Collaborative Buddhist Text Annotation with Layered Document

Yap, Cheah Shen
yapcheahshen@gmail.com
ECAI Workshop, PNC 2016
August 18th, 2016
J. Paul Getty Museum, Los Angeles, California

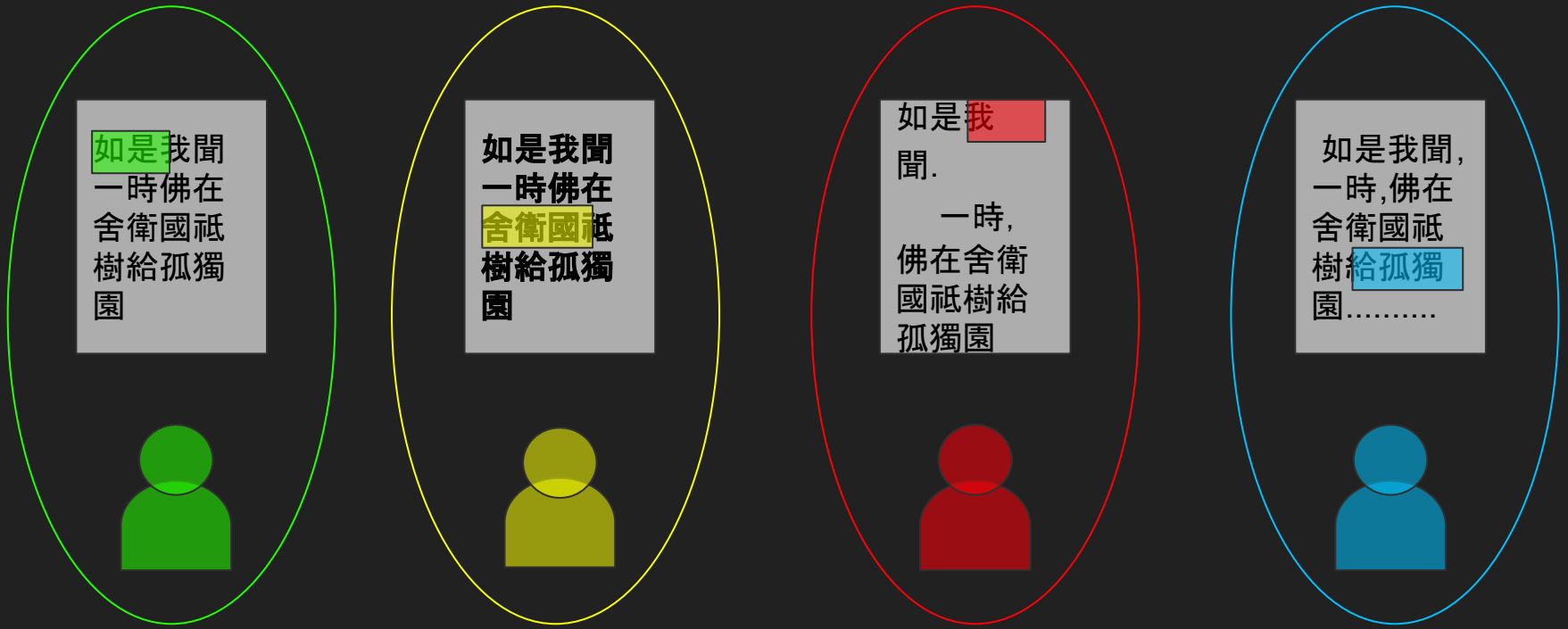
The Insularization caused by Copy/Paste

1) To create text annotation, we make a copy from an authoritative source, e.g CBETA, and paste it to the software (usually a text editor).

2) The copied text is in fact a new independent version, which cannot reflect changes in source text which are made after copy/paste.

3) As annotations are tied up with different version of base text. Annotations cannot be merged without human intervention.

Multiple Insularized File



Single Version Approach

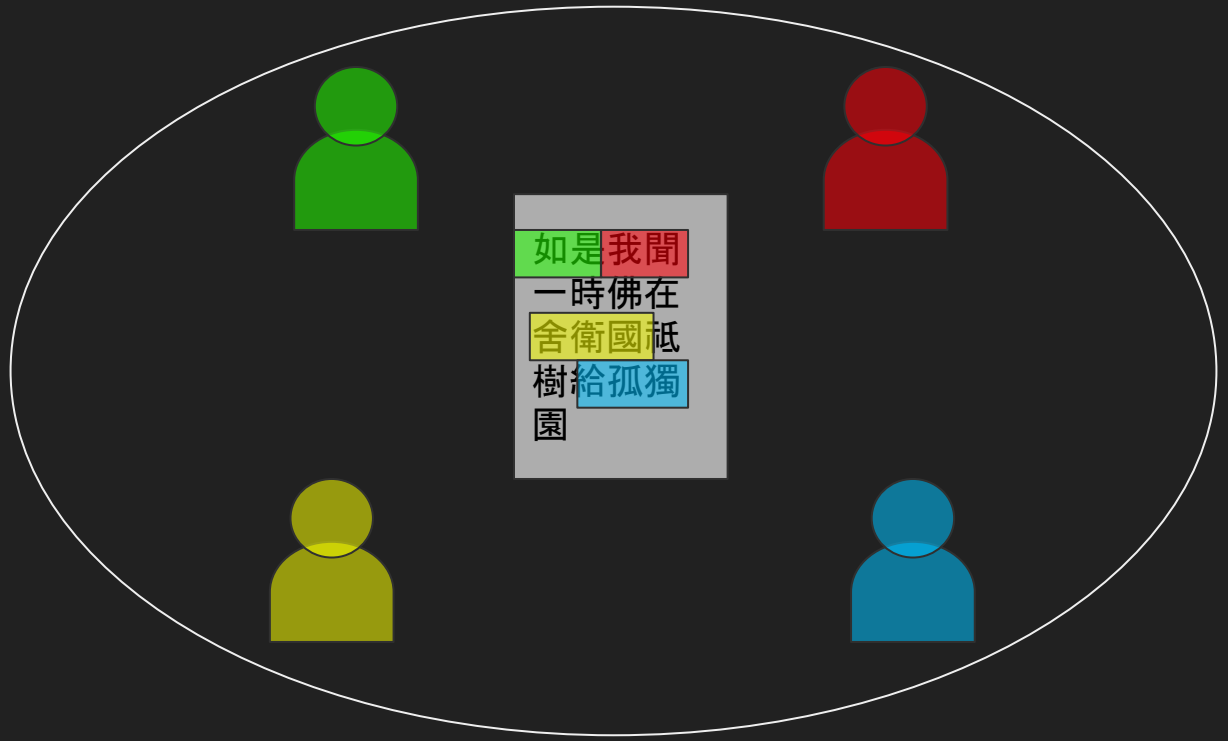
Do not make copy, everyone should work on same XML/TEI file.

To make sure everyone working on same text have identical clones:

1) Single-write-multiple-read policy.
2) Distributed Revision Control System (Github)

Even we can enforce everyone to work on same XML/TEI file. It will become a big mess soon enough. Imagine 50 students writing note on same text book.

Single Version Approach



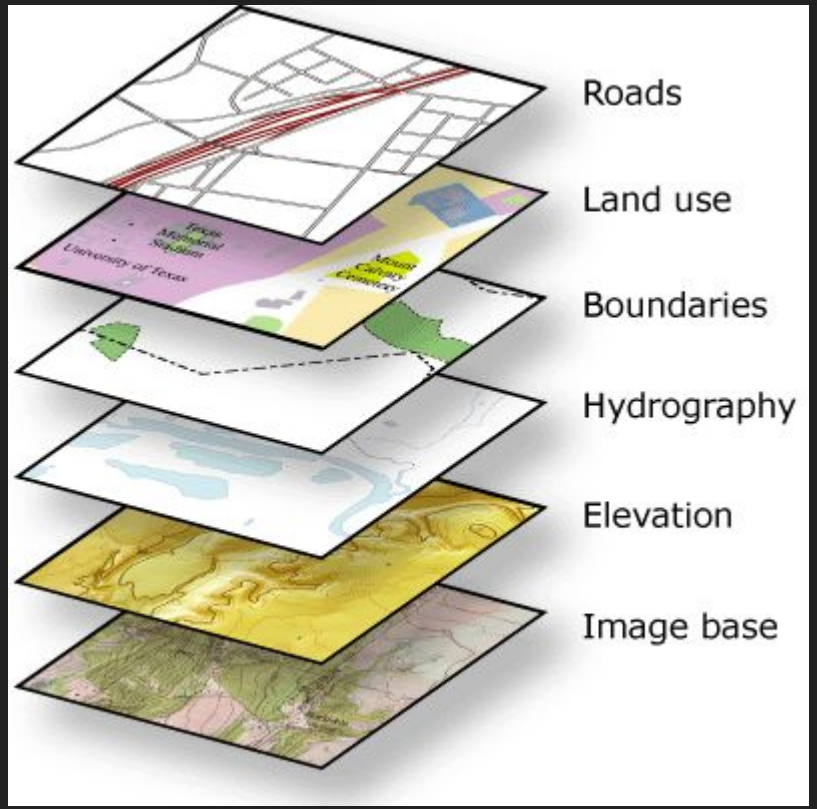
“Too many cooks spoil the broth, Too many tags spoil the file.”

Solution: Layered Document

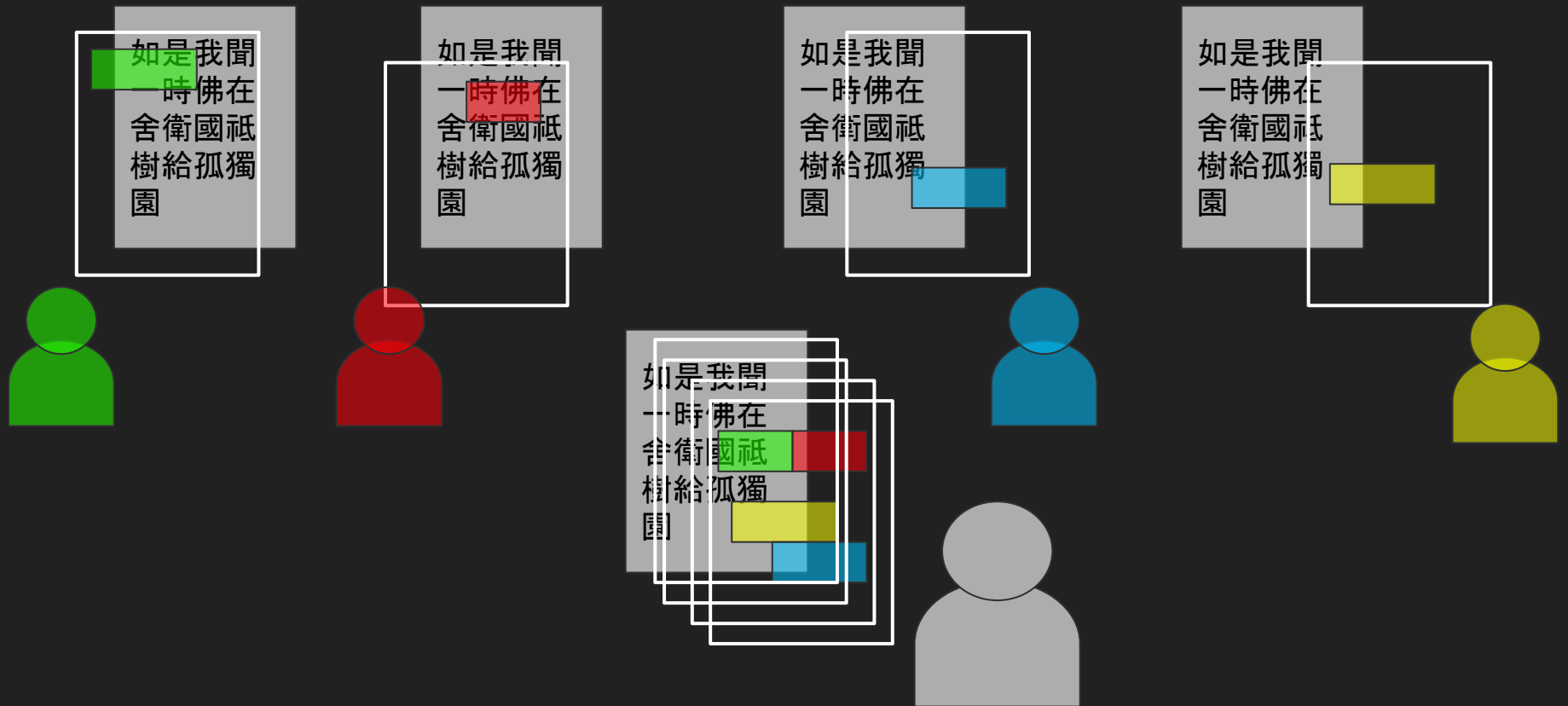
Annotate on a separate layer.

layer works like a transparent film on top of base text, which can be added, removed and combined easily.

Annotator has full control to his own layer, but cannot change the base text and layers created by other annotators without permission.



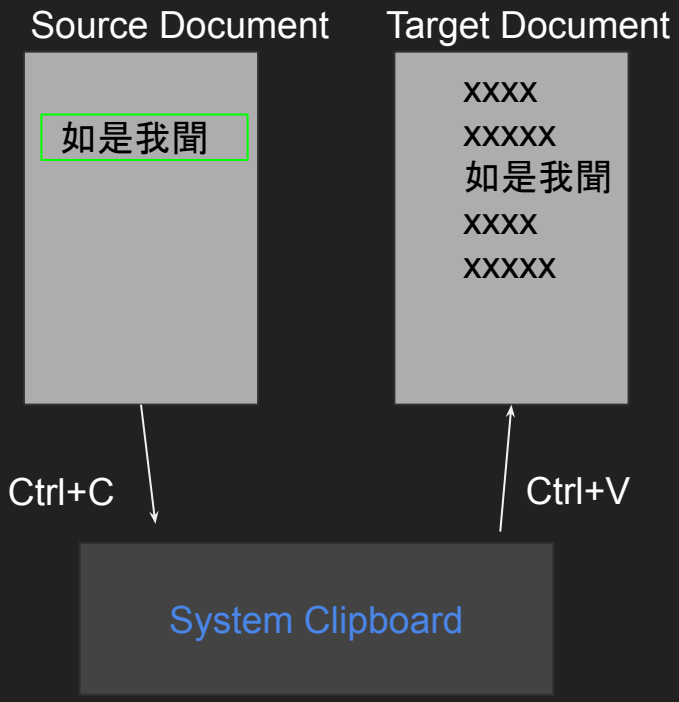
Layered Document

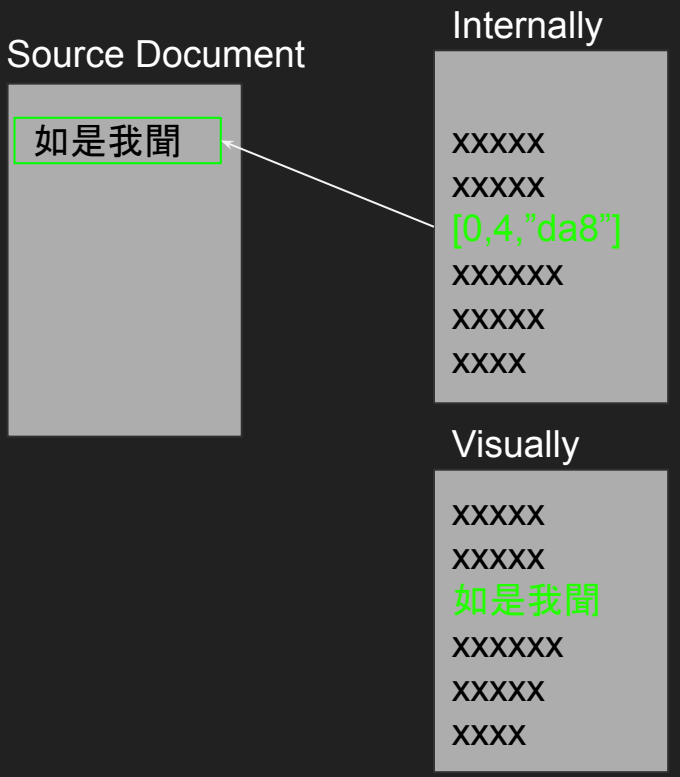


Copy&paste

vs

Text pointer in Layered Document





A Prototype of Layered Document system

1) Web based layered text editor
2) 100% Javascript
3) Highly customizable
4) Optional tree structure (for Kepan)
    a) XML has to be an valid tree structure in memory (as required by Document-Object-Model)
    b) Tree structure prevents simple slicing and merging of XML files.

LIVE DEMO

This Prototype of Ksana Layered Editor is sponsored by

Bodhiyana Foundation http://www.bodhiyana.org

Thank you for your attention

URL of this slide

https://goo.gl/E0jalv