Large-Scale Humanities Computing Projects: Snakes Eating Tails, or Every End is a New Beginning?

William A. Kretzschmar, Jr. <kretzsch_at_uga_dot_edu>, University of Georgia

Abstract

The word "finish" can mean two things that have quite different implications for large-scale humanities computing projects: "to bring to completion; to make or perform completely; to complete" and also "to perfect finally or in detail; to put the final and completing touches to (a thing)." The word "finish" is just not part of the deal for the Linguistic Atlas Project in either sense. However, granting agencies must ask "what do you want money for this time?" and, from this viewpoint, the Atlas Project consists of a series of particular tasks or experiments, each one of which is capable of being "finished" in both senses of the word. This paper discusses the reality of funding, deadlines, and deliverables, as they relate to the sequence of tasks that make up the larger Atlas Project. There are no once-and-done, permanent solutions. The largest humanities computing projects require continuing care and maintenance, and the best way forward is to create some sort of stable institutional setting for large projects that will provide continuity and baseline resources for the work.

One of the motivating questions for this cluster of essays is "What does it mean to finish something?" As it happens, the word "finish" can mean two things that have quite different implications for large-scale humanities computing projects. On one hand, according to the *OED*, "finish" can mean "To bring to completion; to make or perform completely; to complete." On the other hand, the word can also mean "To perfect finally or in detail; to put the final and completing touches to (a thing)." In my own work of this kind, the American Linguistic Atlas project (http://www.lap.uga.edu), we do neither of these things. We cannot come to an end of the work because we are witnesses and archivists of how Americans talk, and they keep talking differently across time and space. Neither do I think that our humanities-computing representation of our research is capable of being finally perfected, of achieving some perfect state, because technology keeps changing and the demands placed upon our research keep changing. If we view the entirety of the Linguistic Atlas Project as a "large-scale humanities computing project," the word "finish" is just not part of the deal. And we are not alone. While the creation of, say, a variorum edition may seem like a project that can be finished in both senses, actually we need to make new editions all the time, since our idea of how to make the best edition changes as trends in scholarship change, especially now in the digital age when new technical possibilities keep emerging.

[1]

However, it is quite reasonable to ask, as our granting agencies must ask, "What do you want money for this time?" or "Did you accomplish what we gave you money to do?" From this viewpoint, the Atlas Project, as an example to stand in for any large scale project, consists of a series of particular tasks or experiments, each one of which is capable of being "finished" in both senses of the word. It is these separate tasks that are fundable, and for which we can claim to have done what we said we would do. In this paper, I would like to discuss the sequence of tasks that make up the larger Atlas Project in order to show the special character of work done deliberately as part of a sequence for a large-scale project, as opposed to work proposed as a singular task. The contextualization of the separate tasks leads to special cases of what it means to "finish" the work in either sense. The point of what follows is not the Atlas Project itself, but instead the way that individual tasks respond to the technical and academic situation at the time. Our own technical work on the Atlas has responded to the microcomputer revolution, to the emergence of the Web, and to the development of text-encoding and computer multimedia. Our work has helped to drive changes in the academic study of language variation, from traditional dialect maps to use of rapid visualization methods and statistical processing for both

[2]

text and audio data. The scope of these changes show how our work and thinking over the years have had to change and must continue to change, just as they must for other large-scale projects like major editions and dictionaries, so that we can avoid the charge of being the snake that eats its own tail.

The Linguistic Atlas Project, per se, has been notable over many years for its twin goals of interactivity for research (including the use of GIS) and making its data sets accessible and available to other researchers and to the public. I first programmed a GIS system for our Linguistic Atlas data in 1990, presented at the ACH/ALLC meeting in Tucson in 1991 Figure 1; [Kirk and Kretzschmar 1991]; [Kirk and Kretzschmar 1992]; [Kretzschmar 1992]. The program, called LAMSASplot, took advantage of work we did with funding from NEH to keyboard words and phrases from our survey interviews so that they would be available for computer-assisted analysis. The LAMSASplot system was widely used for teaching and research on American English in the early 1990s, and it immediately led to breakthroughs in how we were able to think about language variation data [Kretzschmar 1994], [Kretzschmar 1996a]; [Lee and Kretzschmar 1993]. The two most important design elements of that first system were the central column, which shows the frequency of occurrence of a word of interest, both by speaker and by community (most communities have more than one speaker), and the rightmost column, the GIS implementation. That column is composed of two layers, a standard base map layer that shows the state outlines, and a second layer consisting of plotted points at community locations that is generated according to the evidence for the word selected.
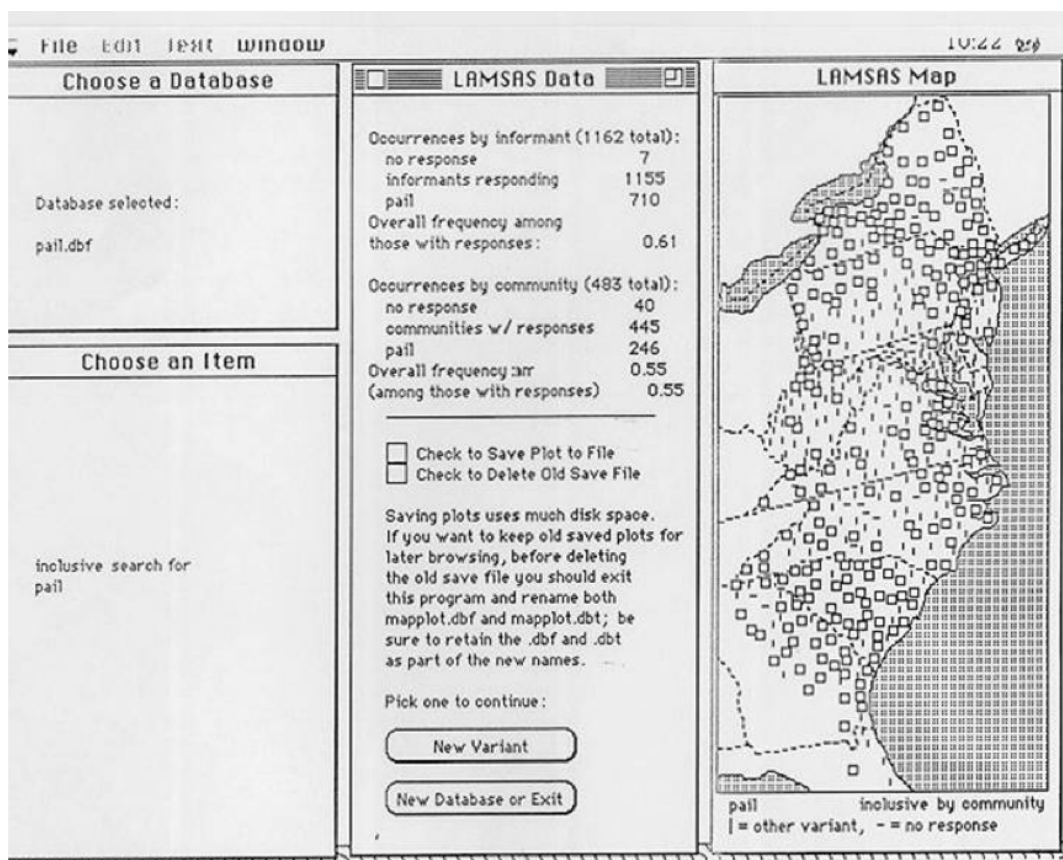

Figure 1. LAMSASplot on the Mac platform, c. 1990

The LAMSAS GIS programs were prepared on Macintosh computers using the Foxbase relational database package. Foxbase permitted us to associate graphical coordinates with our linguistic data within a relational structure. Generation of these plots took only 90 seconds on the desktop computers of the day, at least a 250-fold improvement over the hours required for the charting that had to be done by hand by the pioneers in the field, Hans Kurath and Raven McDavid. The other advantage of LAMSASplot is that it made frequency counts and plotted one variant at a time. This practice was in sharp contrast to the method used by Kurath and McDavid for charting language data, using isoglosses, lines which showed the limit of occurrence of particular variants. As our pail map suggests here, lines just cannot do

justice to the geographical distributions of linguistic features. Our GIS solution thus launched a new line of analysis for dialectologists based on differential frequency in feature distributions. The LAMSASplot humanities computing application not only used available technology to automate mapmaking, it provided a theoretical advance in how the data was analyzed. The Atlas project was thus one of many involved in the burgeoning field of computer-assisted scientific visualization, not just in the humanities, but the particular point of talking about the Atlas is to show how such large trends have particular effects in particular areas of study.

We did not finish keyboarding all of our LAMSAS data with our NEH funding, which had been proposed as the first of a sequence of grants. We could show immediate benefits of getting the process started, in part through LAMSASplot, but unfortunately NEH did not fund our proposals to continue digitization of the data. There were various reasons offered by the panelists over the years, but I suspect that the refusals came down to the fact that the panelists preferred to fund new work over our historical interviews. The immediate analytical task for the Atlas was thus "finished" in both senses, both completed and refined, but only the immediate task; we were shut out from long-term funding to "finish" the whole project.

5

Figure 2. Atlas Web Site, Mark 1

As the next task in our humanities computing implementation, we then ported the system to the Web, which I first demonstrated in 1996 ([Kretzschmar 1996b]; Figure 2 shows the site as of 1998). We had been working on an interactive ftp/gopher system as early as 1993, but when Web technology became available we saw that it enabled perfectly what we had been attempting from another direction. The Web allowed us to make our data available to a wide public, with many additional interactive GIS features for locating speakers and information for the LAMSAS survey, and with static pages for the other regional surveys. We added additional data for our African American LAMSAS speakers plus data from speakers of Sea Island Creole (Gullah) in 1998 with funding from NSF. The GIS maps from LAMSASplot were ported to the Web, as in Figure 3 for the phrase *quarter of* in telling time (as opposed to *quarter till*, *quarter to*), where they now could be created by users almost instantaneously with a server-side script, another improvement in processing time of perhaps fiftyfold. We also offered maps to show where all of the speakers in a state resided (Figure 4). These maps were clickable to allow users to access all of the information about any given speaker and the speaker's survey responses (Figure 5). Besides maps, we enabled browsing and searches of the responses and of speaker information (Figure 6, Figure 7). We also changed the way that we stored our data: flat, comma-delimited files instead of a proprietary relational format.
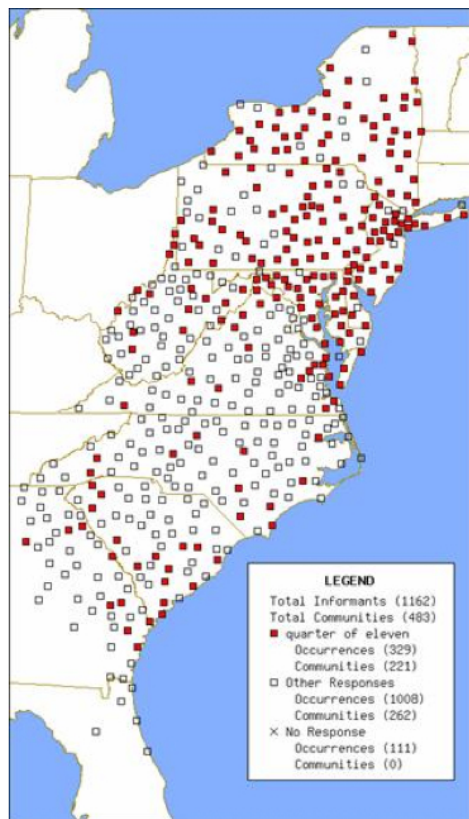
6

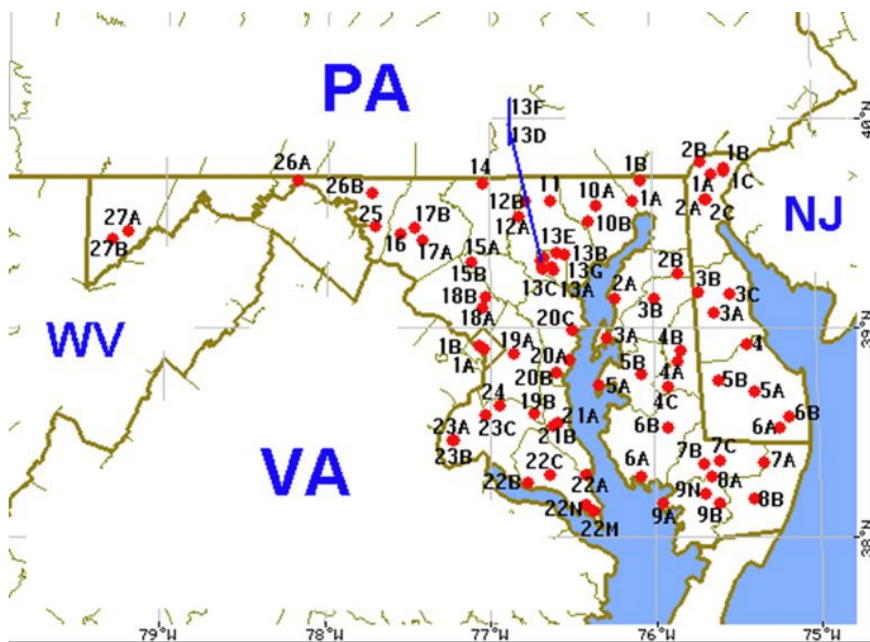**Figure 3.** LAMSAS Web: *quarter of*

Figure 4. LAMSAS Web: state map of speakers in Maryland

LAMSAS

Home

Information

Browse

Search

Maps

Browse

Informant data for record 543:
76-year-old White Female, type Cultivated

Serial Number:	543
Community Number:	563 (Baltimore Co., MD)
Informant ID Number:	MD13E!
Old Informant Number:	58
Auxiliary Informants:	N
Field Worker:	Guy Lowman
Work Sheet:	Preliminary South Atlantic
Year Interviewed:	1934
Type:	III (Cultivated)
Generation:	A (Aged)
Cultivation:	Y
Sex:	Female
Age (when interviewed):	76
Education:	High school: 12 years
Occupation:	Keeping House
Race:	White
Locality Type:	Urban

Figure 5. LAMSAS Web: Baltimore speaker (from clicking MD13E

LAMSAS

Home

LAMSAS

Information

Browse

Search

Maps

Browse

These are links to the raw data files. (*Warning: some of the files, especially with the fancy formatting, are quite large.*) There are two formats for the files, plain and fancy one.

Helpful links:

- Samples of the plain and full formats
- Explanation of the fields

Informant Database:

- ○ Plain format (81.2 K)
- ○ Fancy format (671 K)

Linguistic Databases:

[Quarter to eleven ▾]

- ○ Plain format
- ○ Fancy format

[Submit Query] [Reset]

Figure 6. LAMSAS Web: Browse Screen

LAMSAS Informant Database Plain Format

Links:

- Explanation of the fields
- Search the informants
- Examine the raw data file
- Back to Browse

Inf#	Typ	Community	Local	WS	Year	Sex	Age	Race	Gen	Cult	Educ	Occup
NY1	I	Suffolk Co	R	E	1933	M	53	W	B	N	3	F
NY2A	I	Suffolk Co	R	E	1933	M	63	W	A	N	0	F
NY2B	I	Suffolk Co	R	E	1933	M	85	W	A	N	3	F
NY2C	II	Suffolk Co	U	E	1933	F	80	W	A	N	2	K
NY3A	I	Suffolk Co	R	M	1941	M	71	W	A	N	4	F
NY3B	I	Suffolk Co	R	M	1941	M	70	W	A	N	3	F
NY3C	I	Suffolk Co	R	M	1941	M	50	W	B	N	4	F
NY3D	I	Suffolk Co	R	M	1941	M	48	W	B	N	3	F
NY4A	I	Nassau Co.	R	M	1941	M	69	W	A	N	2	F
NY4B	I	Nassau Co.	U	M	1941	M	79	W	A	N	2	O
NY4C	I	Nassau Co.	R	M	1941	M	51	W	B	N	3	F
NY5A	I	Brooklyn	U	M	1941	F	70	W	B	N	4	K
NY5B	I	Brooklyn	U	M	1941	M	62	W	B	N	4	R
NY5C	II	Brooklyn	U	M	1941	F	51	W	B	N	3	K
NY5D	II	Brooklyn	U	M	1941	M	64	W	B	N	2	R
NY5E	II	Brooklyn	U	M	1941	F	82	W	A	N	2	K
NY5F	II	Brooklyn	U	M	1941	M	53	W	B	N	2	M
NY5G	II	Brooklyn	U	M	1941	M	50	W	B	N	4	M
NY5H!	III	Brooklyn	U	M	1941	M	68	W	A	Y	6	P
NY5I!	III	Brooklyn	U	M	1941	M	54	W	B	Y	6	M
NY6A	I	Richmond C	U	M	1941	F	76	W	A	N	2	K
NY6B	II	Richmond C	U	M	1941	F	48	W	B	N	4	K
NY7A	I	Manhattan	U	M	1941	M	45	W	B	N	2	L
NY7B	I	Manhattan	U	M	1941	M	39	W	B	N	2	O

Figure 7. LAMSAS Web: Speakers (click of Figure 6, plain format informants

The Atlas Web was a significant advance for both teaching and research ([Kretzschmar 1997], [Kretzschmar 2002a]), in line with the goals for an electronic atlas first set forth nearly a decade earlier [Kretzschmar 1988]. The plotted maps and much of the data were the same as what was available for the Mac LAMSASplot package — besides the African American data, we had keyboarded some additional files, a few at a time as resources permit, as we are still doing — but instead of an incremental improvement, the Atlas Web site was actually another sea change for the project. For the first time the whole Atlas project was represented together, not just as separate regional parts, and for the first time we could offer real public access to as much data as we could digitize. Again, the example of the Atlas just shows the particular significance in our area of our participation in the main international trend for Web representation of information.

These were good things, revolutionary in their way, but the fact is that we *had* to do something: the world had changed from mainframes and desktops to UNIX servers. We had actually been using both mainframe processing (for intensive statistics jobs) as well as the desktop Macs and PCs. If we had not acted, the proprietary software that we were using would have gone out of use (FoxBase was sold, and the database for our PCs, RBase, also disappeared), and so we would have lost our investment in preparation of proprietary data files. We would have missed the Web revolution. So, not to change would have meant that the Atlas project would be stuck on paper, where it had been before we developed any computer applications at all. We did benefit from the NSF grant I mentioned, and from a small internal University of Georgia grant that allowed us to set up our own server, both limited tasks that we could propose as achievable within a short time and for which we could show results. We still could not get funding to do the whole project, and if we had waited for that we would still be waiting. Every large humanities computing project faced similar problems: adapt, or become a footnote in academic history.
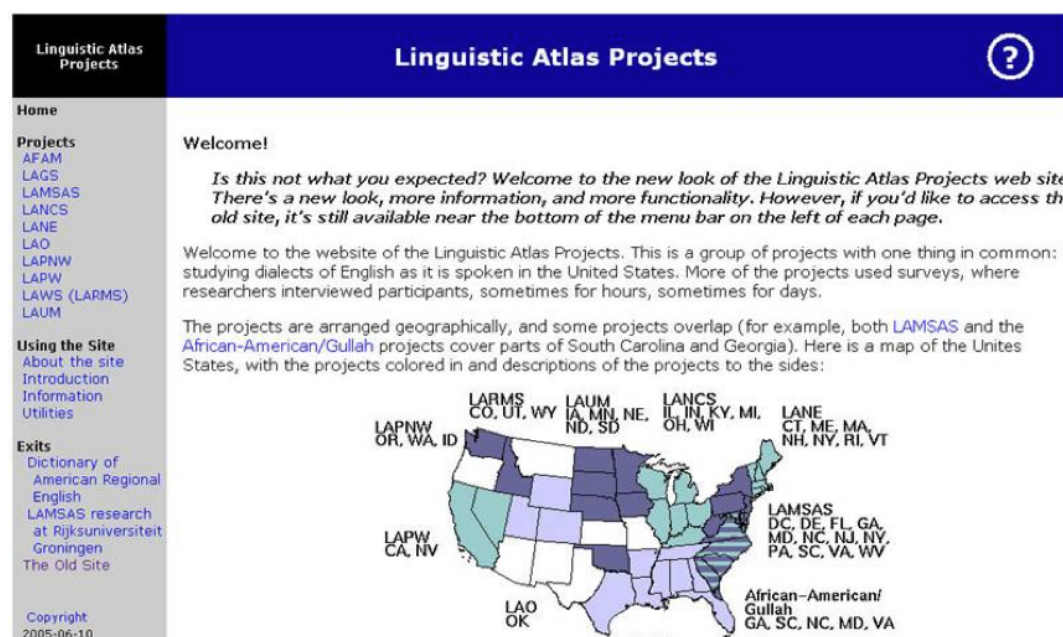
Figure 8. LAMSAS Web, Mark 2, c. 2005

After a while, we wanted to do new things, so we began work on a major revision of the Web site that came on line in 2003 (Figure 8, shown as of 2005). Our earlier GIS visualizations were just not as flexible as they needed to be to satisfy the demands of sociolinguistics, which had come to expect association of linguistic features with particular social variables (e.g. sex, age, social status). We kept the interactive plotted maps and layered GIS access to information, but added even more interactive choices such as more flexible searches and tallies of the speakers and language data set. Browse screens are now composed on the fly from separate data files, as in Figure 9 from the quarter of data and speaker characteristics, so that the linguistic information is immediately associated with the social characteristics of those who said it. For the first time we enabled searching by pronunciation features, not just words or strings, and we offered the ability to do sub-searches in sequence so that users could zero-in on features and groups of interest (Figure 10). In order to accomplish these goals, we again changed our manner of storing and manipulating data from a homemade phonetic font to Unicode, and from flat files and scripted regular expressions to a MySQL implementation. However, we kept "The Old Site" as a link on the new one, so that long-time users would find what was familiar to them, and also for users who did not want the greater complexity that came with greater flexibility of use (see the link on the bottom left corner of Figure 8). We could not just move it, however, because "The Old Site" had to be compatible with Unicode and with the extensive Python scripting that ran functions on the New Site. The task of importing the Mac-based GIS system to the Web was complete by 1996, but was finished, in that second sense, in 2003 with the platform change and the new touches of the more flexible site. Again, other large-scale humanities computing projects faced similar problems; we all had to adapt to new scripting languages and environments, new tools like MySQL and Unicode, and new scholarly trends and demands in our particular areas.

9

The transition to the Web and the further refinement of our Web tools in the late 1990s and early years of the new century were not accomplished by means of external funding from federal sources. We did win awards to record the African American data and to conduct new kinds of research (e.g., NSF awards to explore neural-network analysis of Atlas data, and to conduct a pilot project for random-sample survey research in Atlanta), and some of these resources were applied to computer work, but no award specifically funded our Web innovations. Development of our Web site was primarily accomplished by two graduate assistants, Rafal Konopka and Eric Rochester, who worked on Atlas computer projects for twelve years between them, first Konopka through our initial Web implementation, then Rochester through the first major revision. They held graduate assistantships, sometimes fully funded for research but more often cobbled together from different sources including both research and teaching. The consistent, stable element in this funding came from the Hans Kurath Fund of the American Dialect Society. The Kurath Fund is an endowment for the Atlas created by Raven McDavid and maintained by the American Dialect Society [Kretzschmar 2003]. The Dialect

10

Society hold title to the Atlas research materials, and its Executive Council ratifies editors and advisory board members for the project. While the Kurath Fund could not support a complete graduate assistantship, it could pay enough so that other funding could round out a position. Similarly, while the Kurath Fund could not pay all of the operational expenses of the project, it could provide a key piece of equipment from time to time, repairs, or specialized supplies. The University of Georgia had agreed to provide space for the project and some operational support, in conjunction with the author's faculty appointment (a more permanent agreement may in time be possible). Thus the funding that allowed continuous development of the Atlas Web site has come from multiple sources, and still does. We are always in the position of putting together the pieces of the funding structure year by year. The central role of the American Dialect Society, however, provides the essential stability that has made continuous operation of the editorial site possible.

Figure 9. LAMSAS Web 2: Browse Screen

Figure 10. LAMSAS Web: Sub-Search

Figure 11. LAMSAS Web: Work by Others

The additional changes for the LAMSAS Web Mark 2 may seem more incremental than revolutionary, but again we feel that they responded to a changing environment that could not be ignored, new scholarly imperatives. For the first time our linguistic survey data was now fully available for sociolinguistic research. Dialectology is a venerable pursuit in linguistics, but it had been overtaken by sociolinguistics in recent decades and declared essentially to be irrelevant by many sociolinguists. The changes that we made to our Web site integrated social and geographical analysis of language variation in a single visualization that, in itself, showed graphically that the two approaches were not only compatible but inseparable. In addition, we also began to link our site with other sites, so that the Atlas Project could be seen in its connections with other online resources. Finally, we began to post completed analyses on the site in addition to raw data and associated information; some of these are results from our own funded research, others, such as those from John Nerbonne in Groningen (Figure 11), are from the research of people who used our freely-available data. In this way we could make and defend the claim that our Atlas survey research was a member of a wider community of those studying language variation. We did respond to emerging computational opportunities in the New Site, like Unicode and Python and MySQL, but the most significant part of the new changes spoke to issues of linguistic theory and analysis. And again, nobody funded completion of digitization of all of the data from the Atlas.

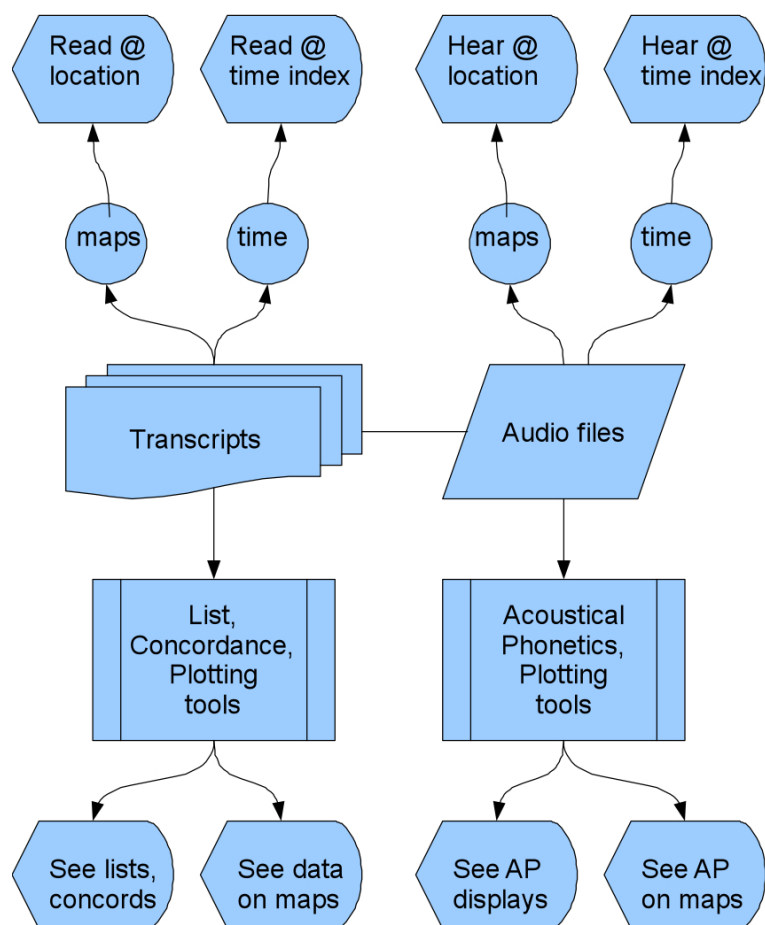# Information Flowchart for Linguistic Interviews

Figure 12. Future LAMSAS Web

Still, the larger Atlas Project is nowhere near at an end. We are now rethinking what the site should do, from a text-based system to one that features audio and stored images along with text and GIS. This change has become possible only in the last two years, as much greater network-attached storage has become available (measured in Tb, before long Pb). Because of our archival audio files, we are now one of the largest clients at the University of Georgia institutional storage array (which we share with bio-informaticists, physicists, and others usually considered to be power users). We now conceive of our new interviews as conversational corpora, in which text transcriptions serve as time-linked indices to audio files [Kretzschmar 2005a]; [Kretzschmar 2006]. Figure 12 charts the flow of information that we envision for the next-generation site. While many users will want to listen to our speakers, others will want to perform acoustical analyses, now a strong trend in language variation research, as we ourselves now perform them (e.g., [Kretzschmar 2004]; [Kretzschmar 2005b]). Our next task is to integrate sound with text and to enable acoustical research functions, while maintaining our interactive GIS functions — a whole new set of tools and problems from the previous task [Kretzschmar 2002b]. As the flowchart suggests, we now envision eight different outputs from a new Web presence: full text transcripts, linked sound, acoustical phonetic data in lists and plots, tallies of feature variants in lists and on maps, and technical statistical results in both lists and maps. While any single task from this list is possible, some will be difficult to automate, and the hardest thing of all will be to integrate all eight outputs. We have not yet made these changes; they have been designed for some time now, and await the right people and circumstances and resources to make them real (our major revision cycle seems to be 5 to 10 years). When we can do that, we will have integrated our existing modes of study for language variation, surveys and sociolinguistics, with corpus linguistics, the latest major contemporary approach to language variation. We will have provided not just words and phrases extracted from interviews, but a full complement of information in our interviews, from sound recording to transcripts to analyses of

acoustical characteristics.

So, will we ever be "finished"? Yes, with the GIS of LAMSASplot and of "The Old Site," though still (as always) tweaking the current site. We have plans to make yet another thorough revision. But we will never be finished either with keeping older sites and available or with creating new visualizations for the information we keep, as new technical possibilities and research demands appear. We can complete particular tasks, and often we can even "finish" particular tasks in the sense of polishing them for improved use. Yet one research proposal does not make the whole research program. While we can often fund and succeed with individual tasks, we must always see tasks as part of the larger process that probably will not be funded completely but still must continue for future users. We would be the snake eating its own tail if all we did was keep polishing eternally a single task (or worse, the dog that pointlessly chases his own tail), but we do well to make every end a new beginning as new technological possibilities become available, and new theories and styles of analyses take hold.

13

After twenty-five years of trying to apply humanities computing to the problems of the Linguistic Atlas, it has become clear that there are no once-and-done, permanent solutions. The largest humanities computing projects are likely to require continuing care and maintenance, if not more radical representation and reinterpretation in light of the advance of scholarship, and yet they seem unlikely ever to be funded comprehensively for these tasks. The best way forward is to create some sort of stable institutional setting for large projects that will provide continuity and baseline resources for the work. This we have done for the Linguistic Atlas, through the American Dialect Society and our association with the University of Georgia. Stable institutional settings allow for additional resources to be sought for individual tasks, as the need for each one becomes apparent. In my general area, the Dictionary of American Regional English project at Wisconsin and the Dictionary of Old English project at Toronto, both innovative users of humanities computing, have found long-term funding through NEH and have also had strong institutional support. We can hope that Wisconsin and Toronto will not end their association with these projects when their first editions are completed. A good model for what might happen has occurred at Michigan, where the completion of the (print) *Middle English Dictionary* was followed by the creation of the electronic *Middle English Compendium* [McSparran 2000] in conjunction with the University of Michigan Library. Conversion was assisted by a grant from NEH. We are now attempting something of the kind at Georgia, with our new grant from NEH which will allow us to make digital conversions of the existing audio-taped interviews from the Linguistic Atlas, and which will also help us to make a new public audio archive of these materials.

14

Thus, at the end of the day, I recommend that, besides grants for new tasks and new work, granting agencies should consider assisting in the formation of enduring institutional structures to support large-scale projects in humanities computing. Some existing programs, like NEH Challenge Grants, may already be helpful for this purpose. The establishment of common interests between the library community and researchers in humanities computing, already well started in the ACH and ALLC and exemplified by projects like the *Middle English Compendium*, is an excellent trend to continue. We will probably never be able to fund big projects like mine at the level needed to finish them completely. But that is OK, because the idea of finishing big humanities computing projects once and for all is just an illusion. We will always need to address changes in technology and changes in our disciplines, even for subject materials that are historical and unchanging. We need, therefore, to accept the rule for Talmudic scholarship that my colleague Virginia McDavid once passed along to me: "You do not have to finish the work, but neither may you desist from it." We just have to keep at it, and find the resources we need along the way.

15

Works Cited

Kirk and Kretzschmar 1991 Kirk, John, and William A. Kretzschmar Jr. "The Analysis and Interpretation of Dialect Databases by Interactive Mapping". Presented at *ACH/ALLC 1991*. *Proceedings of the Joint International Conference of the Association for Computers and the Humanities and the Association for Literary and Linguistics Computing* (1991).

Kirk and Kretzschmar 1992 Kirk, John, and William A. Kretzschmar Jr. "Interactive Linguistic Mapping of Dialect Features". *Literary and Linguistic Computing* 7 (1992), pp. 168-175.

Kretzschmar 1988 Kretzschmar Jr., William A. "Computers and the American Linguistic Atlas". In Alan Thomas, ed., *Methods in Dialectology: Proceedings of the Sixth International Conference on Methods in Dialectology*. Clevedon: Multilingual Matters, 1988. pp. 200-224.

Kretzschmar 1992 Kretzschmar Jr., William A. "Interactive Computer Mapping for the Linguistic Atlas of the Middle and South Atlantic States (LAMSAS)". In Joan Houston Hall Nick Doane, Dick Ringler and Frederic Gomes Cassidy, eds., *Old English and New: Essays in Language and Linguistics in Honor of Frederic G. Cassidy*. New York: Garland, 1992. pp. 400-414.

Kretzschmar 1994 Kretzschmar Jr., William A. "Linguistic Theory and Computer Modeling of Linguistic Survey Data". Presented at *ACH/ALLC 1994*. *Proceedings of the Joint International Conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing* (1994).

Kretzschmar 1996a Kretzschmar Jr., William A. "Quantitative Areal Analysis of Dialect Features". *Language Variation and Change* 8 (1996), pp. 13-39.

Kretzschmar 1996b KretzschmarWilliam A., Jr. "The LAMSAS Internet Site". Presented at *NWAVE* (1996).

Kretzschmar 1997 Kretzschmar Jr., William A. "Computer-Assisted Study of American English Lexical Data". In Udo Fries Viviane Müller and Peter Schneider, eds., *From AElfric to the New York Times: Studies in English Corpus Linguistics*. Amsterdam and Atlanta: Rodopi, 1997. pp. 239-247.

Kretzschmar 2002a Kretzschmar Jr., William A. "Teaching American English Online". *Journal of English Linguistics* 30 (2002), pp. 318-327.

Kretzschmar 2002b KretzschmarWilliam A., Jr. "TEI and Linguistic Interviews". Presented at *TEI Annual Conference*, sponsored by Text Encoding Initiative Consortium (2002). http://www.tei-c.org/Membership/Meetings/2002/kretzschmar.pdf.

Kretzschmar 2003 Kretzschmar Jr., William A. "Linguistic Atlases of the US and Canada". *Publications of the American Dialect Society* 88 (2003), pp. 25-48.

Kretzschmar 2004 Kretzschmar William A., Jr., Sonja Lanehart, Betsy Barry, Iyabo Osiapem and MiRan Kim. "Atlanta in Black and White: A New Random Sample of Urban Speech". Presented at *NWAVE* (2004).

Kretzschmar 2005a KretzschmarWilliam A. Jr., Betsy Barry and Nicole Kong. "Publication of Full Interviews from the Atlanta Survey Project". Presented at *ADS/LSA* (2005).

Kretzschmar 2005b KretzschmarWilliam A. Jr., MiRan Kim and Nicole Kong. " Vowel Formant Characteristics from the Atlanta Survey Project". Presented at *ADS/LSA* (2005).

Kretzschmar 2006 Kretzschmar Jr., William A., Jean Anderson, Joan Beal, Karen Corrigan, Lisa-Lena Opas-Hanninen and Bartek Plichta. "Collaboration on Corpora for Regional and Social Analysis". *Journal of English Linguistics* 34 (2006), pp. 172-205.

Lee and Kretzschmar 1993 Lee, Jay, and William A. Kretzschmar Jr. "Spatial Analysis of Linguistic Data with GIS Functions". *International Journal of Geographic Information Systems* 7 (1993), pp. 541-560.

McSparran 2000 McSparran, Frances, et al., ed. *Middle English Compendium*. National Endowment for the Humanities, 2000. http://quod.lib.umich.edu/m/mec/.