

Converting legacy data to CLDF: A FAIR exit strategy for linguistic web apps

Robert Forkel, Daniel Swanson and Steven Moran

The Heyday of Linguistic web apps

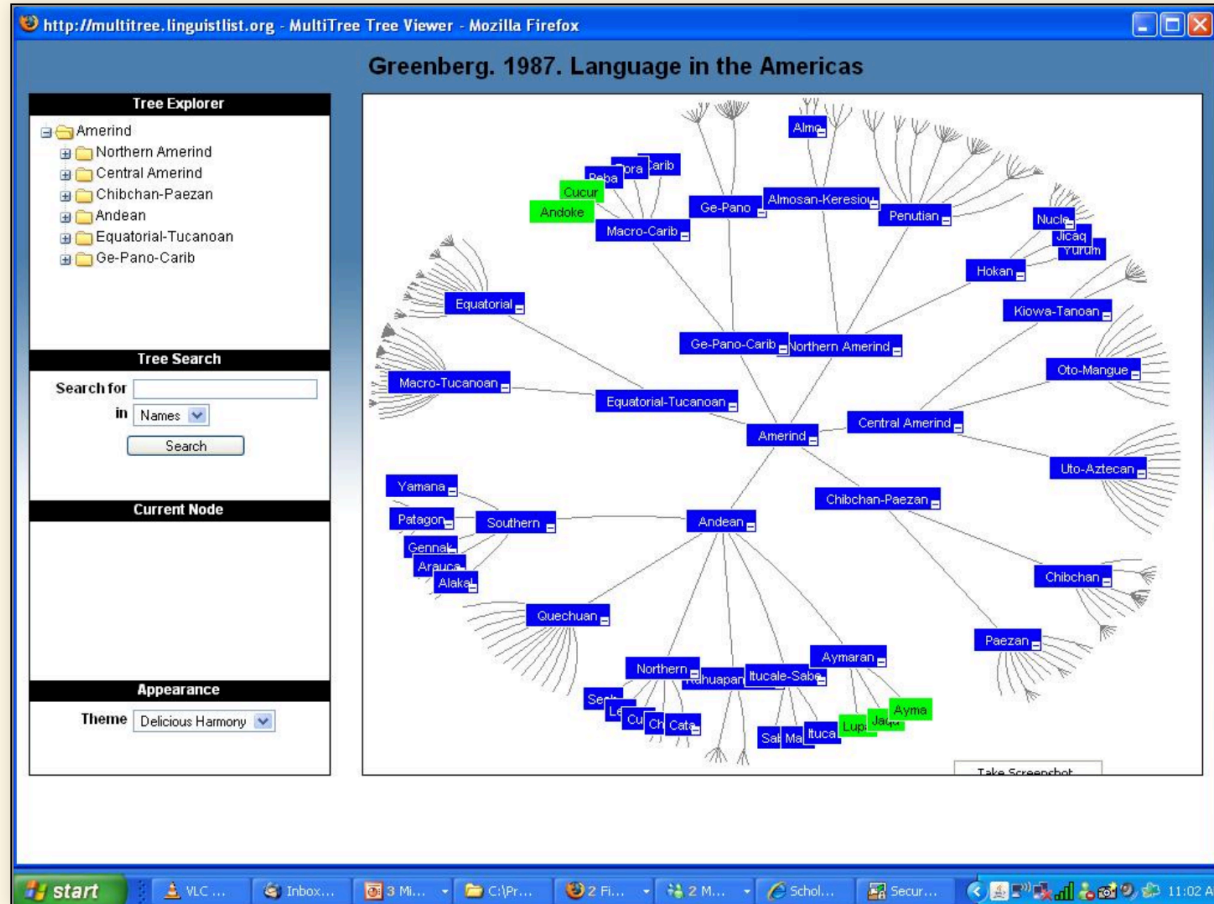
- Web dev became easier
- Browser-based access promised widest reach
- NSF and other funders were on board
- mid 2000s Linguistic Data goes On-Line!
- WALS Online 2008, etc.

Was that FAIR?

- Findable?
- Accessible? Longterm?
- Interoperable?
- Reusable?

MultiTree

- hypotheses about the genealogical relationships among languages
- web app displayed language trees for the browser
- “raw” data available in custom XML



LL-MAP

- **L**anguage and **L**ocation: **M**ap **A**nnotation **P**roject
- Geographic information about languages was digitized
- and put in a GIS-enabled web app for exploration

[Edit](#) [View](#) [History](#) [Bookmarks](#) [Tools](#) [Help](#)

<http://linguistlist.org/llmaptest/home.cfm?tab=google.html#loc=-117.7734375,37.37015718405753,6&layer>

[http://linguistli...?tab=google.html](#) [LL-MAP Data Browser](#)

LL-MAP
[Maps & Data](#)
[Scholar's Workbench \(Login\)](#)
[Become a Data Partner](#)
[About LL-MAP](#)
[Links](#)

[graphy](#)
[Languages](#)
[Political Divisions](#)
[Flora & Fauna](#)
[Find new layers](#)
[Clear Map](#)
[File](#)
[Search](#)

[Map](#)
[Satellite](#)
[Hybrid](#)

[Yokuts - yok](#)

[Contact Us](#)
[Help](#)

Map data ©2007 LeadDog Consulting, Tele Atlas - [View Map Data](#)

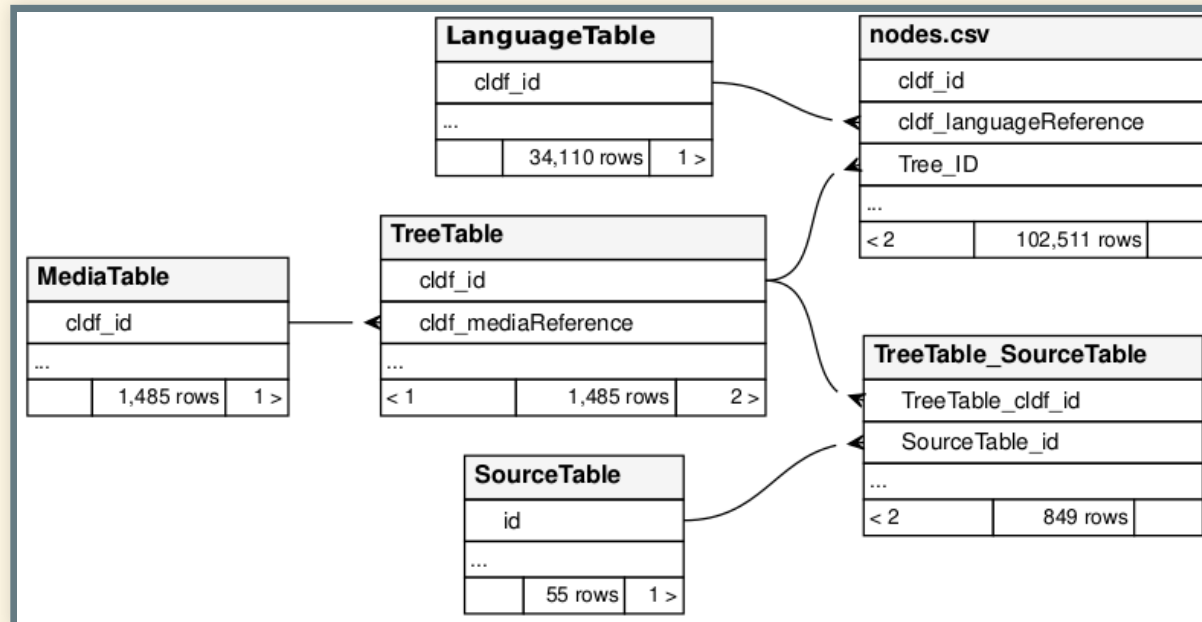
Next Exit: CLDF

- Model data as inter-related tables.
- Map to CLDF components and properties when appropriate.
- Implement conversion as automated pipeline.

Automate with cldfbench

- Legacy data serves as input,
- and is enriched with expert annotations by an editorial team
- and converted to a CLDF dataset including transparent provenance information.

MultiTree as CLDF dataset



```
(ll) robert@lingn35:~/projects/ll/multitree$ cldfbench multitree.show 12067
```

Xincan: Campbell 1997

Source:

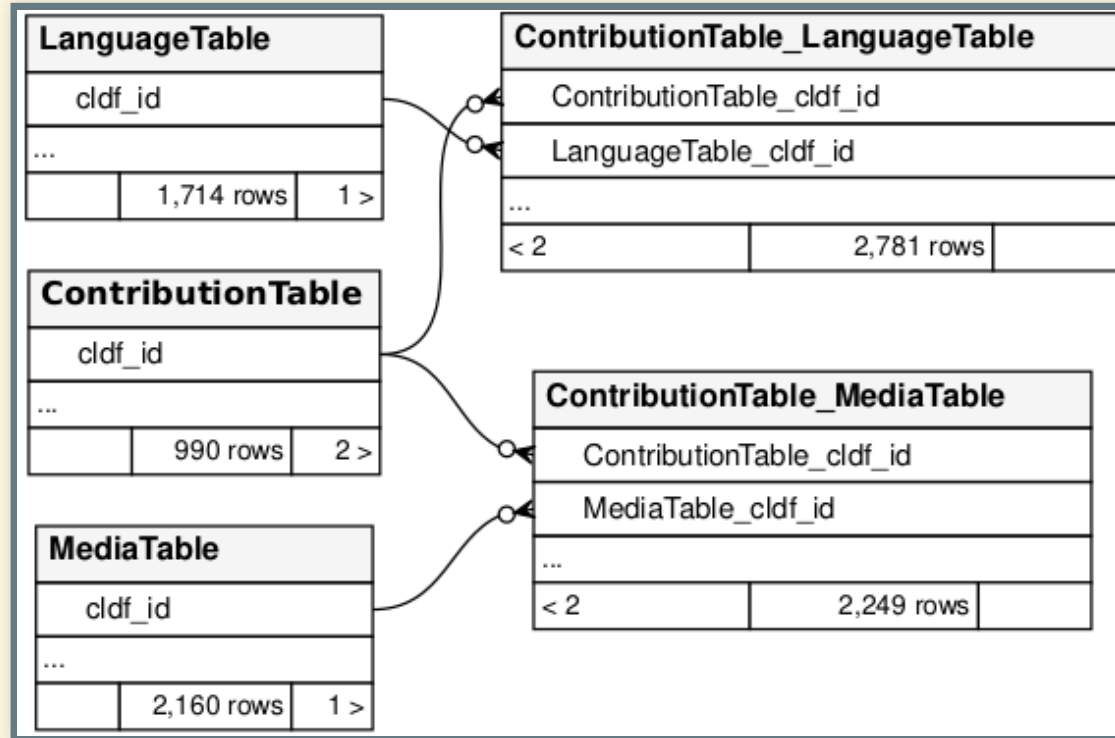
```
    Campbell (1997) [166]
      /-qco_jut
      /-qco-----|
      |             \-qco_yup
--xink-----+qhq
             | -qda
             \-2df
```

Label	Name	Glottocode	Type	Geography	Status
-----	-----	-----	-----	-----	-----
xink	Xincan		Subgroup	Guatemala	
qco	Yupiltepeque		Language		Extinct
qco_jut	Jutiapa		Dialect		
qco_yup	Yupiltepeque		Dialect		
qhq	Jumaytepeque		Language		
qda	Guazacapán		Language		
2df	Chiquimulilla		Language		

References:

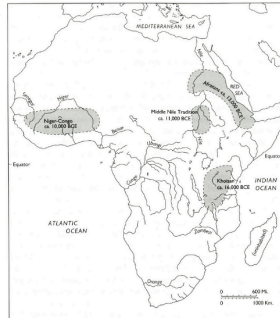
Campbell, Lyle. 1997. American Indian Languages: The Historical Linguistics of Native America. Oxford: Oxford University Press.

LL-MAP as CLDF dataset



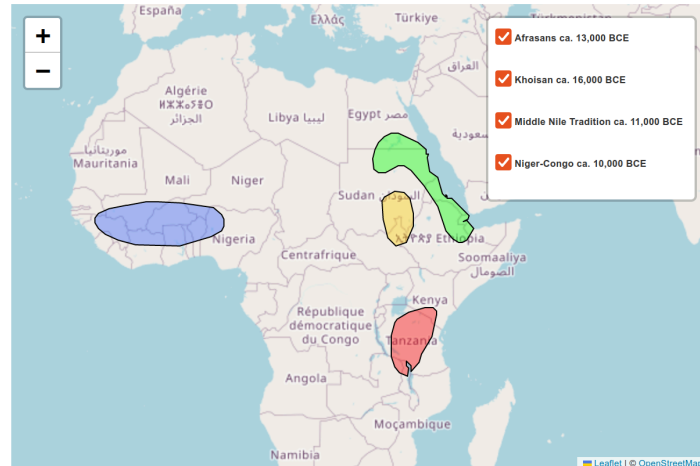
Africa Before the Agricultural Age (Ehret)

Outlined on this map are the locations that Ehret hypothesizes gave rise to the Afrasan, Middle Nile, Khoisan and Niger-Congo traditions. In his book, he describes how each area affected the development of these ancient cultures as well as some of their differences, citing examples such as the Afrasan reliance on wild grass collection for food, and Middle Nile tendency to hunt large game and fish.



MAP 4 Proposed early lands of Afrasan, Middle Nile, Khoisan, and Niger-Congo traditions

over resources, the local advantage in numbers would recurrently have favored the Afrasans.
At the same time, the economy of the early Afrasans offered a way out of the food crisis that their expansion had brought about: the new economy carried with it the ideas and practices of wild grain collection. Faced with an expand-



Conclusion

We demonstrated that data, formerly only available through - now defunct - web apps, can be

- converted to CLDF and archived
- thereby providing FAIR longterm access
- allowing essential analysis functionality to be rebuild easily

