Project Proposal: Ingestion of Professional Fonds

INFO 284: Digital Curation

Professor Darra Hofman

San Jose State University, School of Information

Fall Semester 2024

Chris Reynolds

November 9, 2024

Ingestion, Curation, and Preservation of the Terry Scientist Professional Fonds

As a continuation of a tentative agreement to acquire valuable digital and non-digital assets from Professor Terry Scientist, proud alum of Teeny Tiny University and renowned Harvard geneticist, I submit the following project outline for the successful ingestion of these fonds. As part of this proposal, estimates for funds needed, staffing requirements, possible risks associated with the project and their mitigation, as well as a timeline for completion are included.

Purpose

The donation of Professor Terry Scientist's professional fonds offers TTU a rare opportunity to acquire a world-class archival collection with significant scholarly value. This collection spans decades of groundbreaking genetic research and has the potential to elevate TTU's reputation as a hub for academic excellence, attract external funding, and foster interdisciplinary research collaborations. In addition, these materials fall in line with TTU's Archive Policy (TTU, 2017) in that they are manuscripts and papers which reflect the life of the University community (ie alumna work).

Potential outcomes associated with the acquisition of Professor Scientist's fonds include:

Enhancement of TTU's Institutional Repository:

By integrating the collection into the TTU DSpace repository, we will reinforce our university's commitment to data sharing and re-use.

Attract Scholars and Funding:

The addition of Dr. Scientist's expansive academic collection establishes TTU as a top research destination for genetics, in turn attracting future research, donations, grants, and partnerships.

Preserve a Valuable Legacy:

Preserving the work of a Nobel Prize-winning scientist bolsters TTU's reputation as a university that prioritizes scientific innovation and research excellence.

Resources

Collection & Contents

The collection includes ten linear meters of analogue records, five computers and six external hard drives (including an old Linux box and an old Macintosh, with unknown contents), two databases hosted on Amazon Web Services (AWS) and one database on a server under Professor Scientist's control. File formats include:

- Word Processing: .DOC, .DOCX, .WPD
- Database: .NS2; .SQL, .SSD
- Spreadsheet: .WK1, .WKS, .XLS
- Statistical: .SAS7BDATA, .RDATA
- Discipline specific: .FASTQ, .BAM, .AB1, .GEDCOM
- Other: .CSV, JSON, XML, PDF

Collection contents include material from 1986-2020 and range in subject type from contracts and grant applications to correspondence and data sets and comprehensive lab notebooks. Digital records since 2010 are well documented in accordance with data management plans associated with their various grants. Materials previous to 2010 have never been fully inventoried.

Personnel and Budget Estimate

Ingestion and Curation Phase (2 Years):

2 Full-Time Equivalents (FTEs): Digital Archivist and Data Curation	\$120K/yr
Contracting service for hardware and format migration	\$200K
Digital infrastructure upgrades for 1.5 PB of storage	\$300K

Ongoing Preservation Phase:

Additional 0.5 FTE for maintenance and user support \$50K/yr

Plan for Processing Materials

Following the OAIS model (Higgins, 2008) the fonds will follow steps to ensure proper

integration into physical and digital archival collection.

Receive	- Gain the physical and digital items from Professor Scientist
	- Conduct a more indepth donation interview and clarify appropriate
	rights statement or complete paperwork for transference of rights to TTU
	- Contact Harvard archivist who is also working on Professor Scientist
	collection there
Appraise & Select	- Review items to understand the details of the collection
	- Initial organization to provide structure for the collection
	- Gather data including a more detail inventory of the items, notes that
	describe the items and make connection to other resources
	- Weeding duplicate and unnecessary items and rehoming items that don't
	fit the collections purpose
	- *Create specialized plans for digital formats that may require specific
	technical support (such as databases, spreadsheets, statistical models,
	email, and other discipline specific formats.)
Ingestition	- Documentation & metadata creation
	- Rehousing physical items to archival grade folders and boxes
Preservation	- *Reformatting selected digital items. These items should be either of
	particular interest to our campus or be easily reusable and they aren't
	easily accessible with our current technology; this may include migration
	or emulation for materials that require legacy hardware or software.

~ \$1MM

	- *Digitization plan for physical items that should be accessible through
	our website or that are endangered of degradation, or are requested.
	- Allocation of space within the archive for physical materials.
Storage & Backup	- *Contact IT to talk through long-term storage and access needs
	- *Ensure approximate 1.5 PB of space (this includes backups) across
	multiple servers
Ensuring Access	- Creation of finding aids and linking to Harvard's collection as required
	- Outreach via physical and digital displays, exhibitions, and events
	*Special technology needs & considerations

External Funding

Grants from resources like the National Records Administration (NARA), National Science Foundation (NSF), or National Institute of Health (NIH) could contribute to our financing strategy. In addition, private donations may be available due to the prestige of the collection. Research and connection with potential donors and funding organizations is a priority.

Partnerships

A collaboration with Harvard's library and archive team will be essential to the success of this project as part of Professor Scientist's collection will be housed there. We intend to cooperate with their archivists in sharing collection information, expertise, and finding aids to create links between the collections.

Sustainability

Prioritizing sustainability in a project of this size and expense is essential. TTU's DSpace platform will be the starting point in ensuring standardized, scalable, and secure access to the archives. As "sustainability of preservation systems relates directly to their usefulness," (Gerrard, Mooney & Thompson, 2018), it will be vital to ensure proper maintenance to the repository and the collection on a regular basis. This is reflected in the suggestion for additional personnel requirements, a dedicated 0.5 FTE will guarantee the upkeep of project materials. Additionally, reaching out to the IT team and establishing clear communication should be done soon. Protocols regarding the needs of the collection, patron and personal access to files, file structure, collection security, and digital storage space requirement should be discussed in terms that are understood by both parties (Prater, 2017). Finally, we will want to leverage the collection's visibility to attract donors and sponsorships to ensure ongoing financial support to sustain the archives going forward. As outlined in the processing plan this may include physical and digital displays, exhibitions, and events.

Risks and Challenges

No project is without risk, but we truly believe the benefits outweigh any challenges we may encounter during this project. Foreseen primary risks include financial risks and ongoing costs, technical complexity of the materials, and a significant amount of unindexed content.

For financial risks contracting, full-time staffing, and additional financing means a high upfront cost for TTU. However, this risk may be mitigated by securing adequate grants and donations and adopting a carefully planned, phased-in approach. There is also potential long-term financial burden of storage and staffing for TTU, but this is addressed in the sustainability plan detailed above which includes using the reputation of our new genetics archive to attract ongoing donors and sponsorships. Fundraising campaigns may also play a part in mitigating long-term financing requirements.

Many of Professor Scientists resources include obsolete file formats and hardware, which entails a few complex technical solutions added to the integration process. Again, however, our plan includes a budget for specialized contractors to help streamline this part of the project. Utilizing best practices in digital preservation via ongoing monitoring, management and use of resources within the collection, as suggested by Gerrard, Mooney and Thompson (2017), will help avoid this type of obstacle in the future.

Unindexed Content

Certain resources offered by Professor Scientist include contents uncertain. Sifting through potentially redundant or irrelevant materials, will add time and labor to the task and necessitates a level of scientific expertise to appraise the materials. Initial appraisal may be done by a TTU scientific graduate student intern, who can sort and inventory unindexed materials.

Advantages to Teeny Tiny University

Ultimately, we feel the work and financing required to integrate Professor Scientist's professional fonds will be well worth the effort because of the advantages it will give our university. First, the acquisition of a Nobel Prize-winning researcher's collection will elevate TTU's institutional repository by providing a comprehensive and well organized collection that's available for genetics research. Moreover, the breadth of materials within the fonds promises to stimulate interdisciplinary research across fields such as genetics, data science, history, and digital humanities, and can foster collaboration among diverse academic departments. The collection will support our mission of open-access scholarship, by reflecting TTU's commitment to data sharing and benefiting future generations, as emphasized by Professor Scientist's desire to make her work widely available.

By showcasing the achievements of distinguished alumni, this collection underscores TTU's dedication to preserving their legacies, potentially encouraging future contributions from alumni across various disciplines. Additionally, the visibility of this high-profile acquisition is poised to attract new funding opportunities, which could support the project's sustainability and further the university's broader academic goals. This increase in academic prestige may be applied to the project itself as part of its sustainability or elsewhere as needed for our school.

Conclusion

The Terry Scientist Professional Fonds is an opportunity for TTU to secure a significant scientific research collection. While the initial investment is substantial this plan demonstrates that the long-term benefits in institutional reputation, scholarly impact, and fundraising potential far outweigh the costs. With a strategic approach to funding, partnerships, and implementation, TTU can successfully preserve and promote this invaluable resource for generations to come.

References

- Gerrard D. M., Mooney J. E., & Thompson, D. (2017). Digital preservation at Big Data scales:
 Proposing a step-change in preservation system architectures. Library Hi Tech, 36(3):
 524-538. <u>https://doi.org/10.1108/LHT-06-2017-0122</u>
- Higgins, S. (2008). The DCC curation lifecycle model. The International Journal of Digital Curation, 3(1): 134 - 140. <u>https://doi.org/10.2218/ijdc.v3i1.48</u>

Oliver, G., & Harvey, R. (2016). Digital Curation (2nd ed.). ALA Neal-Schuman.

- Prater, S. (2018). How to talk to IT about digital preservation. Journal of Archival Organization, 14(1): 90-101. <u>https://doi.org/10.1080/15332748.2018.1528827</u>
- Teeny Tiny University. (2017, March 20). Archives, Recorded Information Management, and Freedom of Information and Protection of Privacy.