The charter is the foundational document that describes the rationale, goals, plan of work, resources needed, terms and conditions, and outcomes of a Center for Digital Humanities at Princeton (hereafter CDH) project. Charters are written by core members of a project team in a series of planning meetings taking place over the course of a month. The planning process is intensive, collaborative and requires substantial input from everyone on a team. Charters serve as formalized agreements among all team members on such crucial questions as scope, technical design, infrastructural needs, and success criteria.

A draft of each project charter is peer-reviewed by all CDH staff, and optionally by additional partners or stakeholders, at a "design review" before the start of project work. It is circulated at least one week before the review takes place in an open comment period. Questions and concerns from this period may be raised at the design review. Project teams have two weeks after the design review to address any issues raised and make any requested changes. Project work only begins (and funds are released) once the charter has been finalized and signed by the Project Director (PI) and the CDH Faculty Director. Charters are amended as necessary throughout the project lifecycle to document major changes and note when "Built by CDH" Software Warranty and "Built by CDH" Long Term Service Agreement take effect, and serve as part of the CDH project archive.

CDH charters and their planning documents exist in several forms as we have refined them over the years and tailored them to the several types of projects we have supported. For more about CDH project management, including the charter process, visit: https://cdh.princeton.edu/research/project-management/

Cite this document:

PEMM Charter (2019-20)

Part I: Project Overview

Stories have been told for almost two millennia about the Virgin Mary, the mother of Christ, and the miracles she has performed for the faithful who call upon her name.

One of the most important collections of such folktales is the body of over 700 Ethiopian Marian miracles, written from the 1300s through the 1900s, in the ancient African language of Gəʿəz (also known as classical Ethiopic).

These story collections, called the *Täʾammərä Maryam* (The Miracles of Mary), are central not only to the ancient church liturgy of Ethiopia, but to the daily felt and religious life of 50 million Ethiopians and Eritreans. Princeton University has in its Firestone Library one of the largest and finest collections of Marian miracle manuscripts anywhere in the world outside of Ethiopia, with over 130 codices and hundreds of textual amulets. Worldwide, at least 100,000 *Täʾammərä Maryam* manuscripts exist, some with just a handful of stories, some with hundreds, and many with different versions of the same stories.

While the *Täʾammərä Maryam* is one of the most important African archives of texts, basic information about it and its stories are lacking; as a result, scholars can authoritatively state almost nothing about them. How many are there? When was each written? What themes do they have? Have these African stories grown and changed across regions, languages, and periods? Princeton Ethiopian Miracles of Mary (PEMM) project will collect and collate information about these hundreds of stories across hundreds of manuscripts as the basis for an open access resource that will enable researchers and Ethiopian community members around the world to conduct in-depth research on this vital corpus. Wendy Laura Belcher, professor of African literature in the departments of Comparative Literature and African American Studies will serve as the project's Principal Investigator (PI). PEMM was begun with a Center for Digital Humanities at Princeton (CDH) Dataset Curation grant.

Description and Objectives

With the guidance of the CDH, the PEMM project team will collect and collate data about hundreds of Marian miracle stories in hundreds of Ethiopian manuscripts. Our aim is to enable computational analysis of this vital corpus of African folktales and to generate answers about their number, dating, origin, provenance, themes, recensions, translations, sources, placement, and diachronic change. The CDH and PEMM will design a robust data structure to migrate, store, connect, validate, and query the data. We will discuss a preliminary web interface with sample data visualizations that will make that data available to scholars in the United States, Europe, and Ethiopia; this interface may not be possible in AY20.

Relevant Resources and Projects

PEMM builds on the work of previous scholars, using resources created by others: the most important for PEMM is the Macomber Handlist, and the finding aids of the Princeton University Rare Books and Special Collections. We hope to reference the manuscripts in the British Library. We will also make use of and share data with both the Oxford Cantigas de Santa Maria (CSM) database and the University of Hamburg Beta Maṣāḥǝft (Hamburg BM) Database with the intention of avoiding overlap as much as possible. For complete information about what has been collected and catalogued thus far about the *Täʾammǝrä Maryam,* please see Appendix A.

Research Questions

We are collecting and collating data in AY20 in order to answer three main research questions.

- **How many Ethiopian Marian miracle tales are there?** Scholars have not been able to arrive at an accurate number of how many Ethiopian Marian tales there are despite a century of labor on the issue. One scholar says 540, another says 643; a current database project (the Hamburg BM) has likely identified over 700. Meanwhile, Princeton has some stories in its manuscripts that appear on none of those lists. PEMM has access to what those scholars did not have: thousands of digitized manuscripts (instead of dozens) and sophisticated ways of curating and analyzing the data in those manuscripts (see below).
- **What are the themes of the Ethiopian Marian miracle tales?** Macomber's 1980s catalog provided keywords for some of the tales, but many of the terms were dated, insufficient, and inconsistently applied. By enhancing dataset with better keywords, by refining and standardizing keywords from a controlled vocabulary and consistently applying them, PEMM will give scholars access to an accurate dataset of tale themes and ways of studying how tales correlate with those keywords. We will build a controlled vocabulary from combining and refining Macomber's handlist, the Hamburg BM, the Oxford CSM, and consulting the Index on Medieval Art.
- **What is the origin of each Ethiopian Marian miracle tale?** No one has clearly established which of the tales were originally from Europe or the Middle East. Some say only 33 of them, others say 75 of them, but no one has done the work to be certain. This matters not only to Ethiopianists, but to scholars working on the European Marian tales. Our work correlating the Ethiopian Marian tales with the tales in the Oxford CSM database may enable scholars to discern patterns across and analyze indigenous, European, and Middle Eastern Marian tales.

Project Significance

Significance for African and Literary Studies

PEMM is a historic project with a range of scholarly contributions:

- **Makes a disciplinary contribution**: These folktales about the Virgin Mary are rich repositories of cultural knowledge and literary practice, providing a matchless comparative literature site to study tales across continents, languages, and periods. Comparative literature remains a largely Eurocentric discipline and the Marian miracle tales have seldom been studied outside of their European iterations. PEMM provides a useful corrective to such limited approaches and does so through pairing two innovative comparative literature methodologies: distance reading and world literature.
- **Fills a scholarly gap**: African literature in general, and eight centuries of Ethiopian written literature in particular, are criminally understudied. PEMM will enable more scholars to do more research on African literature. It will also bring greater global visibility to this vital corpus through a web interface.
- **Serves an underserved community:** The number of digital humanities projects that focus on African literature is miniscule. Indeed, perhaps the only other initiative is the Programme in African Digital Humanities, 2018–2023, at the South African universities of Cape Town, Pretoria, Stellenbosch, Western Cape and the Witwatersrand, which "[aims to examine](#)the current forms and practices of reading and digital publishing in order to encourage and support self-directed, digital literary enquiries in the South African humanities environments." PEMM is part of increasing the number of digital humanities projects that focus on Africa.
  - For instance, the annual [Digital Humanities conference in Utrecht for 2019](#) decided to have an Africa focus, working to give funding for Africans to attend and noting that African DH projects "cover the spectrum of DH topics in a somewhat different way than elsewhere." However, only one panel was about Africa: "African Languages And Digital Humanities: Challenges And Solutions," (which has a linguistics focus). Only one of the paper presentations seemed to be about Africa. Isabelle Alice Zaugg's "Global Language Justice in the Digital Sphere: The Ethiopic Case" is "an instrumental case study of Unicode inclusion and the development of supports for the Ethiopic script and its languages." Thus, the DH focus seems to be on methods, underscoring the need for a project like PEMM that focuses on literature.
- **Collects scattered information in one place.** Ethiopian literature has been the subject of study for some centuries. There are large repositories of Ethiopian manuscripts inside and outside of Ethiopia, and massive cataloging and digitizing projects have been underway for the past sixty years. But, little of this information is available online, in one place, in English, for computational analysis.

- **Focuses on literature.** Most projects that focus on Ethiopian manuscripts do not attend to literature. They are linguistic or philological in nature, focusing on manuscripts as material objects, and tend to prioritize biblical books. PEMM is focused on African stories and their themes, providing a useful corrective to the overemphasis on influence and apparatus and underemphasis on African thought and creativity in Ethiopian studies.
- **Increases information about and access to stories.** Of the tens of thousands of *Täʾammərä Maryam* in existence, only a few hundred have been catalogued, and only a few dozen of those have been cataloged with any detail; that is, naming the exact tales in that manuscript. PEMM will increase the number of cataloged *Täʾammərä Maryam* manuscripts.
- **Provides foundation for Belcher's book.** Belcher works on Ethiopian literature in general and has a book in progress on *Täʾammərä Maryam*, titled *Ladder of Heaven: The Miracles of the Virgin Mary in Ethiopian Literature and Art.* It is a book of literary analysis, which will appear with many gorgeous illuminations of the tales from Princeton's manuscripts. Given the dearth of information about these stories, PEMM will provide a necessary basis for the writing of this book.

Significance for Digital Humanities

The CDH's approach to this particular project will also serve to make existing tools and approaches more robust and more useful to Digital Humanities researchers who do not have the support of a development team.

- **Google Sheets as a simplified relational database.** We will develop and document a model for working with Google Sheets as a simplified relational database, and exploring the possibilities offered by an exportable static site based on relational data exported from Google Sheets. Working with spreadsheets and Google Sheets is obviously not new for data work or for Digital Humanities.[1] However, it seems clear that there is a need for a data curation and management solution that sits somewhere between a spreadsheet and a relational database.[2] By applying CDH Development & Design Team skills and expertise, we will push these technologies forward in a way that will benefit others, including those doing data curation and graduate students working on their own Digital Humanities projects. Our approach will include structuring the data across multiple sheets as a simplified relational database, considering the spreadsheet as a user interface, providing enhanced functionality via scripting, and documenting the data structure and the implementation. We will also write scripts to automate data export, data validation and reporting--tools which have the potential to be generalized for wider use.

---

[1] For example, see Matthew Lincoln's post on using Google Sheets as part of a Getty data migration project.
https://matthewlincoln.net/2018/03/26/best-practices-for-using-google-sheets-in-your-data-project.html
[2] See for instance the popularity of products like Airtable or the existence of projects like NodeGoat.

- **Furthers DH work with static website technologies**. Using static website technologies, as championed by the Minimal Computing working group, is also not new for Digital Humanities, although it is new for a CDH sponsored project supported by the Development & Design Team. Our commitment to development best practices and documentation will help further work being done by others to make static sites more accessible to scholars. In addition, for this project the possibility of sharing the results of our work through alternate means is particularly appealing.
- **Takes advantage of new initiatives.** Scholars have access to only a fraction of Ethiopian manuscripts, as most are in remote monasteries. Only a few have been digitized; some estimates put the number as low as 10 percent of all Ethiopian manuscripts. Fortunately, over the last decade, we have seen a huge push to digitize, as Ethiopian manuscripts are globally recognized as an important endangered archive. PEMM takes advantage of this newly available archive.
- **Strengthens partnerships** with scholarly communities in Oxford and Hamburg by establishing a model for sharing data across different projects using different technologies.

Audiences

The audiences for this project are multiple and overlapping.
- **Scholarly audiences for the data:** Scholars of Ethiopian literature (mostly scholars in Europe and North American, Ethiopian and non-Ethiopian) will find this information useful for doing their own research, in particular, the three existing comparative projects will be interested in the data we produce: the Hamburg BM project, the Oxford CSM Database, and the Miracula Mariae project.
- **Scholarly audiences for a public interface:** Clerical scholars (including Ethiopian clerics); non-clerical scholars (scholars of Marian miracles); and students (undergraduate and graduate)
- **Non-scholarly audiences for a public interface**: Ethiopian priests interested in sources and themes for writing sermons.

Project Team

Project Director (a.k.a. Project PI): Professor Wendy Belcher
- Leads and champions project
- Learns enough about project's technical components to be able to describe it on a basic level
- Oversees, participates in, and delegates project work
- Attends regular project meetings (approximately twice a month during periods of active development)
- Supervises and guides Project Manager; keeps open line of communication, responds to emails and questions in a timely manner; alerts in advance of any disruptions or PI unavailability

- - Support for PM may take the form of writing and commenting on charter, helping with project team coordination, approving project workflows, approving project publicity, or other duties TBD in consultation with PM, Technical Lead, and CDH Project Manager and Project Coordinator.
  - If necessary, attends quarterly check-in with Technical Lead
  - If necessary, submits quarterly data progress report to CDH
  - Participates in acceptance testing on software development work
  - Responsible for approving the work and progress of the data team
  - Responsible for project budget, including and overseeing payment of students.
  - Responsible for final project summary of accomplishments

Project Manager: Evgeniia Lambrinaki
- Maintains regular communication with team members, partners, and groups engaged in project work
- Helps to design and implement project workflows with PI approval
- Schedules and facilitates project check-in meetings (including creating an agenda); captures meeting notes
- Tracks progress on project goals and outcomes and communicating with CDH on project progress and/or issues
- Prepares and updates project documentation
- Responsible for overseeing the day-to-day work of the data team
- Responsible for acceptance testing on software features
- Responsible for project publicity (project page, blog) with PI approval

Technical Lead: Rebecca Sutton Koeser
- Oversees design and implementation of project's technical aspects
- Acts as main CDH decision maker on project
- If necessary, modifies decisions about software tools and approach in order to more efficiently and effectively complete the project
- If needed, holds quarterly check-in meeting with PIs
- Has authority to make project decisions if PIs are unavailable
- Responsible for technical documentation at the conclusion of project

CDH Project Manager: Gissoo Doroudian
- Helps manage and coordinate development and design work
- Serves as a resource and point of contact for Project Manager
- Attends and helps create agenda for project meetings (with Project Manager)
- Helps design and implement project workflows
- Supplies periodic updates on project status and progress
- Decides in collaboration with Project Manager who will document meetings

User Experience (UX) Designer: Gissoo Doroudian

- Collaborates on data structure and architecture for project data and advises on configuration and customization for data entry user experience
- Recommends the appropriate types of data visualizations to try with the project data (diagrams/maps)
- Thinks through and conducts user research on access for target audiences specific to this project
- Consults on collaborative and iterative design for content structure and website architecture
- Helps iteratively design a usable, and accessible interface

CDH Project Coordinator: Rebecca Munson
- Advises CDH Project Manager on coordinating development and design work
- Serves as a resource and point of contact for CDH Project Manager
- Advises on project workflows
- Supplies periodic updates on project status and progress
- When applicable, attends and helps to document project team meetings

CDH Developers: Rebecca Sutton Koeser, Nick Budak
- Develop or consult on data architecture and implementation
- Contribute to and review custom software developed for the project
- Document custom software and data architecture
- Write automated tests for custom software
- Experiment with and assesses potential project technologies (e.g. Jekyll, Wax, Hugo) and advise Technical Lead on decisions
- Prototype and implement custom data visualizations
- Provide consultation and training on tools such as OpenRefine to empower Project Director and other project team members to work with the project data

Data Team (Student Researchers)
- Have skills in at least one of these languages: French, Italian, Amharic, Gəʿəz
- Catalog uncatalogued *Täʾammərä Maryam* manuscripts using Macomber handlist identifiers and Gəʿəz incipits

Budget

[Budget available upon request.]

Part II: Grant Year 2019-2020 Plans - Data

Data Status

Types of Data and Storage Format

The data are currently in structured text files, pdfs, Google Docs and Sheets, XML, and MS Word.

Past FY19 Data Work

Macomber Handlist cleanup. In 2018-2019, Belcher and Lambrinaki converted the handlist from a PDF (a scan of a hand typed manuscript with many hand emendations in pen) into a structured text file. They cleaned up the file in Sublime Text, but there are still some errors, since it was incredibly garbled. This file will be converted into a Google Sheet titled "Macomber Canonical Stories." Here is the structure of the text file:

- **MAC###:** Macomber Marian Miracle Canonical Story identifier (some of these stories are various parts of the same story, so the identifier for a story may be something like MAC034A)
- **Title:** English Title from Macomber
- **Text**: Secondary source that discusses this particular Canonical Story
- **English Translation:** Translation of that story as it appears in one or two manuscripts (unfortunately, this is not currently a standard category in the structured file because most entries don't have English translations; not sure if this needs to be made standard in all entries before transferring to Google Sheets)
- **PEth**: Shelf number, beginning folio, and ending folio for where this particular Canonical Story appears in Princeton's RBSC *Täʾammᵊrä Maryam* manuscripts (shelfmark and folios needs splitting out). This field is often empty.
- **EMIP**: Shelf number, beginning folio, and ending folio for where this particular Canonical Story appears in the EMIP digital repository. This field is often empty.
- **MSS**: Shelf number and beginning folio only for where this particular Canonical Story appears in other repositories. Macober uses abbreviations for these (see list)
- **EMML**: Shelf number and beginning folio only for where this particular Canonical Story appears in the HMML digital library.
- **Keywords**: keywords from Macomber handlist with a few additional ones that Belcher and Lambrinaki came up with (problematic and needs to be updated, or categories missing)
- **Incipit**: Current text in this field is garbage and will be deleted. It will be globally replaced with Brown's list of incipits, matching up using the Macomber identifiers.

Earlier data work. Some earlier work was done correlating tales and manuscripts and themes in Excel (such as matching Princeton manuscripts to Macomber based on data from the Princeton finding aid), but those files are now out of date.
Create a data structure. We designed a preliminary data structure for Canonical Stories, Story Instances,, as well as for manuscripts and incipits.

Planned FY20 Data Work

Over the next year, we will:
- **Migrate data.** Project data (namely, Macomber's handlist) currently managed as a structured text file will be migrated to Google Sheets by the CDH team, structured in multiple sheets based on planned data architecture (see Technical Design Plan section on Preliminary Data Structure, below) with data validation to share information across sheets.
- **Enhance data**. Project data will be enhanced by the CDH team by importing data provided from other teams (e.g., those from the Hamburg BM project and Oxford CSM database).
- **Match data.** Use Oxford CSM story data to identify Canonical Stories that came out of Europe and the Middle East, not Ethiopia (to distinguish between foreign and indigenous).
- **Configure data validation.** Connect different sets of data without duplicating information (i.e., Gəʿəz Marian miracle manuscripts and Gəʿəz Marian miracle tales)
- **Develop a controlled vocabulary.** Wendy and Evgeniia work with Macomber, Oxford CSM, Hamburg BM, and Index on Medieval Art's controlled vocabularies to develop a controlled vocabulary for PEMM Canonical Stories.
- **Develop a simple incipit tool (or outsource).** Create a tool for searching Macomber's standardized incipits so that research assistants can catalog manuscripts, preferably a dialog box in Google Sheets. Two challenges are homophones and recensions. The tool for searching must account for homophones, treating certain fidal letters as exchangeable (say ʋ and ጎ) because scribes easily substitute one for the other. Also, when searching incipits, research assistants must be careful about using proper noun searches. That is, a Canonical Story may give a name for the main character (e.g., Barok from Finqe [Phoenicia]), but a Story Instance in a particular manuscript may refer to him only as a "deacon" or as a "sinner."

Data Standards and Capture Procedures

- **Controlled vocabulary**. We are developing and using controlled vocabulary lists (see above).
- **Abbreviations.** We are developing and will use abbreviations for repositories (e.g., BN not Bibliotheque Nationale). The data currently in the structured text file uses codes for repositories holding Ethiopic manuscripts. The repository abbreviations will be used to generate brief repository records in the new Google Sheets data structure, which will be used to document current locations of materials.

- **Organize data.** We use Google Drive and Google Docs and Sheets to organize the project. After migration to Google Sheets, data will be stored in multiple sheets of a single document in order to allow data validation and autocomplete-style lookups for related data. A regular, automated export will be set up to export Google Sheets data to a GitHub repository. We use Slack to communicate about it.
- **Validate data.** In some instances, we have multiple researchers typing the same data, or cataloging the same manuscripts, to check accuracy.

Grant Year Objectives – Data

Main Outcomes/Deliverables:
- Dataset of Canonical Stories (publishable and citable)
- Dataset of 100+ MSS cataloged using list of Canonical Stories (publishable and citable)
- Import relevant data from Princeton Ethiopian manuscripts to Google Sheets
- Documentation of the new data structure
- Documentation of data entry workflow & processes[3]
- Cataloging of the MSS to provide to Hamburg BM project
- Linking to Hamburg BM IDs
- Export data from Google Sheets (possibly as XML) to share with Hamburg BM

Possibly in Scope – Data:
- Linking PEMM Canonical Stories to other project by cross-linking identifiers from Hamburg BM, Oxford CSM database project (European stories)
- Linking to Index of Medieval Art
- Linking MSS records to Princeton University Libraries (PUL) digital editions in DPUL (Treasures of the Manuscript Division)
- Linking MSS records to EMML, if the digital edition is available

Out of Scope – Data:
- Cannibal of Khmer project with The Textual History of the Ethiopic Old Testament (THEOT)
  - 90 MSS with the story carefully selected from different times and regions and typed to run through the software to compare versions. Won't be done until October or November, but then all of that data will be available. May provide interesting information to help with this work but out of scope for this phase.
- Describing characters, objects, places, and other subject matter that occur in stories
- Annotating PUL digitized images or crowd-sourcing

---

[3] CDH developers expect to write code to do the migration, but the code itself is not a deliverable sinces it's a means to an end and not something we are likely to reuse or generalize.

Project Needs – Data

- Research assistants with the language skills (either reading-level or at the level of recognizing the letters)
- Data structure for Google Sheets (see Technical Design Plan section on Preliminary Data Structure, below)
- Incipit tool
- Structured text file migrated into Google Sheets with customized data validation and formatting, including:
  - autocomplete on incipits with homophone search functionality, for matching particular stories to Canonical Stories
  - data validation for field types as appropriate, e.g. numeric values or sequential / increasing numbers for folio numbers within a manuscript.
- Regular check-ins with CDH staff including dev team and project management time.
- Training and support to query the data (e.g., OpenRefine training)

Concerns – Data:

Risks

- Possible overlaps with Hamburg BM; will use their data where possible, and hope to supply ours back to them to be incorporated, but want to avoid duplication of effort. In some cases we may use their data for checking and comparison. We have consulted them with this concern and they are willing to assist.
- Handling Fidel in Google Sheets and in search. One challenge is that the data will use Latin letters with diacritics (about 20 Unicode characters, like ə, ṭ, ṣ, ś, ä, ə, w, ḍ, ǧ, ḥ, ḫ, ḵ, č, ñ), as well as over 300 Ethiopic fidəl characters (also available in Unicode, like ወ፣ቀ፣ምፘተ).
- Different types of software that researchers use to handle and input Fidel
- Finding students with necessary language skills
- Changes with the scale of data may make Google Sheets unusable

Interdependencies

- Sharing / linking to data without other projects
- Manuscript access (digital or otherwise)
  - Firestone Library began digitizing its Gəʿəz manuscripts; prioritizing the *Täʾammərä Maryam*. All ten are now digitized and online in Digital PUL.

Data security consideration**s**

- Structured text file is currently stored and shared via DropBox; a copy has been added to the PEMM Google Team Drive for backup.
- This project does not include any personal or sensitive data

Data management plan

- After data migration, Google Sheets will be the primary canonical data source; GitHub will be a secondary source for backup and experimentation.

Long-term preservation plan

- A released version of data exported from Google Sheets will be deposited with Zenodo or other repository for long term secure storage, and also to make it citable.
- Where appropriate, data will be exported and shared with other relevant projects.

Future Plans

Future Data Work

- **Write precises of Canonical Stories.** A huge and difficult task will be writing short summaries of the 700+ indigenous Canonical Stories. Only those with an excellent understanding of Amharic, French, or Ge`ez will be able to do this work. Maybe 100 can be done from stories available in English translation. Or, perhaps, if the keywords are good enough, no precis is needed?
- **Track word length of Story Instances.** This is a way of getting at the possibility of different recensions.
- **Tag Canonical Stories with keywords.** Another difficult task will be using the controlled vocabulary list to better tag Canonical Stories. Macombere did tag 642 of them with keywords, but many remain and his list can be improved. Only those with a good level of Ge`ez and English will be able to do this work.
- **Identify new Canonical Stories.** We need to give new identifier numbers, titles, themes, and incipits for Canonical Stories not in Macomber. We will use Hamburg BM identifiers where possible, but may need to do this a bit ourselves.
- **Translate into Amharic.** Translate titles, keywords, and website into Amharic.
- **Design and write static website.** This can be done in the last year.
- **Compare Cannibal of Qemer transcripts**. Once Steve, Jeremy, Jonah, and Ashlee complete typing up all 90 versions of the Cannibal of Qemer tale, we will do computational analysis.

Eventually, PEMM would like to answer the following question:

- **How did individual Ethiopian Marian miracle tales change over time and region?** PEMM is conducting a textual history of just one of the tales, called the Cannibal of Qəmər. We already know it has three quite different recensions, but we are trying to determine how, where, and when the tale differs. With new denogram comparative software, we can begin to establish recensions and compare them statistically. To do this, we are collaborating with The Textual History of the Ethiopic Old Testament (THEOT) Project, which has worked out the methods and workflow necessary to carry out textual histories of Gə ̔əz texts.

Part III: Grant Year Plans - Interface

Grant Year Objectives – Interface

Main Outcomes/Deliverables:
- Preliminary interface (prototype):
  - Feed from Google sheets → Github
  - Simple data viz including map
  - Showcasing images from PUL MSS via IIIF
  - Lightweight search and browse
  - Article or post to accompany visualizations

Possibly in Scope – Interface:
- Support for multilingual site capacity
- Preliminary web interface design

Project Needs – Interface:
- Geographic data to generate a map
- Script to pull data from Google sheets into static website

Concerns – Interface:

Risks
- Front-facing deliverables are dependent on progress in data work
- Working with fidel and amharic (multilingualism)
- Community push-back on making data on certain stories accessible and visible

- Prototyping with static website technologies may have more constraints than we expect; CDH development team does not have substantial experience with static site technology
- Changes with the scale of manuscript data may impact static site performance
- If we shift from a static site to a database driven site in a future phase of the project, there are likely to be changes in the site structure and architecture

Interdependencies

- Hamburg BM project
- Working with PUL IIIF + digitized content, but not annotating
- The Textual History of the Ethiopic Old Testament (THEOT) Project
- Rights on existing code for searching fidel

Future Plans – Frontend:

- Distribution of static site with data and images on thumb drives
- Assess prototype and data work to determine how to expand, e.g. database driven site
- Apply to follow up Research Partnership grant for next phase of project development

Part IV: Technical Design Plan

Data

In this phase of PEMM, we will migrate the data from the semi-structured text file into a Google Sheets spreadsheet comprised of multiple structured and related sheets with data validation configured to automatically connect data between different sheets within a single Google Sheets spreadsheet. The data for this project is highly relational and certainly could be implemented as a relational database, but given the phase of this project and the amount of data work still to be done, we chose Google Sheets. Sheets supports edits by multiple concurrent users and tracks versions, and working within a spreadsheet will allow project team members to manage and query the data more easily without developer intervention. We will model the data as if for a relational database (see Preliminary Data Model), but implement it in Google Sheets with an eye towards data entry usability and efficiency rather than a fully normalized database structure.

As a first step, we will prototype the Google Sheets structure based on the new data architecture and determine the appropriate data validation, formatting, and any other configuration that is useful and necessary. This will also give us a chance to experiment with Fidel characters to make sure we can get everything working as expected. We expect to write one piece of custom code in Google Apps Script to support homophone searching in Fidel[4] for an incipit lookup, which will

---

[4] The public search interface for Hamburg BM project (https://betamasaheft.eu/as.html) includes an option for homophone searching, and their help text includes a list of orthographic variants. We will use their implementation as a reference and their list of characters as a starting point for our implementation.

enable project team members to match stories in a manuscript with Canonical Stories from the Macomber catalog based on unique words or phrases.

A prototype incipit lookup has already been created by Steve Delamarter and demoed to the team which we could purchase, but we prefer to design and implement something simple based on project needs. This will give us in-house expertise to maintain and support the tool, and we can iteratively refine it if necessary. We also plan to document as an example for others using Google Sheets for dataset work. If implementing the incipit search proves to be more difficult than anticipated, we will consult with Garry Jost or purchase the prototype.

Once the Project Director has agreed to the data structure and tested and accepted Google Sheets functionality, CDH developers will write a script to parse the structured text file and convert into multiple CSV files for import into Google Sheets, which will create preliminary records for Manuscripts, Canonical Stories (i.e. those cataloged by Macomber) and Story Instance (a story as it occurs in a manuscript). Records for archival repositories that hold these manuscripts will be added manually to the spreadsheet by project researchers, since there are a small number and the structured text file does not supply the needed information.

After data is migrated into Google Sheets, that will become the canonical data source for the project, and the project researchers can begin working on the data. After the migration is complete, CDH developers write a script to generate a regular, automatic export of the Google Sheets as CSV and/or JSON which will be added to a GitHub repository. The data in the GitHub repository will serve as both a versioned backup and as a data source for querying, visualization, and a prototype interface; it may also eventually be used to publish a citable version of the data via Zenodo or a similar service. The export will be powered by the "publish to web" functionality available in Google Sheets, if it is sufficient; otherwise it will be implemented with an existing Python Google API client to access the data. Additionally, if the Google APIs allow it without too much difficulty[5], we will use revision information to credit the project team members who have made edits to the data as co-authors of the commit using the GitHub co-author syntax[6], as a way of making the contributions of project team members a visible part of the record of the data.

We may also implement continuous validation and reporting on the data in GitHub, making use of continuous integration tools that are usually applied to software code in order to automate regular data validation.

Interface – Prototype Website

If time permits, we will develop a prototype website as proof of concept which will allow us to experiment, try new technologies, get familiar with the data and working with Fidel characters. The ultimate goal is to know enough to decide how to proceed in the next phase of the project.

---

[5] Documentation on the Google Drive API indicates this should be possible (https://developers.google.com/drive/api/v3/reference/revisions), but it's unclear how difficult it is.
[6] See https://help.github.com/en/articles/creating-a-commit-with-multiple-authors

The GitHub data repository generated from the Google Sheets data will be used as a starting point for experimentation, creating a prototype static site⁷ which could allow the project team to browse and search the data, and will give the development team a chance to become familiar with the data and working with Fidel characters. We have chosen to work with static site technology because it should allow for quick prototyping and experimentation based on the data from the Google Sheets without making a heavy investment in a particular technology stack for the next phase of the project.

We hope to experiment with the following static site technologies:
- Jekyll (https://jekyllrb.com/; implemented in ruby and commonly in use for Digital Humanities projects)
- Hugo (https://gohugo.io/ implemented in Go; newer and more powerful than Jekyll)
- Gatsby (https://gatsbyjs.org/ implemented in JavaScript)
- Wax (http://marii.info/projects/wax, software for generating exhibit sites with IIIF, spreadsheets, and Jekyll)
- Elasticlunr.js (http://elasticlunr.com/, browser-based searching)

The prototype website will be implemented with a responsive design that supports mobile devices by choosing an existing theme to allow us to focus on the more innovative aspects of the project. Creating a site that is usable, accessible, and welcoming to the diverse audiences for this project will require user research, but because this is a prototype website we may begin that research during the project year to guide later phases of the project. The static site will be hosted on GitHub pages as we prototype. If we determine we want a Princeton URL for the prototype site before the end of the current grant phase, we will request a hostname and possibly a virtual machine from OIT.

Alongside the static site development, dependent on data delivery, CDH developers and UX Designers will contribute to data visualizations and maps as appropriate to help answer the project research questions. These may or may not be part of the static site; they may be included in an essay to be published at the end of this phase of the project.

If time allows, CDH developers will experiment with internationalization, with the hope of making the prototype site available in both English and Amharic. This is not only something we're technically interested in, but also something we feel ethically challenged and compelled to do, based on Wendy Belcher's comment that no Ethiopic manuscript materials or data are currently available in Amharic. There are existing solutions for multilingual sites implemented with Jekyll, including notably The Programming Historian, in the Digital Humanities space.

---

⁷ Any static site code committed to GitHub will be put in a separate repository from the data to allow the data to be easily deposited with Zenodo without including static site software or content.

We are also interested in leveraging minimal computing principles and techniques to provide a version of the static site as a standalone package, including project data and any image content with permissions that allow redistribution, to be shared and distributed via alternate means, such as USB drive or inexpensive hardware.[8]

Preliminary Data Model

Princeton Ethiopian Miracles of Mary Project – Data Model
July 03, 2019

Contributors
Wendy Belcher, Rebecca Sutton Koeser, Rebecca Munson, Evgeniia Lambrinaki, Gissoo Doroudian
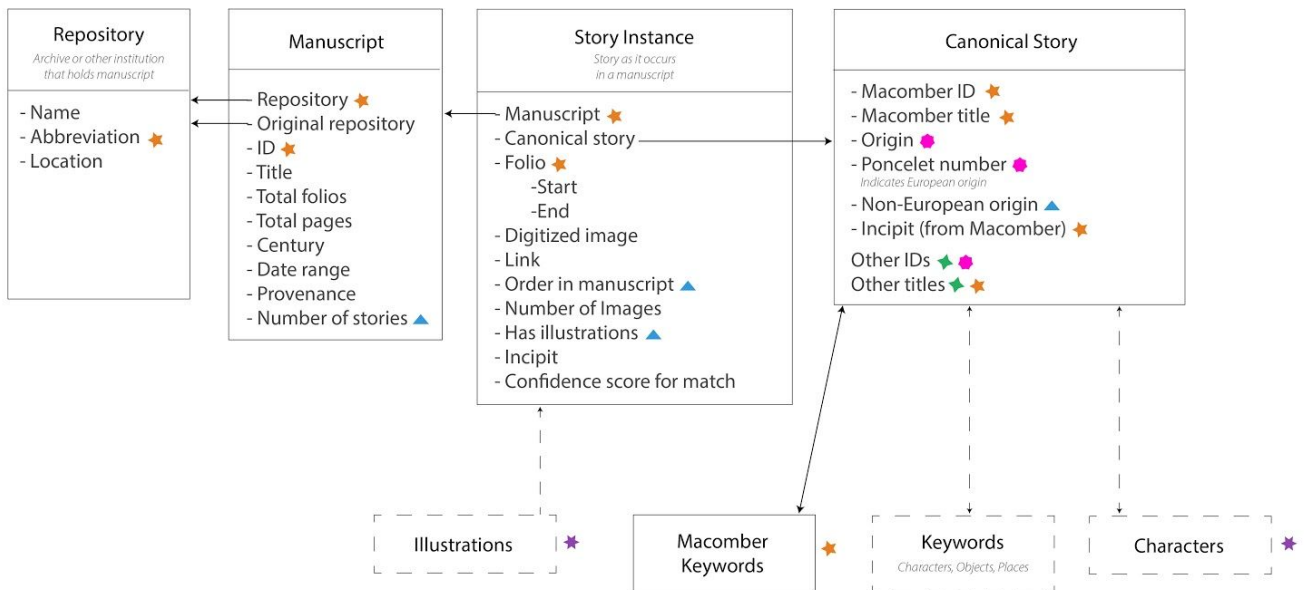
Part V: Deliverable Timeline

Summer 2019 (by September 1):

Data:
- Documentation & diagram describing data structure

---

[8] For example, see Ed Summer's post about building an offline static site with React
https://inkdroid.org/2018/01/10/offline-react/

- Google sheets spreadsheet that implements data structure
- Data from Sublime Text file migrated to Google Sheets

Interface:

- None

Fall 2019 (through December 31):

Data:

- Incipit lookup with phonetic searching
- PUL finding aid data added to existing Google Sheets
- BM data added to existing Google Sheets
- Oxford CSM data added to existing Google Sheets
- Trained student researchers (at least orient to MSS and project, orient to Google Sheets; more if data structure is ready)
- Automated feed from Google Sheets to a GitHub repository
- Continuous integration for data validation on GitHub (nice to have)
- 40 Story Instances catalogued

Interface:

- None

Spring 2019 (through the end of the grant period)

Data:

- Data visualizations
- 20 manuscripts catalogued

Interface:

- Prototype static site, if sufficient data is available and time permits

Part VI: Grant Year Wrap-up

This section is completed after the grant year concludes. Describes the goals reached and outcomes of the project, and explains major changes and discrepancies with planned work. Continuing projects will include this in the charter for their subsequent project phase. Completed projects must provide to the CDH Project Coordinator within one month of the conclusion of the grant period.

Part VII: Agreement

Project pause policy

To ensure that all projects receive sufficient and equitable development time, time-sensitive queries and requests must be addressed within two weeks of initial (email) request. The PI is responsible for communication with development team. If PI is does not respond to a task that has been indicated as time-sensitive by the CDH team within 2 weeks of initial request, further project development will be paused until the project can be reasonably integrated back into the CDH development schedule.

Rights, Permissions, and Attribution

Site content and data will both be licensed under Creative Commons Attribution 4.0 International (CC-BY 4.0). If any of the datasets consist solely of factual data where authorship cannot be claimed, they will be licensed as CC0.

Any software developed by CDH that merits release will be licensed under Apache 2.0. The Technical Lead will fill out an invention disclosure form in order to gain approval from the Office of Technology Licensing in order to release the code. Before approval is granted, the code will be owned by the Trustees of Princeton University.

Web Presence and Project Publicity

The PI will create a project page on the CDH website, keeping it up-to-date and accurate during the grant year. The Project Manager will submit at least two blog posts per year, to be published on the CDH website. The schedule for publishing blog posts will be determined in consultation with CDH staff. In the case of a public site launch, or similar event, the PI and PM will work closely with CDH and Princeton University Library staff as needed on publicity, communication, and outreach. Currently, the PI has a web page for the project at https://wendybelcher.com/african-literature/pemmproject/

Credit

All team members will be credited on the project's website and CDH project page. The project's website will include a sponsorship statement (indicating the CDH as well as any other supporting groups, departments, agencies) and will include a citation statement indicating how the project assets should be cited. The site will also list and link to other projects that contributed data.

Project PI

_____

CDH Faculty Director

_____

Date:

_____

Appendix A: Relevant Resources and Projects

Data

Currently, the project data exists in seven separate, uncorrelated formats:

1. Macomber Handlist of Marian Miracles in the Ethiopian tradition with identifiers for each of 642 Canonical Stories, translations and analyses, and its keywords, and shelfmarks and folios for Story Instances in about 200 manuscripts).
2. Brown's list of incipits for Macomber's Canonical Stories: (typed up in fidäl and available in Google doc or sheet; checked by Hamburg BM for accuracy; used to catalog stories in manuscripts.)
3. Oxford CSM database for controlled vocabulary for themes (700+ terms, which need to be cleaned up and merged with the Macomber Handlist themes, which need to be updated,) (available in text file.)
4. Princeton RBSC finding aid for its ten *Täʾammərä Maryam* manuscripts (not yet cataloged with the Macomber Handlist identifiers,) (available in EAD XML). The project will also reference data and images from PUL digitized editions of relevant Ethiopic manuscripts held by PUL, managed by PUL and provided via IIIF Presentation and Image APIs .
5. Hamburg BM identifiers and cataloging data for about 75 *Täʾammərä Maryam* manuscripts (cataloged with the Macomber Handlist identifiers,) (available in TEI XML.)
6. Delamarter's list of 90 manuscripts across six centuries (1400s, 1500s, 1600s, 1700s, 1800s, 1900s) and five regions (north, south, central, east, west Ethiopia) (in Google Sheets, uncataloged) at EMIP / HMML
7. Archives of manuscripts from EMML, Bibliotechque Nationale, the British Library (none cataloged with Macomber handlist).

Archives and Databases

Lots of information has been collected about the *Täʾammərä Maryam*, but is spread across hundreds of obscure print catalogs, French and German articles, Italian books, and Gəʿəz monasteries. Each uses different numbering systems and tale titles, almost none have keywords, and many catalogs do not enumerate the tales in a manuscript, rather simply stating that something is a *Täʾammərä Maryam* and moving on. Thus, a crucial aspect of PEMM will be collating information using the following archives and databases.

Macomber Handlist

The most important PEMM resource is William Macomber's unpublished handlist of 642 Marian miracles, based on his study of 100+ manuscripts, including each story's title,

translations, themes, and incipit (the unique first sentence of each story; used with medieval manuscripts as an identifier, as they have no titles). (This was quite an extraordinary accomplishment, before digital work in the humanities was common. It gives PEMM a huge leg up.)

- Macomber, William F. n.d. [1980s]. *[Handlist of] The [Ethiopian] Miracles of Mary*. Collegeville, MN: Hill Monastic Museum and Library, St. John's Abbey and University.

Lombardi Handlist

Chiara Lombardi disagrees with Macomber and thinks there are only 530 Canonical Marian Miracle Tales. Where available, and depending on time, we may include her identifiers in addition to Macomber's.

- Lombardi, Chiara. 2009. "Il Libro etiopico dei Miracoli di Maria (The Ethiopic Miracles of the Blessed Virgin)."BA thesis, Archeology, Università di Napoli.
- Lombardi, Sabrina. 2010. "Miracoli di Maria."MA thesis, Anthropology, Corso di laurea in lettere, Università di Firenze.

Archive Catalogs

An indispensable source for PEMM are extant catalogs of the *Täʾammərä Maryam* manuscripts. Macomber worked with about fifteen (see list of abbreviations). While most catalogs do not use Macomber's numbering, these catalogs often provide manuscript dates, provenance, and folios of each story. The most important catalogs are as follows:

1. **Princeton University Rare Books and Special Collections.** Belcher and Qesis Melaku catalogued the RBSC's ten *Täʾammərä Maryam* manuscripts, although they did not use Macomber's numbering for their stories. ll of ten of these manuscripts have been digitized and are available online.
   a. Treasures of the Manuscripts Division, Ethiopic manuscripts https://dpul.princeton.edu/msstreasures/catalog?f%5Breadonly_collections_ssim%5D%5B%5D=Ethiopic+Manuscripts&q=
   b. Melaku Terefe, and Wendy Laura Belcher. 2009. Princeton Collections of Ethiopic Manuscripts, 1600s-1900s: Finding Aid. Princeton, NJ: Princeton University Library, Department of Rare Books and Special Collections, Manuscripts Division. https://findingaids.princeton.edu/collections/C0776
2. **Hill Museum and Manuscript Library (HMML), St. John's Abbey and University, Collegeville, MN.** This is a digital library; that is, it archives microfilms and digital images of manuscripts that are elsewhere. A huge project of HMML in the 1960s and 1970s was the Ethiopian Microfilm Manuscript Library (EMML), which microfilmed 8,000 manuscripts in monasteries and churches in Ethiopia. William Macomber, author of the handlist, was one of the directors for this project, and his handlist used many EMML manuscripts. HMML also hosts other digital collections of Ethiopian manuscripts, such as the Ethiopian Manuscript Imaging Project (EMIP, by Stephen Delamarter). At least 22 *Täʾammərä Maryam* are currently available for free online, but HMML has over 535 *Täʾammərä Maryam* manuscripts awaiting transfer

from microfilm to online digital form. In general, they only digitize microfilm if a client wants access to it and pays for it. Anyone can access the microfilm for free, but of course only on site, in Minnesota. Belcher is in correspondence with them about gaining digital access to more of their manuscripts. Developing a relationship with HMML may be useful to PEMM more broadly as they have digital copies of over 250,000 handwritten manuscripts in many languages from around the world.

   a. One can search parts of this collection at https://www.vhmml.org/readingRoom/

3. **British Library**. This archive has eighteen of the most splendid *Täʾammərä Maryam* manuscripts in existence, most looted from the royal scriptorium. Manuscripts from the center of power, from the clerics of the royal house, will be most useful for establishing the canon of stories, as they would be the ones to make it. The regions will have their own interesting trends; and perhaps have their own standard collections. Both are interesting, but tracking what is canon is vital. This collection has been catalogued and recently digitized, but has not used Macomber's numbering.

   a. http://www.bl.uk/manuscripts/BriefDisplay.aspx?source=advanced
   b. Wright, William. 1877. *Catalogue of the Ethiopic Manuscripts in the British Museum Acquired since the Year 1847*. London: British Museum.

4. **UCLA Library.** Belcher and Qesis Melaku also catalogued UCLA's collection, but before a big acquisition. It has very few *Täʾammərä Maryam* manuscripts, however.

   a. http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0009gx3x

5. **Catholic University of America, Institute of Christian Oriental Research (ICOR).** Its catalog of its Ethiopic manuscripts is not yet online, but it has 12 *Täʾammərä Maryam* manuscripts.

   a. Weiner Codex 233 – EMIP 2175 (17th cent.?): 178 miracles
   b. Weiner Codex 260 – EMIP 2259 (1930–1974): 5 miracles
   c. Weiner Codex 308 – EMIP 2340 (20th cent.): 11 miracles
   d. Weiner Codex 335 – EMIP 2370 (18th cent.): 80 miracles
   e. Weiner Codex 364 – EMIP 2399 (20th cent.): 31 miracles
   f. Weiner Codex 366 – EMIP 2401 (20th cent.): 81 miracles
   g. Weiner Codex 370 – EMIP 2405 (19th cent.): 3 miracles
   h. Weiner Codex 395 – EMIP 2660 (20th cent.): 3 miracles
   i. Weiner Codex 403 – EMIP 2668 (20th cent.): 36 miracles
   j. Weiner Codex 428 – EMIP 2716 (18th cent.): 3 miracles
   k. Weiner Codex 449 – EMIP 2737 (18th / 19th cent.): 34 miracles
   l. Weiner Codex 463 – EMIP 3238 (19th cent.?): 45 miracles

6. **Other Archives.** Many other archives exist in Europe and North America, including the Bibliotheque Nationale, the Vatican Library, St Petersburg Library, and so on.

   a. Platt, Thomas Pell. 1823. *A Catalogue of the Ethiopic Biblical Manuscripts in the Royal Library of Paris, and in the Library of the British and Foreign Bible Society: Also Some Account of Those in the Vatican Library at Rome, to which are Added, Specimens of Versions of the New Testament Into the Modern Languages of Abyssinia*. London: R. Watts.

Oxford Cantigas de Santa Maria (CSM) database

This database, created by Stephen Parkinson of the Oxford University Centre for the Study of the *Cantigas de Santa Maria* (manuscript) and PI for the Oxford CSM database launched in 2005, is the most authoritative on the world-wide collection of Marian miracle stories, but based largely on manuscript collections in Europe. The Oxford CSM database is "designed to give access to a vast range of information relevant to the processes of collection, composition and compilation" of the Marian miracle stories. It provides a fully searchable electronic version of Poncelet's list of Marian miracles (*Index miraculorum B.V. Mariae quae saec. VI-XV latine conscripta sunt*); brief descriptions of all European Marian miracles; and a controlled vocabulary list for Marian miracle story themes. The principal investigator is Stephen Parkinson, who is eager to share any data he has, in return for better information about the Ethiopian Marian miracles for his database. This database will help us to identify which stories originated outside of Ethiopia and to develop our own controlled vocabulary theme list.

- http://csm.mml.ox.ac.uk/

Beta Maṣāḥəft (Hamburg BM) Database

This database project is hosted by the Hiob Ludolf Centre for Ethiopian Studies at the Universität Hamburg in Germany and with Prof. Alessandro Bausi as the principal investigator. It is a very long-term project (from 2016–2040) with large German government funding to create a "virtual research environment" for collecting and managing data about "the predominantly Christian manuscript tradition of the Ethiopian and Eritrean Highlands." They are also using Macomber's Handlist to catalog Marian miracle stories, although they are not currently tagging themes. They are also collating data on manuscripts (date, provenance, total folios), developing controlled vocabulary lists for people and places, and arriving at their own identifier for Canonical Stories (since Macomber missed some). They are open access and so we will be exchanging information as much as we can: the Hamburg BM giving PEMM its data in accessible forms and PEMM giving the Hamburg BM whatever data we create.

7. Beta masāhǝft: Manuscripts of Ethiopia and Eritrea (Schriftkultur des christlichen Äthiopiens und Eritreas: eine multimediale Forschungsumgebung)
8. "See the section on contributing and reusing data below."

Miracula Mariae project

This comparative project, with principal investigators Ewa Balicka Witakowska and Anthony John Lappin, is also called "Miracles of the Virgin: Medieval Short Narratives Between Languages and Cultures." Begun in 2015, it will compare six to ten individual Marian miracle stories across many languages and regions as a way of studying transmission (in image and text) (in Arabic, Armenian, Croatian, Dutch, Ethiopic, French, Georgian, Greek, Hungarian,Latin, Middle English, Old Icelandic, Old Swedish, Polish Italian, Romanian, Slavonic, South Slavonic, Spanish, Syriac, and Ukranian). It will also compare sociological and psychological aspects of

the stories in different cultural contexts. The eventual aim is a database, but there is no evidence of this having been initiated yet.

> "The overall tradition [of Marian miracle stories] offers a rich and vast body of literature, which, in its totality, has not been studied, and whose intertextuality offers a number of interesting problems and resources for further study. … The project will seek to analyse this complex set of interrelated traditions from three successive standpoints. The <u>first</u> will consider manuscript transmission and the physical distribution of miracle-tales; the <u>second </u>will compare collections and versions, in order to understand the cultural pressures that led to variation and re-elaboration of a set number of miracle-tales; and the <u>third </u>will look at the resulting texts from a narratological point of view, and aim to establish the limits and development of a story within a primarily manuscript culture. The first stage of the project will be a<u> text-critical study of a selected number of miracles</u> from the core collection, tracing their development across manuscripts, enabling sub-families and recensions to be established, and allowing the evolution of the collections to be precisely identified."

- https://hildefonsus.wordpress.com

Index on Medieval Art

This Princeton institute may have useful controlled vocabularies.

Appendix B. Planned FY20 Data Steps

Step 1. Create PEMM Canonical Stories Dataset. The Macomber handlist will be used as the basis for a <u>Google Sheet titled PEMM Canonical Stories Dataset.</u> It will have more data than the Macomber one. Below is its format (with invented data for one story to give a sense for the appearance of the data):

- PEMM miracle tale **identifier**: MAC0007
- Hamburg BM identifier: LIT3640Miracle
- Lombardi miracle tale identifier: 53
- Oxford CSM miracle tale identifier: 5 [From BM]

- Macomber **title** of Marian miracle tale: The monk of Dabra Qaləmon who did not fast.
- Lombardi/ Cerulli title of Marian miracle tale: Il monaco che non ha digiunato
- Dillmann title of Marian miracle tale: Monachi non ieiunant
- EMIP title of Marian miracle tale: The monk who did not fast
- Oxford CSM title of Marian miracle tale: none
- Colin title of Marian miracle tale: Le moine qui n'a pas jeûné
- Tsegaye title of Marian miracle tale: ያልበስበው መነኩሴ

- Budge miracle tale **edition/text** [translator, title, page]: Budge, *Hundred*, item 86; Budge, Miracles, p. 42.

- Tsegaye miracle tale edition [translator, title, page]: None
- Tasfā Giyorgis edition [translator, title, page]: Tasfā Giyorgis *TM*, item 77, page 274-275.

- English **translation** of tale [translator, title, page]: Budge, *Hundred*, p. 87.
- French translation of tale [translator, title, page]: Colin, *TM*, p. 10
- Amharic translation of tale [translator, title, page]: Tsegaye, *TM*, p. 402
- Italian summary of tale [translator, title, page]: Cerulli, Il libro, p. 166.

- Princeton Ethiopic **manuscripts** with the tale and its beginning and ending **folios**: 47 (23r-24v)
- Other repositories known manuscripts with the repository name, shelfmark, and beginning (and sometimes ending) folios: G-7; ZBNE 60-29; 61-27; CRA 52-35; 54 (91r); 55 (91r); SBLE 32-28; BM 2-31; 3-36; VLVE 267 (56v); 298 (23r, 50v); SALE 23-27; 43-29; LUE 30-40; 32-28; CBS-28; CCBE 951-28; AECE 1 (30r).
- EMIP (Ethiopian Manuscript Imaging Project) digitized manuscripts with the tale and their beginning and ending folios: [to come]
- EMML (Ethiopian Microfilmed Manuscript Library) digitized manuscripts with the tale and its beginning and ending folios: 2275 (194r); 6938 (29v); 5520 (31v); 2378 (20v); 6640 (49r); 2060 (195r); 2066 (31v, 135v); 2802 (22v); 6196 (81r); 7543 (17v).

- **Keywords/themes** (controlled vocabulary list TBD):

- Story length (how many characters or words): 1,250
- Story number of total stories in manuscript (order): 231
- Story precis (100 words or fewer):
- Story translation (if in public domain):

- Story Instance **illustrations** no.: 4
- Story Instance illustrations characters: farmer, abbott
- Story Instanceillustrations objects: bow and arrow
- Story Instance illustrations dating (if later): same

- **Incipit** 1 with manuscript shelfmark (imported from Brown): ወሀሎ፡ በደብረ፡ ቅዱስ፡ ዐቢይ፡ አቡ፡ ሳሙኤል፡ ዘቀልሞን፡ ቤተ፡ ክርስቲያን፡ ሥናይት፡ በስመ፡ እግዝእትነ፡... ወኮነ፡ ውስተ፡ ዛቲ፡ ቤተ፡ ክርስቲያን፡ ስዕል፡ ዐቢይ፡ ወመንክር፡ (6938).
- **Incipit 2** with manuscript shelfmark: ወሀሎ፡ አሐዱ፡ ብእሲ፡ መነኮስ፡ በደብረ፡ አቡ፡ ሳሙኤል፡ ዘቀልሞን፡ ወያፈቅራ፡ ለእግዝእትነ፡ ማርያም፡ ወያነብቦ፡ ወትረ፡ ተአምኖተ፡ መልአከ፡ ሌሊተ፡ ወመዐልተ። ወዝንቱሰ፡ ብእሲ፡ ኢይጸውም፡ ወኢይጼሊ፡ ወይትሜሰል፡ ከመ፡ አብድ፡ ወእንቡዝ፡ (2378).

Macomber's abbreviations for repositories are as follows (note that some repositories appear twice because Macomber is using catalogues of collections, not the actual collections):
- AECE = Abbaye d'En Calcat, Dourgne, France (but microfilmed for HMML)
- CBS= manuscript of the Berlin Staatsbibliothek (described by E. CERULLI)

- CCBE = manuscripts of the Chester Beatty Library (described by E. CERULLI)
- CF= manuscripts of the Biblioteca Nazionale of Florence (described by E. CERULLI)
- CL= manuscript of the Academy of Sciences of Leningrad (described by E. CERULLI)
- CRA= manuscripts of the d'Abbadie Collection of the Bibliotheque Nationale in Paris (described by Conti Rossini)
- DULE = Ethiopian manuscripts of the Duke University Library, Durham, North Carolina (but microfilmed for HMML)
- EMML= Ethiopian Manuscript Microfilm Library, of Hill Monastic Manuscript Library (HMML), St. John's Abbey and University, Collegeville, Minnesota
- G = manuscript of the Biblioteca Giovardiana in Veroli (described by E. CERULLI)
- GBAE = manuscripts of the Biblioteca Ambrosiana in Milan (described by S. Grebaut)
- GVE = manuscripts of the Vatican Library (described by S. Grebaut and E. Tisserant)
- HBS = manuscripts of the Staatsbibliothek in Berlin (described by E. Hammerschmidt)
- LUE= Ethiopian manuscripts of the Uppsala University Library (described by O. Lofgren)
- SALE= Ethiopian manuscripts of the Conti Rossini and Caetani Collections of the Accademia Nazionale dei Lincei in Rome (described by S. Strelcyn)
- SBLE = Ethiopian manuscripts of the British Library (described by S. Strelcyn)
- SGE = Manuscripts of the Griaule Collection of the Bibliotheque Nationale in Paris (described by S. Strelcyn)
- SWE = Ethiopian manuscripts of the Seabury-Western Theological Seminary (described by W. F. Macomber)
- VLVE = Ethiopian manuscripts of the Vatican Library (described by A. Van Lantschoot)
- WBLE = Ethiopian manuscripts of the British Library (described by W. Wright)
- ZBNE = Ethiopian manuscripts of the Bibliotheque Nationale in Paris (described by Zotenberg)

Step 2. Create keyword fields for PEMM Canonical Stories Dataset. When consultants who read Gəʿəz start cataloging stories, they will need access to the controlled vocabularies through dropdown menus for many keyword/theme fields. Although this particular type of cataloging task will not be done until FY21, we want to be aware of the types of fields we will need in the future. For now, the dataset will have fields for characters and settings; we can add others later. Below are the possible fields.
- Story themes
- Story main human character, including proper noun [e.g., Barbara, Simon]; profession [abbess, beggar, wife]; nation/town; status [noble, royal, commoner]; gender; age [infant, child, teenager, young adult, middle age adult, the old]; type [protagonist, antagonist, and/or generally bad, generally good, and/or, evil, sinning believer, good nonbeliever, saintly]; religon [Muslim, Jew, Christian, pagan]; type of conflict [against self, against society/group, against another, against nature]; problem/challenge/conflict [lame, blind, castration, deaf, poor, disbelief, disease, exile, false accusation, famine, childlessness, away from home, pregnant]; sin [man-eater, adultery, jealousy, arson, blasphemy, drunk, frivolity, heresy]; virtue [celibacy, chastity, belief in Mary, doing

something for Mary, fasting]; threat [hell, ambush, discovery, death, drought, drowning, hanging]; activity [plowing, bathing, childbirth, dream, fall, travelling]; body part [ear, hands, penis]

- Story human character 2 [same as above]
- Story human character 3 [same as above]
- Story human character 4 [same as above]
- Story human character 5 [same as above]
- Story human character 6 [same as above]
- Story human character 7 [same as above]
- Story other characters (nonacting): wife, children, servants, friends
- Story human characters group(s): monks, children, Muslims, Cistercians, family, enemies, demons
- Story Divine Character 1 : Mary
- Story Divine Character 2: angels, demons, Christ, Holy spirit
- Story plot (maybe): travelling away from home, committing a sexual sin, healing
- Story emotions: hate, envy, terror
- Story animal(s): dog, dragon, birds, frog
- Story food(s): bread, grain, honey, beer
- Story four elements: water, earth, air, fire
- Story Mary mechanism: icon, vision, apparition, milk, hand, baptism, her belt, fragrance
- Story national setting/location: Ethiopia, Egypt, Israel, Syria, Cyprus, France/Europe/Farang
- Story province setting/location: Gojjam, Tigray
- Story town/village setting/location:
- Story landscape setting/location: mountain, sea, lake, farm, field, bridge, cave, garden, heaven
- Story building setting/location: monastery, church, home, castle, boat, furnace, gallows
- Story religious rite: baptism, prayer, burial, confession, Easter, eucharist
- Story texts: Gospel of John, Hail Mary
- Story religious objects: icon, Bible
- Story domestic objects: gourd, table, candle
- Story fighting objects: bow, arrows, sword
- Story other objects: alms, bell
- Story sources/intertextuality: seems to be in relation with foreign story
- Story origin: France, Germany, England or Europe

Step 3. Create PEMM Manuscripts Dataset. The Macomber handlist will be used as the basis for a Google Sheet titled PEMM *Täʾammərä Maryam* Manuscripts Dataset. The sheet will have more data than Macomber. It will be important to include information on manuscript dating and region, where available. Below is its format (with invented data for one manuscript to give a sense for the appearance of the data):

- Manuscript title: *Täʾammərä Maryam*
- PEMM Ms No.: PEMM 105

- Others' Ms No.: BN Ms. No. 23
- Manuscript original repository: Dabra Libanos, Ethiopia
- Manuscript provenance (lat., long.): 9.712177, 38.848075
- Manuscript current repository now: BN (Bibliotheque Nationale)
- Manuscript total no. of folios: 405
- Manuscript total no. of pages: 202
- Manuscript total no. of images: 202
- Manuscript total no. of stories: 103
- Manuscript century: 15.25
- Manuscript date range (if available): 1517-1543
- Manuscript illustrations no.: 28
- Manuscript illustrations size: 25 full page; 3 quarter page

Step 4. Create controlled vocabularies for PEMM Canonical Stories Dataset. We need to develop a controlled vocabulary of our own because (1) Macomber's is outdated (e.g., it uses "Moslem" instead of "Muslim"); (2) Hamburg BM's is designed for the Ethiopian environment, but not the Marian miracles specifically; (3) Oxford CSM's was not that well controlleds, so needs to be cleaned up; and (4) the Index's doesn't account for the Ethiopian environment specifically. Wendy and Evgeniia plan to combine all four controlled vocabulary sets and then comb through them for redundancies, to create cross references, and to identify hierarchies (e.g., we have "oxen" but also "animal" and the first is a type of the second). For instance, since catalogers may not think of the exact same word, we should have cross references (e.g., "angels" and "divine messengers [angels]"). With tightly controlled vocabulary lists, we can do better analysis of story themes.

Step 5. Mark beginning of incipts of each Story Instance in each manuscript. The research assistants will not know where the incipits are for each Story Instance. Those with an excellent level of Ge`ez will need to mark up manuscripts so that research assistants using the incipit tool and matching incipits don't have to read for the beginning of the incipit on the page.

Step 6. Catalog manuscripts. The biggest task will be increasing the number of Story Instances in the Google Sheet. With a bigger data set, we will be able to better answer the research questions. The research assistants will use the Incipit tool to catalog manuscripts first, giving each Story Instance a Canonical Story identifier number and then marking their degree of certainty that the incipits match. If they are not confident, someone who reads Ge`ez will go after them, checking.

Alternate Tasks

Currently out of scope is the following, but it might come into scope if we run into problems with students doing cataloging.

- **Write precises of Canonical Stories.** A huge and difficult task will be writing short summaries of the 700+ indigenous Canonical Stories. Only those with an excellent

understanding of Amharic, French, or Ge`ez will be able to do this work. Maybe 100 can be done from stories available in English translation. Or, perhaps, if the keywords are good enough, no precis is needed?

- **Track word length of Instantiation Stories.** This is a way of getting at the possibility of different recensions.
- **Tag Canonical Stories with keywords.** Another difficult task will be using the controlled vocabulary list to better tag Canonical Stories. Macombere did tag 642 of them with keywords, but many remain and his list can be improved. Only those with a good level of Ge`ez and English will be able to do this work.
- **Identify new Canonical Stories.** We need to give new identifier numbers, titles, themes, and incipits for Canonical Stories not in Macomber. We will use Hamburg BM identifiers where possible, but may need to do this a bit ourselves.
- **Translate into Amharic.** Translate titles, keywords, and website into Amharic.
- **Design and write static website.** This can be done in the last year.
- **Compare Cannibal of Qemer transcripts**. Once Steve, Jeremy, Jonah, and Ashlee complete typing up all 90 versions of the Cannibal of Qemer tale, we will do computational analysis.

Regarding the précises specifically, there would be certain sources:
- Amharic translations (currently out of scope):
  - Täsfa Giyorgis, ed. 1931. *Täʾammərä Maryam bä-Gəʿəz ənna bä-Amarəñña* [The Miracles of Mary in Gəʿəz and Amharic: 111 Miracles]. Addis Ababa, Ethiopia.
  - Täsfa Gäbrä Śəllase, ed. 1996. *Täʾammərä Maryam bä-Gəʿəz ənna bä-Amarəñña* [The Miracles of Mary in Gəʿəz and Amharic: Part Two: 402 Miracles]. Addis Ababa, Ethiopia: Täsfa Gäbrä Śəllase Printing Press.
  - Täsfa Gäbrä Śəllase, ed. 1994. *Sədsa Arattu Täʾammərä Maryam* [Sixty-four Miracles of Mary]. Addis Ababa, Ethiopia: Täsfa Gäbrä Śəllase Printing Press.
  - Täsfa Gäbrä Śəllase, ed. 1968. *Täʾammərä Maryam bä-Gəʿəz ənna bä-Amarəñña* [The Miracles of Mary in Gəʿəz and Amharic: Part One: 270 Miracles]. Addis Ababa, Ethiopia: Täsfa Gäbrä Śəllase Printing Press.
- English translations (currently out of scope):
  - Budge, E. A. Wallis. 1900. *The Miracles of the Blessed Virgin Mary, and the Life of Hannâ (Saint Anne), and the Magical Prayers of 'Aheta Mîkâêl: The Ethiopic Texts Edited with English Translations Etc*. 2 vols, Lady Meux Manuscripts Nos. 2-5. London: W. Griggs.
  - Budge, E. A. Wallis, ed. 1933. *One Hundred and Ten Miracles of Our Lady Mary*. London: Oxford University Press, H. Milford.
  - Zärʾa Yaʿqob. 1992. *The Mariology of Emperor Zärʾa Yaʿəqob of Ethiopia: Texts and Translations*. Translated by Getatchew Haile. Edited by Getatchew Haile. Rome, Italy: Pontificium Institutum Studiorum Orientalium.
- French translations (currently out of scope):
  a. Colin, Gérard. 2004. Le livre éthiopien des miracles de Marie (Taamra Mâryâm). Paris: Les Editions du Cerf.
- Italian translations (currently out of scope):

a. Cerulli, Enrico. 1943. Il libro etiopico dei Miracoli di Maria e le sue fonti nelle letterature del medio evo latino. Rome: G. Bardi.