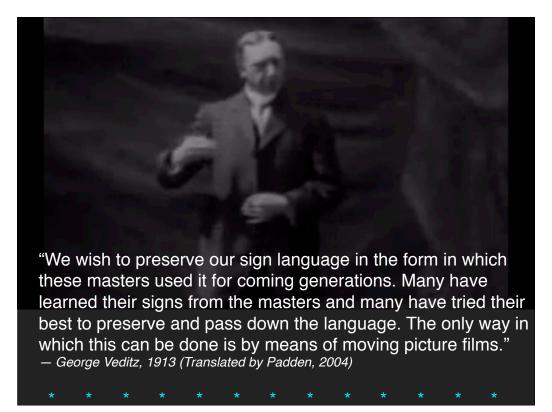
JULIE A. HOCHGESANG AND EMILY SHAW MAINTAINING THE STORIES OF THE DEAF COMMUNITIES AT GALLAUDET



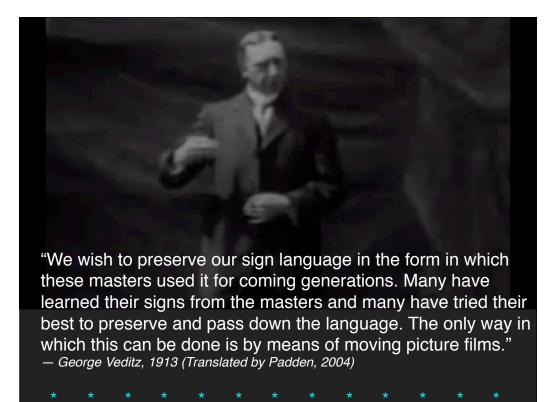
YOU ARE WELCOME TO LIVE-TWEET OR TAKE PICTURES AND SHARE ON SOCIAL MEDIA. CONFERENCE HASHTAG #MIII

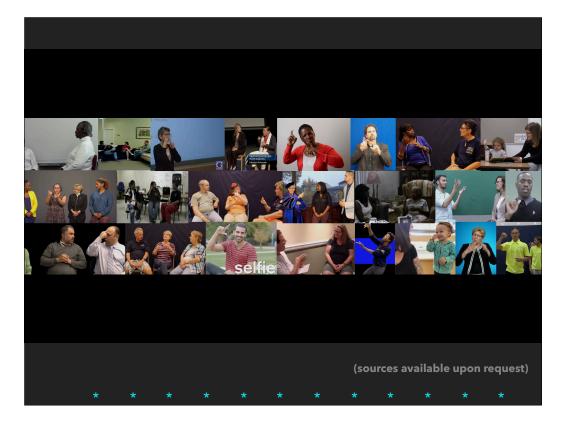


American Sign Language (ASL), the language of the American Deaf communities, is about two hundred years old. But no widespread written language to represent ASL has ever been developed. This means there is no direct textual representation of the stories that have been told. But as soon as video cameras were invented, ASL stories have been filmed. We want to talk about this a bit today.



Deaf people all over the world sign to communicate. There's no one signed language but hundreds. Similarly, There's not one American sign language (ASL) but hundreds of variants. And they exist in the thousands of stories of American deaf people as long as cameras have been around. We want to preserve these stories. Many of them are here. At Gallaudet. One of the oldest institutions serving deaf people in the world.

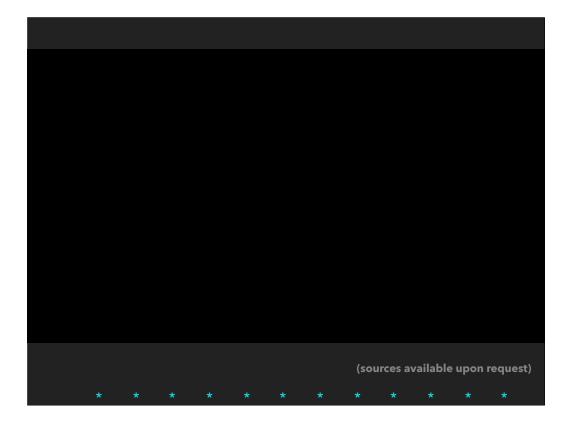




For a minute, I want to look at what we're talking about today. Just a minute of video with all kinds of ASL videos edited together.

#### (Watch video)

So, after having seen all of these, even if you don't know ASL yourself, you can see how rich and nuanced language use is? You can see the different representations - different people, different social contexts, different genres. And if you do know ASL, you probably caught some signs and can remember them now - rocket, funny, etc. Now imagine trying to go back and find what you wanted. You need to be able to search for it. We need to be able to represent this into text so we can use that for that purpose. We can't study language in its raw form - as sound waves or light waves on the air. We need to freeze that behavior somehow. For signed languages, we are unable to rely on already conventionalized written languages like for English or the International Phonetic Alphabet (IPA) that linguists can use to represent any spoken language.





As you've just seen, Here at Gallaudet, since cameras have been existing, people have been filming various things on campus here at Gallaudet. Examples include filming presentations on campus, different student campus events, personal narratives, filmed lectures, filmed tasks for research projects. They all capture some kind of ASL use. All of this is available on Gallaudet Video library or Youtube channel as we can see here. Gallaudet has been around since 1864. The first filmed video was from 1910. It's 2019. That's over 100 years of filming. So much potential here - especially given that these are mostly community-generated. Some are elicited for research too. And there's plenty around North America too. I often get requests to help host collections - home videos, other research projects, and so on.

We have a lot of stories we can use. So collecting them isn't so much the issue. We want to take care of the stories we already have.





So there are thousands of videos, starting in 1900s and exponentially increasing through the years They're scattered across campus, although some are hosted in the Gallaudet Video library

They're in different formats - VHS, CDs, different digital formats

They're not processed consistently - if at all - meaning they're not named in the same way, organized with the same metadata and so on

# ASL VIDEOS AT GALLAUDET

\* thousands of videos
\* some in Gallaudet Video library or YouTube but most scattered across campus
\* different formats

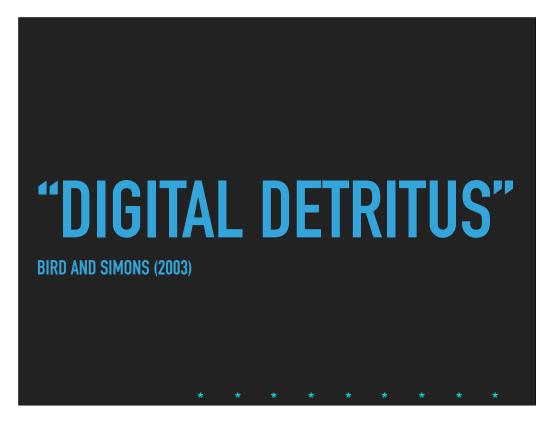
processed differently (if at all)

## ASL VIDEOS AT GALLAUDET

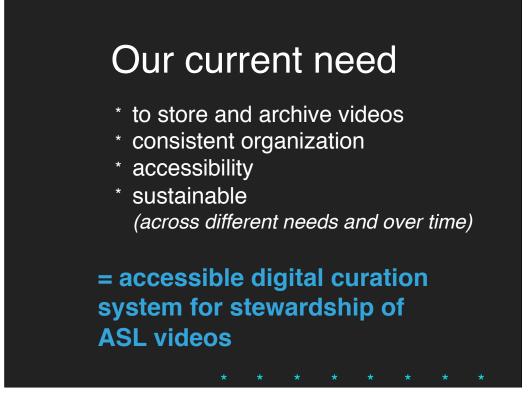
thousands of videos

- some in Gallaudet Video library or YouTube but
- most scattered across campus
- different formats
- processed differently (if at all)

### ¢≠accessible

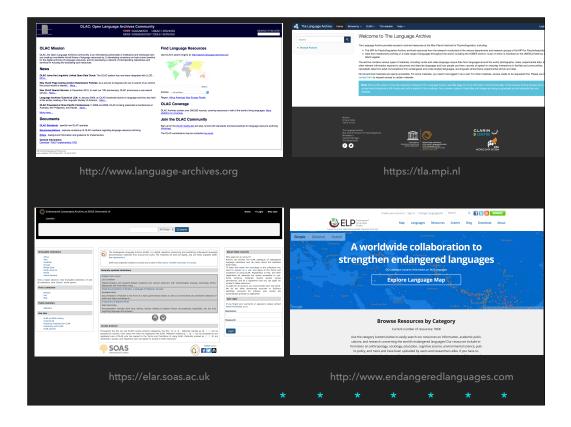


All of this means that they're inaccessible - basically 'digital detritus' as described by Bird and Simons in their 2003 article on portability of data for language documentation

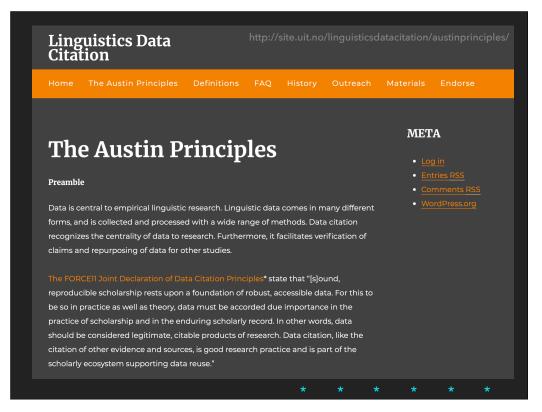


Our current need - and this has been echoed by many - is a place to store, consistently organize, access (make machinereadable) and share our videos. And it's something that needs to be sustainable - maintained over time.

At Gallaudet, we've done this management of our videos somewhat individually in departments and in the Gallaudet Video Library which uses the Telescope Digital Assets Management program. It is currently under the domain of University Communications. But it has no more financial support and needs to be updated or we need to move to other digital resources asset management.



Basically we need language archives something like these - stable repositories with consistent metadata and technicians who update the materials as technology evolves - as well as terms of use, graded permission levels if needed, and so on.



We can also refer to The Austin Principles of Data Citation - which is a general push by linguists to make primary data (spoken, written or signed) accessible. This will be guidance as we develop best practices in making our ASL video collections more accessible to everyone. Personally for me, this is essential as a linguist and as a Deaf woman because it ensures access to the primary data - something that's very hard to do when looking at publications - usually text on paper - which creates challenges for representing signed languages which do not have standardized written forms.

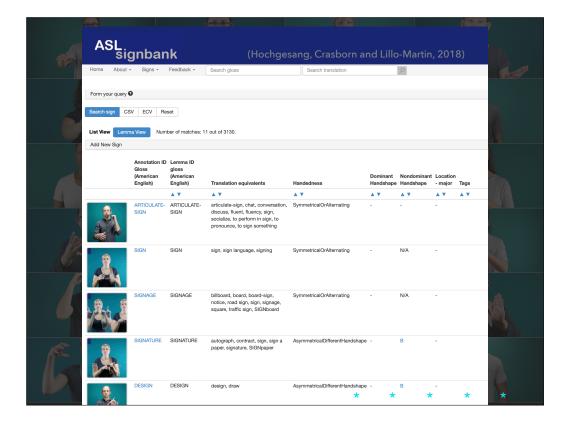


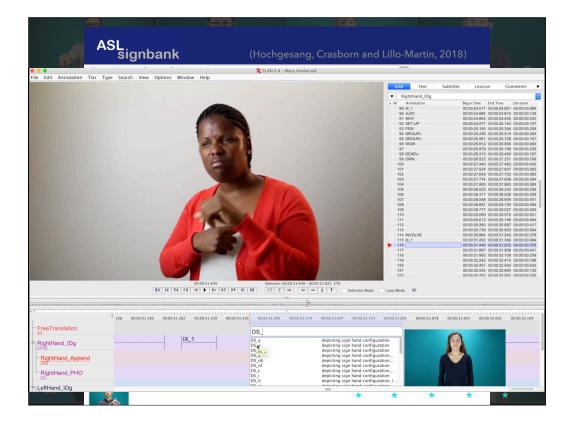
Speaking of not having written systems for signed languages... while we don't have standardized writing systems, we have something called glossing which means we use the written system of the majority spoken language - for us in America that means written English in upper case form is used to represent ASL. We don't have time to get into this very much but I do want to mention that this is problematic because it's a secondary representation of a different system. One way we've gotten around this in our field is to use something called ID-glossing (e.g., Johnston, 2010). It basically looks the same as glossing but when used with a database and video annotation, we're able to use glosses as indexical representations. It helps to reduce the impact of the written system. Still plenty of problems there but when used as intended...

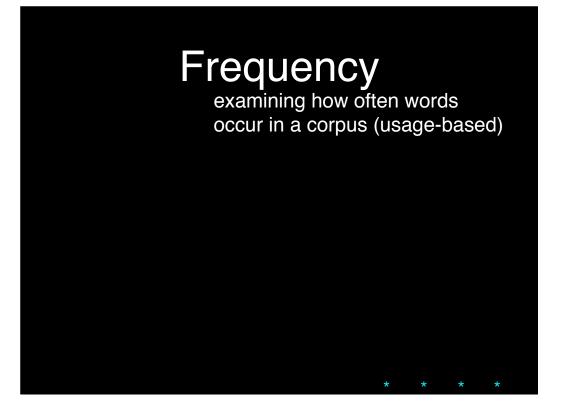
For example, take the ASL Signbank which I use as a maintenance of ID glosses for ASL signs (there are other sign banks for other sign languages - we modeled ours after theirs). ID glosses are "identifying glosses" (unique labels used to represent the same signs)-now it's not the main form of representation like we want but rather a form of labeling data (or ASL signs) so we can find them in the videos we have. This screen recording with the video we showed before but in ELAN, a particular software annotation program, with the ASL Signbank linked shows how ASL becomes more prominent in the representation.







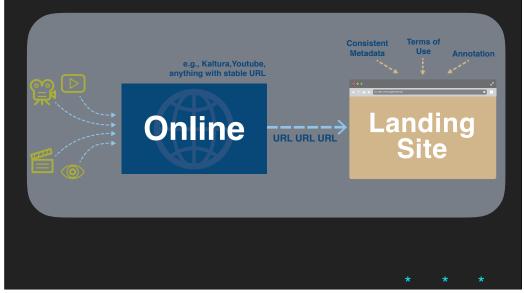




If we annotate the videos themselves - the signs that occur on camera - then that information becomes machine-readable. We can search it like we do text on Google. And we can do things like find out the most frequent words in ASL. This kind of information is interesting in so many ways.

lexical frequency					
THAT		1162	NOT		606
WHAT-PU		995	DEAF		603
LOOK		787	BUT		545
HAVE		754	KNOW	N.	458
LANGUAGE	Contraction of the second seco	628	REAL		321
(Out of 85,674 items from 942 .eafs)					
			*	* *	*

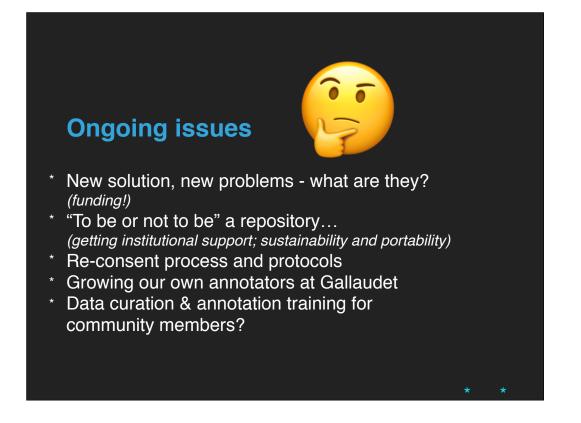
### GALLAUDET UNIVERSITY DOCUMENTATION OF ASL (GUDA)



So with the need to store and organize videos digitally, armed with ideas of language archives and the guiding vision of the Austin Principles of Data citation and ID glosses stored in the ASL Signbank - we can do something like what we've been starting to work on - Gallaudet University Documentation of ASL (GUDA)

See the graphic here in the slide - it shows our overall vision.

We're interested in finding all of those existing video collections and making sure they have a stable URL then pointing to them on a landing site. Once there, we'll have consistent metadata (done in similar format); terms of use; annotation to make the videos searchable.



These ongoing issues (with a "whither a corpus for ASL" larger point) come from of our discussion about the ongoing "devaluing" of our project.

Ongoing issues - funding (we need some); sustainability (this landing site is not an archives - it's not a long-term solution. It's an initial step in something that will take our entire careers to do); portability (we need to consider how to make sure this is accessible across platforms and over time. e.g., Bird and Simons 2003; re-consent (seeking permission from those on video to continue to share their data)



GUDA is then the maintenance of ASL stories if we can actually manage to combine the collection of ASL videos in a stable and accessible digital place with data archiving protocols. If this project could ever happen, it makes total sense to do that here at Gallaudet University, the only university in the world for deaf and hard of hearing people.

Berez-Kroeker, A., Gawne, L., Kung, S., Kelly, B., Heston, T., Holton, G., Pulsifer, P., Beaver, D., Chelliah, S., Dubinsky, S., Meier, R., Thieberger, N., Rice, K., & Woodbury, A., (2018). Reproducible research in linguistics: A position statement on data citation and attribution in our field. Linguistics, 56(1). 1-18.

Bird, Steven and Gary Simons. 2003. Seven dimensions of portability for language documentation and description. Language, 79(3). 557-582.

Crasborn, Onno and Han Sloetjes. 2008. Enhanced ELAN functionality for sign language corpora. In Construction and Exploitation of Sign Language Corpora. [Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages. 6th International Conference on Language Resources and Evaluation, LREC 2008, Marrakech.], edited by Onno Crasborn, Thomas Hanke, Eleni Efthimiou, Inge Zwitserlood and Ernst Thoutenhoofd. 39-43. ELRA, Paris. <u>http://</u> www.lrec-conf.org/proceedings/lrec2008/workshops/W25\_Proceedings.pdf

ELAN (Version 5.4) [Computer software]. (2018, April 04). Nijmegen: Max Planck Institute for Psycholinguistics. Retrieved from <u>https://tla.mpi.nl/tools/tla-tools/elan/</u>

Hochgesang, J.A., O. Crasborn, & D. Lillo-Martin. (2018). ASL Signbank. New Haven, CT: Haskins Lab, Yale University. <u>https://aslsignbank.haskins.yale.edu/</u>

Padden, C. (2004). "Translating Veditz" Sign Language Studies, 4(3), 244-260.

Sources of video available upon request (Translation of Veditz from (Padden, 2004)



**CONFERENCE HASHTAG #MIII**