5. Sustainability Plan

This sustainability plan consists of two sections: in §5.1 we describe the development principles and other efforts by the BPS team that support the sustainability of the project. In §5.2 we discuss those aspects of the institutional context of BPS at the University of California at Berkeley that are relevant for its continuous growth and development.

5.1 Sustainability: Project Specifics

Sustainability of BPS has been a priority from the beginning, and is reflected in our community engagement model, in our software development principles, in the technical architecture of our production system, and in the commitment of the technical and the humanities specialists of the BPS project to nurture and maintain communication between disciplines. Moreover, the BPS team has made conscious efforts to connect to the various campus units and campus initiatives that are actively engaged in Digital Humanities and Data Science, and to contribute to such efforts with our experiences and results (see §5.2).

The community of cuneiform scholars working on the corpora that were the initial focus of BPS has a history of collaboration and cooperation. The ORACC repository for curating the corpora is one artifact of this collaboration, and the CDLI repository at UCLA is another. Notably, scholars of one of the world’s earliest writing systems have quickly embraced the shared standards and best practices of the digital world. Members of this community who participated in workshops and user studies in the initial stages of the project remain engaged and enthusiastic about ongoing development. In the meantime we reached out to a broader range of academic disciplines that are interested in prosopographical research, including Islamic studies, history of zoology, and Russian. Several domains have been identified as strong candidates with which to explore the details of our model, our workflows, and the related tools necessary for scholars to make use of BPS in their research on their corpora. The grant will support this work by funding workshops, further user studies, and related ongoing engagement with them to expand our community. In addition, we will continue to publish and present the tools and the resulting research outcomes on our campuses and in appropriate fora (conferences, journals, etc.) to ensure that potential collaborators are aware of BPS, how it may be useful to them, and whom to contact should they wish to get involved. The engagement with a vibrant, diverse community of scholars is an essential aspect of project sustainability. This activity does require some overhead for facilitating discussion, but Project Manager Pearce and her colleagues see this as part of their role as academicians, and are committed to sustaining this effort.

We believe that best practices in software design can significantly improve the sustainability of a tool like BPS, by ensuring broad utility of the tool-kit and its individual modules and by reducing the cost to maintain and support the tools. To that end, we identified key principles that foster broad re-use of the tools, beyond the initial domains in which we worked. These principles include:

- A strict separation of domain-specific rules, conventions, or practices from the core algorithms, models, and software implementations. Where corpus specific knowledge is important to the functioning of the tools, the functions were abstracted in the model, and then corpus-specific “plug-ins” were developed. As such, the disambiguation rules are configured for each corpus and parameterized by each researcher.

- The choice of common, open communication platforms for project planning and coordination. All project documents, issue tracking systems, etc. are open to the community so that our project management process is transparent. This helps a broader community stay engaged and feel confident that they know where the project is headed.

- The choice of a web-application model (rather than a traditional desktop application model) makes it much easier for new users to take advantage of the tools with no installation required,
and to access them from a variety of locations and devices. This is consistent with the evolution of
digital humanities computing to be a web-based endeavor.

- The choice of commonly used programming languages, platforms, databases, and related tools,
  so that a broad range of software developers can engage with the project. This includes Java-
  based web services using a popular REST platform (JAX-RS) hosted as a standard web
  application in Apache Tomcat, and a standard tool-kit for the user experience (HTML, CSS,
  PHP, and JavaScript+jQuery).

- A RESTful, services oriented architecture ensures that the tools can easily be integrated into
  the larger environment of tools, content management systems, and related IT infrastructure in which
  modern digital humanities takes place. We follow best practices that separate key aspects of the
  infrastructure (e.g., corpus management, workspace management, the disambiguation engine,
  etc.) from one another, so that we can evolve or even replace a given function with minimal
  disruption to the system as a whole. This makes it much easier and more cost-effective to
  maintain and evolve the software over time.

- A open source software license to ensure that others can re-use our models and software.

We believe that all of these principles help to ensure the long-term sustainability of the software and
provide the most effective basis for ongoing development.

Another set of architectural choices was made to ensure the sustainability of production systems that
support BPS functionality. These follow current best practices in IT, and while that may seem unusual for
a digital humanities project, we believe that it will become an essential piece of the sustainability
equation. Rather than deploying the BPS services on a traditional server, such as is typically maintained
in departmental or even central IT locations, we leverage the use of cloud-based platform services. These
are professionally supported, and at a scale that makes them very efficient and extremely cost-effective.

For a small set of corpora in the early phase of the project, a VM costs as little as a few hundred dollars a
year. This includes the costs to keep the system up to date with stable platform releases and security
“patches”, and includes support for hardware monitoring, access and load logging, and regular backups,
all managed by skilled professionals (vs. part-time graduate students, as is all too often the case). As the
user base grows, or as more complex analysis is added that requires increased compute resources, the
cloud services can be easily scaled up with minimal downtime for the service and without risk to the
software and data. This is much simpler than moving software to a new local server, which can incur
considerable IT resources and risk of downtime for the service. The incremental cost of accommodating a
new set of corpora is quite low, and continues to decline as compute resources come down in price. We
have no overhead invested in servers or local system administrative staff. The cloud-based deployment,
and the concomitant cost-effectiveness, stability, and flexible scalability of our production service, are
significant contributors to the sustainability of BPS.

All of these activities provide a basis for sustaining the BPS project. Nevertheless, there are some
ongoing costs that must be paid to keep the system running. While the expectation is that any significant
development work is to be funded through project-based budgets (e.g., through grants such as this),
production costs and some minimal staff activity are required as well. Up to now, UC Berkeley has
covered the modest costs associated with the cloud services, and the ongoing staff time falls within the
campus supported activities of the core personnel. The institutional commitment is easy to sustain when
the costs are so low. To handle the case of considerable growth in the corpora and user-base using BPS,
we are exploring a variation on what has been called a “grant-tax” in which researchers who import a
corpus and wish access to the services for analysis would pay a small portion of their research funding to
help support the project. If they anticipate considerable technical work will be required (e.g., to develop
custom disambiguation rules, or to process their corpora for effective input into BPS), then they would
have to include budget in their research planning for the associated resources (as a development project).
Part of the current project will be to build experience in the cost of such corpus-specific development, as a
guide to researchers preparing budgets for their projects. Finally, it is worth noting that BPS need not be a
single, centralized service. It could be easily deployed by others as well, if another community or domain of users wants to sustain the services for their research activity.

We are confident that the community, the technology choices, and the deployment architecture place BPS on a strong footing for sustainability. We continue to monitor and explore the issues that will arise if we see a significant growth in our user community, and believe we have realistic plans to handle these scenarios.

5.2 Sustainability: Institutional Context at the University of California, Berkeley

BPS is developed in the context of UC Berkeley’s sustained investment in cyber-infrastructure for the arts and humanities. The history of BPS goes back to the participation of Veldhuis and Schmitz in campus explorations of the needs and feasibility of producing digital tools that respond to actual research agendas in the humanities. The campus commitment to developing reusable, extensible tools for humanities research has produced projects that truly serve the research needs of humanities investigators. Continuing activity on DIRT and related DH resources reflect ongoing campus support of and commitment to that effort.

With external funding and considerable campus support, in the form of infrastructure, technical expertise, and funding from a variety of sources, BPS has established a presence on campus; the team is regularly invited to participate in campus-wide presentations of Digital Humanities. BPS joins successful DH projects, such as ECAI (Lewis Lancaster) and ORACC (Veldhuis) that have both a long-standing campus presence and that situate Berkeley prominently in the broader DH community. The steady growth of DH at Berkeley is attributable to the efforts of campus officials at various levels to support and promote such research. This year Professor Anthony Cascardi, Dean of Arts and Sciences, has established the Digital Humanities Council (which includes Veldhuis) to support him in finding ways to create an integrated, extensive DH program and to coordinate faculty DH efforts.

Although conceived, designed, and implemented in the realm of DH, BPS participates in and benefits from the intensive expansion of UC Berkeley into Data Science. In 2013, Prof. Cathryn Carson, Associate Dean of Social Sciences, created D-Lab, a collaborative environment for data-heavy research questions. Most recently, a grant from the Sloan and Moore foundations contributed to the establishment of the Berkeley Institute for Data Science (BIDS). While the missions of the Sloan and Moore foundations focus on the social and hard sciences, BIDS has made a public commitment to include DH as a part of the range of supported domains. The participation of Arts and Sciences Dean Cascardi on the BIDS advisory board assures a strong DH presence in this initiative. The campus environment welcomes and supports the cross-disciplinary conversation, evidenced by the invitation extended to BPS to participate in the Data Science Faire that accompanied the public launch of BIDS.

UC Berkeley, with its long standing excellence in the I School, EECS, and IST, and with an impressive history in humanities scholarship, has embraced the challenge of bringing these two worlds together in the development of Digital Humanities. Within this context, BPS can flourish as a project that will both profit from the expanding interactions between humanists and technicians, and as a project that furthers this ongoing development.