7. Data Retention

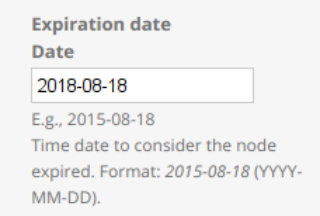
Ethnographic projects tend not to have “embargo periods” and ethnographic data tends not to have “expiration dates” whereas both are quite common for digital data management in science and engineering disciplines. There are particular reasons that account for this difference. First, ethnographers tend not to share “raw data” but drafts of partial and preliminary analyses with other ethnographers and other research groups. The very concept of “raw data” is foreign to most contemporary ethnographic projects since data only acquires meaning in the context of a particular ethnographic project. To put in different terms, data must refer to what we call “conditions of production” to acquire particular meaning and become useful for research purposes. Ethnographic data is data generated in the context of human relationships in general and forms of human and non-human interaction in particular. Without information on these basic foundations of data production, ethnographic research data is not useful and not usable by other researchers. Lastly, the reason why expiration dates are not common for ethnographic data is because ethnographic data represent documents of, not only anthropological and sociological interest, but of historical importance in many cases. They can be used for building archives and for comparative efforts at any point in the future as long as they are properly stored, extensively described, and made available through flexible licensing schemas and interoperable data management systems with open, public interfaces.

In the course of specifying and implementing PECE 1.0, we made design decisions with the goal of questioning and changing the current understanding and usage of data retention policies. The aim was the pose the trade-off between data protection and openness under a different framework with a focus on Open Source technologies, Open standards, and Open Data. Instead of focusing on data protection against competition in the sciences for priority of publication, which tens to be the current norm and practice, we channeled our efforts on the hard task of creating infrastructures and fostering collaborative ties in which data are contributed to a common pool from which many researchers and related disciplines can draw from. PECE, in this sense, aims first and foremost to be a contribution to a digital commons for humanities and social sciences. Therefore, the current notion of “data retention” is not particularly useful nor central to our mission. There are, however, very important exceptions in which “data retention” can be understood for our disciplines under the light of ethical guidelines and privacy issues in the ethnographic practice.

Ethical guidelines and privacy issues (such as the ones we described in the subsequent sections on “Disposition” and “User Agreements of this document) are key topics of debate and concern in respect to retention periods as ethnographic data is kept secure and private given potential privacy concerns or expressed intent of research subjects (whom, in the context of PECE, we identify as “research co-participants” instead of informants). “Retention periods” for ethnographic projects, therefore, are usually established around the sensibilities of our co-participants, observance of their rights to privacy and anonymity and, ultimately, the needs of a particular project to protect, analyze, and then delete a particular piece of data under the request of a research co-participant.

In respect to its technical capabilities, PECE provides users with the ability to identify sensitive pieces of datum and change its status after a certain period of time (from published to unpublished, for instance) and for certain functions to be performed (such as deleting a file or artifact). This is important for the ethical and privacy concerns we mentioned above, but,

particularly to remind our users that certain pieces of data must be deleted after the project is over. Compliance with requests for deletion of data can be accomplished on PECE by setting up a “timer” on PECE artifacts. Under “Publishing Options” for every artifact, the user has the option of setting up an expiration date at the time of submission in the following format: YEAR-MM-DD (year-month-day).



Alternatively, deleting artifacts per requirement of research co-participants can be performed in batches. It is necessary, first, to collect the “Node ID#” of every exception and save it into an unordered list, such as [1. 3. 10. 49. 321. 5423. 43, etc.]. Then, a simple shell script can be used to remove ethnographic data that was requested to be deleted:

```
#!/bin/sh
# Declare the array with the nodes that were requested to be deleted
array = (Node IDs # i.e. 1 2 3 4)

# Iterate over the array items and delete one-by-one from PECE
for i in "${array[@]}"
do
    drush node_delete $i
done
```

There are ways to collect Node IDs with specific expiration dates by executing a query on the PECE database. This can be done using `drush` and Drupal “Entity API” with the following command:

```
# Query for nodes with expiration dates, saving the output to a file:
$ drush php-script expired_nodes.php > expired_node_ids.txt

# 'expired_nodes.php'
<?php
$now = new DateTime(); // time when the query was executed
$query = new EntityFieldQuery(); // make usage of Entity API
$query
   ->entityCondition('entity_type', 'node')
   ->fieldCondition('field_expirationdate', 'value',
       $now->format('Y-m-d'), '<')
   ->addMetaData('account', user_load(1));

$result = $query->execute();
drush_print_r($result); // terminal output as an example
?>
```

It is part of our roadmap to create an automated way of marking and deleting “private” content with “expiration dates” for PECE 2.0. The improvement of this data management policy will include the identification of sensitive data through tagging, regular, scheduled scanning

across the dataset for sensitive, private content, and systematic deletion of data upon completion of a research project as specified on the “Project” section on the platform.

Instruction for Developers: It is necessary to install the following modules in order to implement this data management policy: **“node_expire”** v. 7-x 1.8 and **“rules”** v. 7-x 2.9. Configuration for “node_expire” is as follows: please select [Trigger "Content Expired" event only once when the node is expired, Date popups, Allow expire date in the past]. Please observe that “Date Popup” must be activated for the Date module (and the following option must be selected: [Use default jQuery timepicker].