Bringing Digital Data Management Training into Methods Courses for Anthropology

Linguistic Anthropology:
Principles and Practices of Digital Data Management

Arienne M. Dwyer

2016

Bringing Digital Data Management Training into Methods Courses for Anthropology
is a set of five modules:

General Principles and Practices of Digital Data Management
Archaeology: Principles and Practices of Digital Data Management
Biological Anthropology: Principles and Practices of Digital Data Management
Cultural Anthropology: Principles and Practices of Digital Data Management
Linguistic Anthropology: Principles and Practices of Digital Data Management

Organization

I.   Review of material from "General principles and practices" module
II. Key data management practices
III. Ensuring the future of your data and avoiding catastrophe
IV. The basics: Working with data
V.   What are our responsibilities?
VI. Archiving and re-use of data
VII. Exercises
VIII. Instructor notes
IX. References
X.   Additional resources
XI. Acknowledgments

Review of material from "General principles and practices" module

- What are data?
- What is data management?
- What are the advantages of making data accessible?
- What are ethical dimensions of data management?
- What is a data management plan?

Key data management practices

Key data management practices can be summarized in four points:

1. Data should be put into an enduring format.
2. Data should be discoverable via metadata.
3. Data should be archived.
4. Data gathering, archiving, and dissemination should be fully consultative and done with the permission of involved participants.

Ensuring the future of your data

Workflow: The data lifecycle

- Data in, data out: I/O format; planning for output at the time of input

Native Aymara Language Domain

By Haylli. https://commons.wikimedia.org/w/index.php?curid=4949323

What constitutes data?

- *In situ* ("field") work or other research creates input (recordings, notes, maps, images)
- Analysis and archiving creates output (transcriptions, translations, annotations, lessons, articles, websites, books)
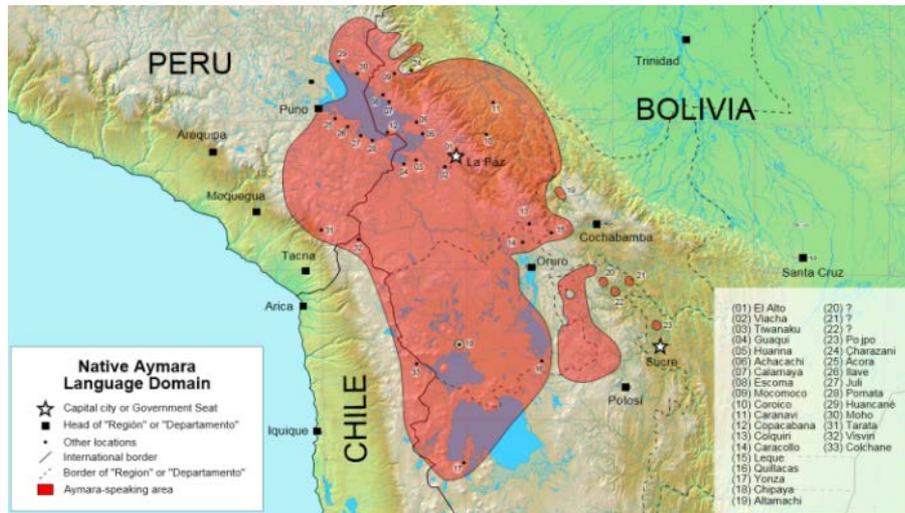
[Optional in-class exercise: Discussion of data workflow]

Ensuring the future of your data

Archival, working, and presentation data formats: a useful distinction (Simons 2006)

- Archival: highest-quality, lossless (uncompressed), non-proprietary, structured
- Working: whatever format and size that allows for data analysis
- Presentation:
  - Optimized for sharing (e.g., web or print)
  - Compressed audio-visual (A/V) and highly formatted text are usually preferable.
- The Library of Congress has prepared [Sustainable format](#) guidelines

Ensuring the future of your data

Ethical and legal considerations (overview)

- Planning data management entails negotiation with research subjects and/or a community, about:
- Informed consent for the collecting, archiving, and sharing of data
- Access to and availability of data
- Issues of intellectual property.

Common data disasters include

- Losing the only copy of your data.
- Storing data in an obsolete format.

Ensuring the future of your data:
Data management plans

Data management plans (DMPs) are closely linked to the data lifecycle through
- Concerns about sharing and protecting data and participants.
- Motivations:
  - Funding requirement
  - Dissertation/publication
  - Better analysis
  - Data reuse
- The anthropologist's responsibilities: sharing maximally and ethically.

(On DMPs, see Linguistic Data Consortium 1992-2016)

[Optional in-class exercise: Discuss data management plans]
[Outside-class exercise: Draft a data management plan]

Ensuring the future of your data
Data management plans

What's in a DMP?

- Ethics, including research design, data gathering, collaboration, sharing, attribution, and privacy
- Data, raw and processed:
  - What constitutes the data and its required accompanying metadata?
- Metadata and project documentation
- Data formats and needed tools:
  - In what formats will the data and relevant tools be archived?
- Data sustainability and archiving plan:
  - How exactly will the materials be archived and preserved?
- A named archive:
  - At which archive will the materials be hosted?

Ensuring the future of your data
Data management plans

- Principles of access, attribution, and privacy for collaboration and sharing, including participants' roles and responsibilities of participants, and credit.
- Production, access, control, and sharing of research products and their re-use:
  - Relevant intellectual property rights (IPR), both nation-states' and indigenous perspectives
  - Local, national and international laws

From left : Bill Toledo, Robert Walley and Alfred Newman, World War II Navajo code talker veterans, April 12, 2013, at Luke Air Force Base exchange

By U.S. Air Force photo/Airman 1st Class Grace Lee.
https://commons.wikimedia.org/w/index.php?curid=25858057

The basics: Working with data
Data types in linguistic anthropology

Linguistic anthropological methods
- reflect the interdisciplinary nature of the sub-discipline
- overlap with qualitative and quantitative methods for
  - linguistics, such as documentary linguistics, sociolinguistics, discourse analysis, cognitive tasks)
  - cultural anthropology (participant observation, interviewing, surveying, and reflexivity).

Linguistic anthropologists are likely to
- work with human subjects.
- generate data that include any or all of the following:
  - audio and/or video (A/V) recordings and transcriptions of them (often with translation and grammatical annotation, which is sometimes time-aligned); notes, sketches, images (photographs, maps (georectified or not), and diagrams), spatial data, artifacts and other physical data; websites, blogs, emails or other Internet-based communication; ultrasound, and MRI; word lists, grammatical paradigms, sentences, grammaticality judgments of them, and texts; the texts may include printed, handwritten, or electronic texts, questionnaires, surveys, and inscriptions, as well as metadata about these primary research data.

The basics: Working with data
Digital and non-digital data

Linguistic anthropological data may be digital or non-digital, structured or unstructured.

Born-digital data: Include "field" recordings, images, geolocations, experimental data

- Require associated metadata to understand the scope and organization of the data
- May be quite large and require external storage planning (e.g., hard drive or Internet-based)
- Archiving may require conversion to open formats.
- Require systematic file-naming, organizing, versioning, and backup

The basics: Working with data

Non-digital data

Non-digital data include print, an extant collection or archive, and pre-existing dictionaries and grammars

- May require hundreds of hours to digitize and structure
- Processes may be different from born-digital data:
  - Creating a dictionary from a digital wordlist only requires structuring and possibly encoding conversion.
  - Digitizing a print dictionary requires either scanning and optical character recognition (OCR) or keyboarding, and then structuring.
- Archival data may also require conversion to open formats.
- Also require systematic file-naming, organizing, versioning, and backup
  - Always retain an original, unchanged version.

The basics: Working with data
Metadata

"Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource" (NISO 2004).

- Like a bibliographic entry (book citation), but for your data
- Provides the necessary context about your data, allowing access and retrieval
- Types of metadata
  - descriptive vs. structural
  - administrative
  - rights management
  - preservation
- Which elements and how much is enough are your decision
- One metadata standard is proposed by the Open Language Archives Community (OLAC 2008).

Data quality
- Data quality: lossless (or less lossy), documented with metadata.
- Data quality requirements do vary depending on research purpose.
  - Phonetic analysis will generally require higher-quality audio than syntactic work.
- Archive the highest possible data quality, regardless of your topic or discipline.

[In-class exercise: Discussion of data]

The basics: Working with data
Data capture, regularization, and organization

- Digitization: representing analog data (e.g., sound, activity, object) in a digital format.
- Regularization: systematically replacing irregular forms with regular ones, in order to make certain data comparable.
  - A version of the non-regularized data is retained.
  - Some transcriptions of conversations might be done in the International Phonetic Alphabet (IPA).
  - Other transcriptions might be done in multiple practical orthographies.
  - To make different transcriptions usable together, one system is regularized to the other.

IPA chart-vowels translated into Basque

By Gorkaazk
https://commons.wikimedia.org/w/index.php?curid=25390004
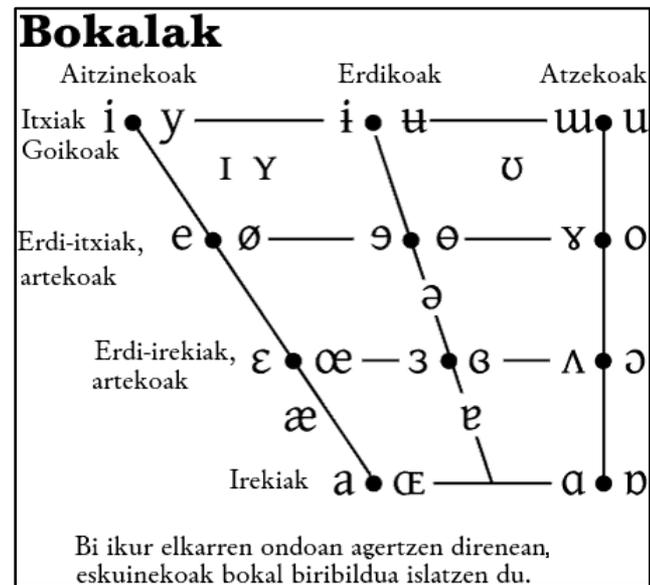
The basics: Working with data
Conventionalization

Communities share certain linguistic conventions, including orthographic and disciplinary.

- It is conventional for discourse analysts to transcribe speech in the language's official orthography, adding conventionalized speech annotation marking (e.g., Santa Barbara system, Dubois 2006).
- Phoneticians or documentary linguistic anthropologists, by contrast, often conventionally transcribe in a phonetic system (e.g., IPA).
- Those interested in grammar would add interlinear grammatical annotation, conventionalized by the so-called Leipzig Glossing Rules (2015).
- Ethnographic convention often encourages the identification of key words.

The basics: Working with data
Data standards and conventions

- Data format
- Character format
- File naming
- File structure

Organize data and metadata

Document the above activities and conventions (See Digital Preservation 101 [2012]; Case Study [n.d.])

Well-formed data are enduring data.

- Distinguishing the form and content of data
  - Well-formedness: durable and reusable
  - Interoperable: document formats and data structures
  - Keep data in non-proprietary (open) formats
  - Proprietary (closed) formats: commercial

The basics: Working with data
Tools (software)

- At start of any project, get the necessary tools that allow you to collect, organize and work with digital data.
- Open-source tools that allow maximal re-use are preferable.
- How tools fits into the workflow: proprietary vs. open-source tools:
  - Transcription tools
    - For discourse and conversation analysis (DA/CA): no special software is needed; one option is EXMARaLDA.
    - For time-alignment: Praat (originally for phonetic analysis), Transcriber
  - Multi-level annotation tools for A/V (including time alignment): ANVIL, ELAN
  - Producing Interlinear Glossed Text
    - Minimally: transcription, grammatical glossing, and translation

The basics: Working with data
Tools (software)

Digital Humanities tools of use to linguistic anthropologists

- To start out, web-based interfaces are easy.
  - Text analysis tools such as TAPoRware allow concordancing and frequency counts.
  - Online visualization tools such as Vidi are used to map or graph data.
- For more functionality and customization, locally-installed tools (which may be command-line or GUI) are useful.
  - R and R Studio for both quantitative research and visualization.
  - Gephi for network visualization.

What are our responsibilities?

Ethical and possibly legal obligation: Consent, whether oral, written, or both; and attribution of value

- Informed consent:
  - Participation in research is voluntary.
  - Research participants must be fully aware of
    - the purpose of the research.
    - how the data will be gathered.
    - how their privacy (and/or recognition) will be handled.
    - how the data will be shared.
- Requesting written permission from "a community" as an IRB requires often sows mistrust (Dwyer 2006, van Driem 2016)
  - Attribution of value means recognizing Indigenous Knowledge, Traditional Knowledge, or Traditional Ecological Knowledge of indigenous and local peoples (UNESCO n.d.; IPinCH 2016; Traditional Ecological Knowledge 2016).

What are our responsibilities?

Ethical and moral obligations

Sharing of data and research results are obligatory in two contexts:

- For community-based research, respecting community norms for data access.
  - Open access mantra: information wants to be free (Stewart Brand)
  - The community, not just the researcher, determines access, ownership, and sharing
- "Giving back" to the individuals and community in which the data were collected:
  - In a format, in a language, and with content that can be used locally, while attending to privacy concerns of participants
  - Sharing with the public research results and at least some of the data, unless restricted.

[Optional in-class exercise: Discussion of access]

What are our responsibilities?

Legal obligations: Numerous committees and organizations review and monitor projects involving "human subjects."

- University IRB, institutional review board, independent ethics committee (IEC), or other committees at most North American institutions
- Separately, approval of other IRBs may be needed, as by
  - School boards.
  - Tribal IRBs.

Moral obligations, beyond what the IRB requires:

- Do no (even unintentional) harm
- Create a research product that is useful to the native-speaker participants.

[Optional in-class exercise: Discussion of obligations]

What are our responsibilities?

Rights

- Intellectual Property Rights (IPR)
  - Ownership of "creations of the mind" (Newman 2007, Levine 2016)
  - Based on Western notion of ownership, but can include TEK
  - Copyright: who owns and can distribute a particular work (Copyright FAQ n.d.)
- Recognition: All participants have the right to receive and credit for their contributions.
- Anonymity: participants have the right to be anonymized in the case of sensitive data.
  - Using alphanumeric identifiers rather than names for speakers
  - Ensuring that speakers are not identifiable by public audio-visual materials
  - Not all data can be effectively anonymized.

Archiving and re-use of data

Why should I archive?
- Many linguistic anthropology research funders require archives.
- The data are backed up in a trusted repository.
- Others, not just the original researcher, can re-use the data.

Data care: backup and data protection
- LOCKSS (Lots of Copies Keeps Stuff Safe)
- Document archival, working, and presentation formats
- Versioning:
  – Keep track of different versions of the data.
  – Use versioning software, such as Subversion, for collaborative projects.

Archiving and re-use of data

Archival concepts
- Users and use case scenarios
- Access:
  - Open
  - [Graded](#)
  - Closed, based on confidentiality agreements
- Ownership: Intellectual property, copyright
- Deposit: guidelines for which language, data and metadata formats
- Original ("raw") vs. edited data

Use the "How to Deposit" guidelines of existing archives to learn more about
- Common data formats.
- Ethical protocols.
- Best practices.

Archiving and re-use of data

Key archives for linguistic anthropologists: An alarmingly short list

Open Language Archives Community, OLAC, participating archives, http://www.language-archives.org/archives

- AILLA, Archive of the Indigenous Languages of Latin America, http://ailla.utexas.org/
- ANLA, Alaska Native Language Archive, https://www.uaf.edu/anla/
- ELAR, Endangered Language Archive, http://elar.soas.ac.uk
- The Language Archive, https://tla.mpi.nl/
- PARADISEC, Pacific and Regional Archive for Digital Sources in Endangered Cultures, http://paradisec.org.au/

Mobilization: Re-using your outputs

- Publishing articles while writing a dissertation
- Publishing a monograph after the dissertation
- Sharing primary data and metadata
- Maximizing re-use potential by others

In-class exercises: The basics: Working with data (Slide 15)

Discussion questions:

1. How do you create data? In what formats? How is it organized?
2. If you start out a project with the goal of open access (meaning that at least some of the primary data must be publicly accessible),
   - How does that choice affect your working with speakers and creating a data collection?
   - In particular, how do you respect community norms, which will almost certainly show that not all data can be freely available to the public?
3. Name at least one born-digital data type and one non-digital data type that you might use.
4. What issues should your digital data management plan take into account? How might these issues differ for born-digital vs. non-digital data?
5. Suppose a Martian (who speaks your language) discovers your miraculously preserved data in 100 years. What documentation would you need to include to make sure the Martian can open and understand your data?

Outside-class exercise: Create a first draft of a data management plan (Slide 9)

Objectives: Draft a DMP (1-2 pp.) and answer two reflection questions.

Use an existing DMP as a kind of a checklist, to make sure all elements are included. (See sample DMPs from NEH and NSF available at http://www.americananthro.org/methods.)

1. Who is your target: what institution's DMP will you use? Funding agency, First Nations/tribal institution, NSF Doctoral Dissertation Improvement Grant (DDIG)
2. What elements must be included? Permission letter from community; permission letter from archive; acceptable archival formats; research locale and community; research scope; data backup and sustainability
3. What elements are specific to your project? Community-specific data access and collaboration requirements; community research product desiderata; political or social considerations; special data types; special strategies to protect media in a particularly humid or cold climate.
   - Not all of these belong in a DMP, but all will affect research design.
4. Your DMP will likely include most of the elements listed on Slides 10 and 11.

Reflection questions:

1. What difficulties (logistical, methodological, or other) did you identify in the process?
2. Do the above requirements change your project design? If so, how?

Optional in-class exercises

Slide 6. Ensuring the future: Data workflow

Discussion: Give examples. Brainstorm data's path through the workflow.

Slide 7. Ensuring the future: Formats

Discussion
1. Is a word processing document (e.g., .docx, .odt) document an Archival, Working, or Presentation format? An mp3 audio? A .tiff image?
   Note, Many A/V formats can be more or less compressed, and for archival purposes, less compressed (less lossy) is always better.
2. Why might an archival format be awkward to work with or share?
3. Why might a presentation format be a poor choice for archiving?

Optional in-class exercise: Ensuring the future: Data management plans (Slide 9)

Discussion of data management plans

Video and notes of Speaker A were recorded with consent by Local Researcher B and analyzed by graduate student C; these need to be placed on Professor D's website.

Scenario 1:
1. What will go into the DMP about crediting participants in a research paper?
2. How about when publishing the video online?

Scenario 2:
1. How will the DMP attend to changes in consent? Examples: a consenting participant wants to be anonymized or recognized; a consenting participant withdraws consent; community leaders or a participant stipulate(s) that part of the archived materials be closed to the public.
2. How does your own relationship with the community affect your research design?

Optional in-class exercise: What are our responsibilities? (Slides 22, 23)

Ethical, legal, and moral obligations
1. Instructor presents examples of legal actions that are ethically dubious and ethical actions that are potentially illegal for the students to debate. Course participants then present examples of how they have shared or will share data in the two contexts, and discuss how intellectual property rights and legal issues interact with their obligations as researchers.
2. Name at least two locally appropriate steps that can be taken to ensure shared data access by the language community.

Closed/Limited/Open Access debate:
1. Imagine or enact a role-playing debate between people who take strong positions on the issue of protecting the exploitation of community knowledge vs. "all information wants to be free."
2. Bring up the strongest arguments for each position (with real-life examples), and then see how best a compromise position that addresses all needs is reached.
3. Possible roles (who may be argue any one or multiple sides of the debate): Indigenous community elders, indigenous linguists, a digital humanist or corpus linguist, an NSF representative, a PhD student (indigenous of the community, indigenous of another community, or non-indigenous), a specialist professor, the university IRB, etc.
4. When might we or our language consultants not want to share data?
5. Will wide data-sharing lead to researchers being "scooped," (i.e., having someone publish your intellectual property before you do)?
6. Not all data users will be uni-disciplinary linguists or even academics; what steps can be taken to make the data maximally accessible to and interesting for multidisciplinary groups as well as non-academics (such as those in public policy, NGOs, unrelated language communities looking for a possible model, and the public)?
7. Data collection can be faulty and preservation imperfect. What steps can be taken to mitigate mistakes?

Instructor notes: Organization and Key data management practices (Slides 3, 5)

For additional information, see also Arienne M. Dwyer, "Expanded Materials for Digital Data Management in Linguistic Anthropology." Arlington, VA: American Anthropological Association, 2016. Available at http://www.americananthro.org/methods

Aim: To introduce best practices in data management for researchers in linguistic anthropology.

Introduction: Data management is crucial for good research. Scholarship creates a range of data forms that require converting, analyzing, storing, and sharing. As scholars, we have a responsibility to make sure that data endure into the future in accessible formats. These data are gathered by researchers (often from participants), and usually receive input from many people. We are therefore also are responsible for the ethical, legal and intellectual property issues arising from these data, including proper attribution/anonymization and adhering to conventions and laws of relevant locales. Archiving and sharing the output of research in print and online venues (a.k.a. publication) requires attending to best practices in data management. Most research funders now require a data management plan, in order that your results and data be enduring and public.

Beyond scope: This module does *not* cover methods of obtaining grants or human subjects permission. It is not a tutorial on intellectual property or other national or international laws. It is also not a guide to digitization or format conversion. (See References and Additional Resources)

Data management and archiving begin at research design, not after the data are collected.

Ethics begins at research design: we have a responsibility to plan and carry out research in partnership with a community; to ensure that the work benefits that community, as well as our institutions, our funders, and ourselves; to ensure that our work conforms to local moral and ethical practices, institutional regulations, as well as national and international laws.

Instructor notes: Key data management practices (Slide 5) and Ensuring the future of your data (Slides 6, 8)

Pretty good practice is good enough. Good practices are not out of reach. Don't let "best practices" or this module keep you from learning pretty good practice (Di Paolo et al. 2006).

Module content: The module presents an overview of the key issues in data management via a tip of the iceberg approach: the digital data workflow from research design and project planning to data creation, data management and analysis to preservation, reusability and publication. The contents in each unit can be covered in more or less detail depending on available time. The accompanying exercises ask participants to come up with examples from their own experience and apply that unit's concepts to those examples. Ideally, the instructor will also provide use cases for each unit.

Ensuring the future of your data (Slide 6): Aim: To provide an overview of the issues; each bullet is relevant, no matter what the project. Each of these introductory issues is revisited later in the course. Bullet points can be exemplified by the instructor.

Ensuring the future (Slide 8). Common data disasters and workflow inefficiencies (exemplified by instructor)
- losing the sole copy of one's data, by failing to back up one's *in situ* research
  - solution: regular backup (LOCKSS)
- storing data in an obsolete proprietary format, such that it can no longer be read
  - solution: use open formats and consider storing in multiple formats
- using a non-Unicode font, and later having a "character salad"
  - solution: use a Unicode-based font (there are hundreds, but e.g. Arial Unicode)
- overwriting a file with an inferior copy with the same name
  - solution: versioning (better) or unique file naming
- a collaborative team naming files every which way, including *wedding.wav.*
  - solution: systematic file naming

Instructor notes: The basics: Working with data (Slide 12) and The basics: Working with data: Naming and conventions (Slide 18)

Overview, Slide 12: Aim: To introduce the many data types and formats, and to describe the minimum a researcher must do to create enduring data. Information regarding software will need regular updating.

Naming and conventions (Slide 18): detailed information
- Data format: What file format are the data stored in? Best: open formats
- Character format: How are the characters encoded? Best: Unicode (UTF); Second best: (lower) ASCII.
- File naming: Consistent, documented, short, no whitespace, no upper ASCII characters
- File structure: consistent, documented, avoid over-use of folders

Well-formed data:
- Distinguishing the form and content of data
  - Well-formedness: durable and reusable
  - Interoperable: document formats and data structures (e.g., in readme.txt), so that data can be opened with a wide range of generic software on any platform
- Keep data in non-proprietary (open) formats
  - Proprietary (closed) formats: encoding is a trade secret of a commercial company; you generally have to purchase their software (the latest version) in order to decode the data. If you don't, you lose your content.

Instructor notes: What are our responsibilities? (Slides 21–23)

Aim: To discuss researchers' ethical and legal responsibilities and Intellectual Property Rights. It is best to confront these issues during project planning, well before an IRB application. Attention to ethics equals good data. Also emphasized are the limits of Open Access: full consultation with communities forms the basis for solid data management plans and sharing arrangements that align with community norms.

Community access: The community, not just the researcher, determines access, in terms of who has access, and how much access.

- project data are co-owned by researcher and community
- what is shared depends on the types of data and the research context
- indigenous communities may also have their own protocols, as elaborated further under Legal responsibilities

Moral obligations, additional elements

- Research products useful to communities (e.g. pedagogical materials, children's books) are usually not allowable expenses by funders
- Such difficulties do not justify avoiding these legal and moral obligations;
- Resolving these issues is usually community- and project-specific.
- Example of key conflict in linguistic anthropology: Withdrawing "informed consent" on dictionary making

Agreements with stakeholders

- "Stakeholders" include participants, research team, local or national bodies, funding bodies, and home institution (see Dwyer 2006).

- Agreements are proposed during research design and re-visited for potential changes throughout the research.

  - Agreements concern many key topics: attribution/anonymization, compensation, responsibilities and division of labor, access to and rights in data and field notes, co-authorship on deliverables including publications, data access, archiving plan and liabilities.

  - Agreements may be written, verbal, or third-party (e.g., via a village leader).

- International and/or interdisciplinary teams must consider all relevant nations and ethical codes/practices of all relevant disciplines.

- The IPinCH project is an excellent resource (IPinCH 2016).

References

Case Study: File Formats. Stanford: Stanford University Libraries, n.d. http://library.stanford.edu/research/data-management-services/case-studies/case-study-file-formats

Dwyer, Arienne M. 2006. "Ethics and Practicalities of Cooperative Fieldwork and Analysis." In *Fundamentals of Language Documentation: A Handbook*, edited by Jost Gipper, Ulrike Mosel, and Nicolaus Himmelmann, 31-66. Berlin: Mouton de Gruyter, 2006. https://kuscholarworks.ku.edu/handle/1808/7058

Di Paolo, Marianna, with the assistance of Gary Holton, Susan Smith, Arienne Dwyer, Steve Moran, Doug Whalen, Julia Good Fox, and Barbara Need. "Collecting Primary Texts, Working Group 1 Report." In *Proceedings of 2006 E-MELD Workshop*. 2006. http://emeld.org/workshop/2006/wg/wg1-report.rtf

Digital Preservation 101. Digital POWRR [Preserving (Digital) Objects with Restricted Resources], 2012. http://digitalpowrr.niu.edu/digital-preservation-101/

DuBois, John W. *Transcription in Action: Resources for the Representation of Linguistic Interaction*. Santa Barbara: University of California, 2006. http://www.linguistics.ucsb.edu/projects/transcription/representing

IPinCH [Intellectual Property Issues in Cultural Heritage Project]. *Think before You Appropriate: Things to Know and Questions to Ask in Order to Avoid Misappropriating Indigenous Cultural Heritage.* Vancouver: Simon Fraser University, 2016. http://www.sfu.ca/ipinch/sites/default/files/resources/teaching_resources/think_before_you_appropriate_jan_2016.pdf

IPinCH. *Traditional Knowledge Fact Sheet*. Vancouver: Simon Fraser University, 2016. http://www.sfu.ca/ipinch/sites/default/files/resources/fact_sheets/ipinch_tk_factsheet_march2016_final_revised.pdf

Leipzig Glossing Rules. Leipzig, Department of Linguistics, Max Planck Institute for Evolutionary Anthropology, 2015. https://www.eva.mpg.de/lingua/resources/glossing-rules.php

Levine, Melissa. "Policy, Practice and Law." In *DH Curation Guide: A Community Resource Guide to Data Curation in the Digital Humanities*. 2016. https://guide.dhcuration.org/contents/policy-practice-and-law/

Library of Congress. *Sustainability of Digital Formats: Planning for Library of Congress Collections*. 2013. http://www.digitalpreservation.gov/formats/

Linguistic Data Consortium. *Data Management Plans*. Philadelphia: University of Pennsylvania, 1992-2016. https://www.ldc.upenn.edu/data-management/data-management-plans

Newman, Paul. "Copyright Essentials for Linguists." *Language Documentation & Conservation* 1(1) (June 2007): 28-43. http://scholarspace.manoa.hawaii.edu/bitstream/handle/10125/1724/newman.html

References

National Information Standards Organization. *Understanding Metadata*. Bethesda: NISO, 2004.
http://niso.org/publications/press/UnderstandingMetadata.pdf

OLAC. *Metadata*. Open Language Archives Community, 2008. http://www.language-archives.org/OLAC/metadata.html

Simons, Gary F. "Ensuring That Digital Data Last: The Priority of Archival Form over Working Form and Presentation Form." *SIL Electronic Working Papers* 2006-003. http://www-01.sil.org/silewp/2006/003/SILEWP2006-003.htm

Traditional Ecological Knowledge. Washington, DC: Society for Ecological Restoration, 2016. SER http://www.ser.org/iprn/traditional-ecological-knowledge

Copyright FAQ. Lexington: Department of Linguistics, University of Kentucky, n.d. https://linguistics.as.uky.edu/copyright-faq

UNESCO. *Best practices on Indigenous Knowledge*. UNESCO, Management of Social Transformations Programme, n.d. http://www.unesco.org/most/bpindi.htm

van Driem, George. "Endangered Language Research and the Moral Depravity of Ethics Protocols." *Language Documentation and Conservation* 10 (2016): 243-52. http://scholarspace.manoa.hawaii.edu/bitstream/10125/24693/1/vandriem.pdf

Additional resources

See also "Resources for Anthropological Digital Data Management" available at http://www.americananthro.org/methods.

Ethics and rights

American Anthropological Association. *Principles of Professional Responsibility*. 2012.
http://ethics.americananthro.org/category/statement/

American Sociological Association. *ASA Code of Ethics*. 1999. http://www.asanet.org/membership/code-ethics

Linguistic Society of America. *Ethics Statement*. 2009. http://www.linguisticsociety.org/sites/default/files/Ethics_Statement.pdf

Panel on Research Ethics. *TCPS 2: Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans*. Government of Canada, 2014. http://www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/Default/

World Archaeological Congress. *Code of Ethics*. 2016. http://worldarch.org/code-of-ethics/

Acknowledgments

Modules: *Writers,* Arienne M. Dwyer, Blenda Femenías, Lindsay Lloyd-Smith, Kathryn Oths, George H. Perry; *Editor,* Blenda Femenías

Discussants: *Workshop One, February 12, 2016:* Andrew Asher, Candace Greene, Lori Jahnke, Jared Lyle, Stephanie Simms
        Workshop Two, May 13, 2016: Phillip Cash Cash, Jenny Cashman, Ricardo B. Contreras, Sara Gonzalez, Candace Greene, Christine Mallinson, Ricky Punzalan, Thurka Sangaramoorthy, Darlene Smucny, Natalie Underberg-Goode, Fatimah Williams Castro, Amber Wutich

American Anthropological Association:
Executive Director, Edward Liebow
Project Manager, Blenda Femenías
Research Assistant, Brittany Mistretta
Executive Assistant, Dexter Allen
Professional Fellow, Daniel Ginsberg
Web Services Administrator, Vernon Horn
Director, Publishing, Janine Chiappa McKenna