WORKFLOW RECOMMENDATIONS FOR DATA COLLECTION

March 2019

The FROGBEAR project is committed to creating an open-access data repository to be hosted by the University of British Columbia (UBC).

> *As texts and other materials are collected, a robust set of metadata will be created for each item…. This … open access database will display photographs of manuscript and printed texts, serving as a unique virtual library for use by scholars, students and the public. The database will also preserve material brought together by our project. Scholars will be able to use this database to discover and trace connections between seemingly disparate geographical and temporal contexts that would have previously been impossible to discern. The collection as a whole, as well as particularly significant items contained within, will become the basis for scholarly symposia and subsequent publications. It will enhance interactions both among team members and with other research units around the globe. It will help disseminate our research findings and will continue to be supported by the UBC funding and/or external funding beyond the life of this SSHRC grant.*
>
> *-excerpt from proposal to SSHRC, Oct 2015*

The repository will be different from much of the material found online, through better:

- Access (we might get access to material not already documented)
- Attention (scholars pay attention to things tourists do not)
- Accuracy (thorough description, in metadata produced by scholars)
- Systematic coverage, allowing users to see complete sets of things and put them in context and in relation to one another through documentation of whole sites
- Reusability, under open license terms

We will aim for high quality, and need to achieve minimum standards, but the above are more important than quality alone.

The UBC Library digital repository, cIRcle, will archive the material and make it available through their Open Collections portal (https://open.library.ubc.ca/.) This meets SSHRC requirements of open-access and data preservation. The project has worked with the library to develop clear metadata standards.

This document contains recommendations for data and metadata collection for FROGBEAR cluster field trips. The process can be adjusted based on the needs, capacities, and working conditions of each group. However, in order to be included in the repository hosted by the UBC Library, the material collected must adhere to the FROGBEAR metadata principles and standards. In particular:

- All data supplied must include complete metadata in the required format (typically, in the form of a spreadsheet in each folder of image/multimedia files)
- All multimedia material should be in open (non-proprietary) formats
- All material must be accompanied by a cIRcle license signed by its creator(s), releasing the material under a Creative Commons licence (must answer YES to both questions)
- Any depictions of identifiable persons must be accompanied by a signed release form

Before Field Trips

At least one month before departure, each group designates at least one (preferably two or more) participants to upload files and metadata ("uploader") and emails their names to the Project Manager, Vicky Baker. These people will also serve as contacts if issues arise with the material collected after the trip.

Each uploader signs up for a Compute Canada/WestGrid account and sends their WestGrid user ID to Vicky Baker. This is the system we use to transfer the material collected to FROGBEAR staff at UBC. (See the separate instructions for this, and note that it may take up to a week to finalize the creation of an account.) The uploader will also be sent instructions on how to use the OwnCloud interface to upload material.

Cluster leaders designate one or more participants to collect release forms (for individuals shown/recorded in files), and ensure that a sufficient number are printed before the trip. These cover not only researchers, but also people who are appear in images or whose voice is recorded, such as experts or religious practitioners who are interviewed or filmed.

Cluster leaders designate one or more participants to collect permission forms/cIRcle License (from those who create media files) and ensure that a sufficient number of blank forms are printed before the trip.

If paper metadata forms will be used (Frogbear_Printable Creator Metadata.docx and Frogbear_Printable Metadata.docx), cluster leaders ensure that a sufficient number will be available.

Cluster leaders should also inventory the equipment that the participants will bring, and determine whether additional equipment will be needed. In particular, we recommend identifying the available still and video cameras, audio recording devices, lighting equipment, supports (tripods and light stands), and power (batteries, backup USB power supplies). Ensure that enough data storage is available: we highly recommend bringing along at least one high-capacity external (USB) hard drive. We also recommend polling participants to identify particular skills that may be useful. Please see Appendix B for full list of suggested equipment.

During Field Trips

Start of Trip

At the start of each field trip, clusters should set aside time to review data and metadata principles and standards, the goals of the particular trip/cluster, material to focus on, expected challenges, and collection procedures and roles. All participants should be familiar with the data collection goals of the trip and the basic FROGBEAR data and metadata requirements and recommendations. We suggest setting aside time to watch our training videos on YouTube/Youku and to discuss your plans and intended outcome.

Participants plan their data collection in terms of the **records** that will ultimately be created in the UBC library's collection. Each record may contain several related files – for example, if the record is about a statue, it may include an image from each side of the statue. If a particular document will constitute a record, then it might include one or more photos of the text, as well as a transcription in the form of a text file. We recommend limiting each record to at most a handful of files, and breaking them into multiple records when there are more than 5-10. For example, if there are only six photos of a temple complex,

they may all be included in a single record, but if there are 20, they may be grouped into multiple records, such as views of the main halls, views of side buildings, and views of the interior altar.

Each record will have information associated with it that describes specific characteristics of the files and what they represent. This is the **metadata**. The metadata must be provided in a format designated by the UBC Library. Instructions for creating proper metadata appear in the **metadata template** (an Excel spreadsheet) and our wiki.

If your cluster would like to record information that does not fit into the existing fields, it is possible to include this information in a supplementary file stored in the repository. If your cluster would like to discuss this possibility, please advise Vicky Baker.

Clusters should also review their equipment, making sure that everyone who will be using particular devices is familiar with their operation. Device clocks should be double-checked to ensure that they are recording the correct local time; location recording should be enabled for GPS-enabled devices. If possible, a trial run can be made to practise using the equipment and the metadata recording procedure.

Daily

For each day of data collection, cluster leaders designate teams of data and metadata recorders, ensuring that each person recording data (images or audio/video) is paired with one or more recording metadata

During data collection, each small team should record metadata as they collect data, in draft form. This is particularly important for information not evident from the images/recordings, such as the identity and significance of objects depicted, the names of people represented, and the location, if it is not automatically recorded by the device. It is also essential to record who produced each file (name of the photographer/videographer). Depending on the working conditions, this may be done electronically, but especially when working outdoors it will usually be easiest to keep paper notes. A dry-erase board can be used at the beginning of a series of images, listing the metadata that is common to that set of files. By taking a photo of the board, each new set of material will be easily identifiable when scrolling through images.

> *Note on releases from people depicted*
> Under Canadian privacy law, in order to make public any identifiable representation of a living person, permission must be obtained from that person. This includes any image/video in which an individual is recognizable (e.g., showing their face) or any recording of their voice. So we must ensure that (a) the person depicted is identified by name in the metadata spreadsheet and (b) we have received a release form, signed by that person, permitting the publication of the material. If you have recorded images or audio of any people—including project participants—please obtain a signed release form as soon as possible. We will not be able to post any material for which we do not have a signed release, or in which unidentified people appear.

As soon as possible after data is collected--ideally, at the end of each day--teams should meet to assemble the data that was collected and to input metadata based on the draft recorded in the field. The data should be transferred from devices (cameras, etc.) to a single computer/hard drive.

For clarity, we recommend creating separate folders for each day of activity. You may also further divide the files into subfolder for each creator (photographer/videographer). Each folder to be uploaded to the repository should contain a single metadata spreadsheet listing all the files in the folder.

Once the files are organized into the appropriate folder structure, metadata spreadsheets should be created and as much of the metadata as possible should be filled in, based on the draft recorded in the field. Any uncertainties or questions should be highlighted. Team members should also review data files to decide which to include in the final repository. For example, if multiple shots have been taken of the same subject, only the best should be kept, and any blurry or misaligned photos should be excluded. Some problem images can be corrected in software, for example those that are misaligned (upside down, turned 90 degrees), or over/underexposed. Files can be grouped into four categories:

1. Files designated for inclusion in the repository--this is material for research and educational purposes, and it must be accompanied by full metadata
2. Files to be shared with the FROGBEAR project for other purposes, such as publicity images, but not meant for the repository. This might include, for example, group photos of participants, which can be shared in promotional materials such as newsletters.
3. Files the researchers will keep for themselves, but not send to FROGBEAR. This might include redundant material--for example, if multiple photos were taken of a single object, only the best one or two would be included in the repository, while the researcher may keep all of them; personal photos also fall into this category.
4. Files to be discarded (out-of-focus/poor quality, etc.)

Only material in categories 1 and 2 will be shared with FROGBEAR, via the Compute Canada/WestGrid system. Files in both categories require the appropriate releases and permissions to have been obtained from all creators and all individuals depicted. Without these, we cannot include the files in the repository or use them for other purposes such as promotion of the project.

Full metadata is only required for material in category 1. Each folder of material *must* include a spreadsheet with complete metadata or it will not be included in the repository.

Material in category 2 does not require complete metadata, since it will not be going into the repository collection. However, it will be helpful to the FROGBEAR staff to provide descriptions of the contents. To do so, either use the metadata spreadsheet (without needing to fill out all the fields, only those that will be useful in understanding the images/videos/etc.), or supply a separate document such as a Word file. If there is material you especially want to highlight or that you think would be good for publicity purposes, you can email Vicky Baker to notify her.

End of Trip/After Trip

Participants should set aside time at the end of the trip to review their data and metadata, resolving any questions and filling in gaps. If a large amount of data was collected, it can easily take a full day to select which files to use, to write complete, accurate, and consistent titles and descriptions, and to transfer all the files to a single hard drive.

Once it has been finalized, teams can share their material with the uploader(s)--the easiest method is to copy the material to a portable hard drive. Depending on their schedule and preferences, teams may choose to provide files to their uploader(s) all at once at the end of the trip or in smaller daily batches. Depending on schedule and the available internet connection, uploaders may upload material at the end of the trip or as soon as possible after returning.

Once all of the trip's material has been uploaded, the uploader should notify Vicky Baker by email.

Appendix A: Possible types of material

1. Photos
    a. Subjects
        i. People (consent form is required)
        ii. Objects
            1. Statues, ritual implements, etc. (indoor, outdoor, 360)
            2. Texts (indoor)
            3. Shrines (indoor, outdoor)
            4. Cliffs (outdoor)
            5. Caves (indoor)
            6. Stone inscriptions (indoor, outdoor)
            7. Buildings (indoor, outdoor)
            8. Context (combining multiple objects)
    b. Special Techniques
        i. Lighting
            1. Flash (if permitted)
            2. Side illumination (for reliefs/inscriptions)**see video on photographing epigraphy*
        ii. 360 degree (best with diffuse light)
            1. Exterior (landscape etc.)
            2. Exterior of object
        iii. Panoramic

2. Videos (consent form is required if people appear)
    a. Ceremonies / rituals
    b. 'making of'/ project promotion
    c. Conversations with key individuals
    d. How-to/training/examples of techniques
    e. 360 videos

3. Audio
    a. Conversations (consent form is required)
    b. 'other sounds'/ambient

4. Maps
    a. Sketches
    b. GPS coordinates/tracks

Appendix B: Suggested Equipment list

☐ Tripod/light stand

☐ Camera (DSLR, point & shoot, or phone)

☐ Cell phones with good cameras, plus GPS and compass apps

☐ Video Camera / GoPro (LED light and external mic best for interviews)

☐ Audio Recorder + sock (to muffle wind)

☐ External Hard Drive (USB power only, preferably ruggedized)

☐ Battery backup (USB power)

☐ Power Adapter

☐ Flash/LED Lighting & batteries

☐ SD Cards and readers

☐ Extension Cord/Power bar

☐ Notebooks/paper (preferably water-resistant)

☐ Rulers/colour cards (to measure & scale in photos, and for colour correction)

☐ Umbrellas/plastic sheeting (for rain protection)

☐ HDMI cable (for viewing images on larger screens)

☐ Clipboards (for groups filling out forms by hand)

☐ Dry erase boards (small)