव

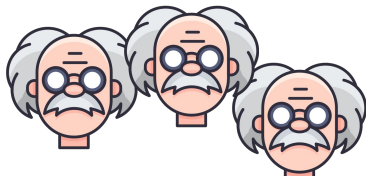
Gather and collect Kashmiri Literary sources



Recruit a data entry team



Create a Subject Matter Expert team



Build Non-Unicode to Unicode Converter



upload

convert



Google Drive on Cloud

data entry

verify



**Thank you.**

