The production log: A digital imaging capacity replication project

Received (in revised form): 16th March, 2018

Marcia McIntosh
is the Digital Production Librarian at the University of North Texas Libraries Digital Projects Lab, where she helps supervise the digitisation of materials relating to or about the University of North Texas and the history of Texas. She received her MSIS in information studies from the University of Texas at Austin.

University of North Texas Willis Library, 1155 Union Circle #305190, Denton, TX 76203-5017, USA
E-mail: marcia.mcintosh@unt.edu
Tel: +(940) 369-7809

Shannon Willis
is Digital Projects Lab Manager for the University of North Texas Digital Project Unit. She received her MA in film studies from Emory University and her MLS from the University of North Texas.

University of North Texas Willis Library, 1155 Union Circle #305190, Denton, TX 76203-5017, USA
E-mail: shannon.willis@unt.edu
Tel: +(940) 369-7809

Abstract The University of North Texas Digital Projects Unit Imaging Lab (DPU) took on the challenge of providing project completion estimates to its partners. The DPU replicated the Points Process system developed by the Northwestern University Libraries' Repository and Digital Curation Department. The project has enabled the lab to begin estimating the time necessary to complete imaging projects that come through the lab, thus giving staff members the ability to start measuring the capacity of the lab. This paper will report on the process and steps taken by the lab and best practices for repeating the process elsewhere. The DPU intends to extend the measuring of capacity by adding inventory and metadata numbers as well as developing a more automated method for capturing production statistics. Being able to quantify lab capacity will enable lab staff to better arrange for the scheduling of projects with both internal and external faculty and staff.

KEYWORDS: project management, digitisation, workflows, replication, process

INSPIRATION

The question of how long something will take is often a challenge to answer with digital collections development. To try and provide verified answers to such questions for its various partners and stakeholders, the University of North Texas' (UNT) Digital Projects Unit Imaging Lab (DPU) replicated a new statistics gathering and reporting process developed by Northwestern University Libraries on its imaging production.

The DPU is one of several units in the Digital Libraries Division within the University of North Texas Libraries that provides digital services including imaging, archival storage of electronic files, metadata development and other activities relating

to digital preservation. Together with the Digital Curation, Digital Newspaper, Software Development and User Interfaces Units, the division creates and contributes content for two digital libraries: The Portal to Texas History (http://texashistory.unt.edu) and the UNT Digital Library (http://digital.library.unt.edu). The UNT Digital Library contains over 400,000 items and is the main institutional repository for university scholarship and university-owned materials and collections. The Portal to Texas History has over a million items created from content from both within the university and from external partners across the state of Texas. The DPU's main function in building these digital libraries is the digitisation of materials. It has a variety of equipment to capture photographic prints, film materials, bound and unbound documents, books, oversized items and three-dimensional objects (a fuller look at equipment currently used in the DPU is available on the UNT Digital Projects Unit website[1]). All projects that come through the lab are on loan from over 300 partner organisations, departments on UNT's campus and inside UNT Libraries.

While there are sometimes built-in deadlines for projects to begin and end with the fiscal year (eg with the DPU administered mini-grant programme, Rescuing Texas History, or state and national grant projects), many projects enter the lab informally or on an individual request basis. Naturally, the question of how to prioritise and schedule the projects that flow through the lab year-round, and to get them all done either by firm deadlines or in fulfilment of partner expectations, is a challenge for DPU staff to manage. Without a formal process for committing the DPU to projects and without an understanding within the Division of the DPU's capacity for digitisation, the DPU frequently suffers from either too few or (more frequently) too many projects than it can handle in a given time period. The ensuing backlog queue makes it difficult to estimate completion dates for both old and new projects. This queue makes it difficult to communicate to partners or individuals interested in starting new digitisation projects. The Points Process system, developed by Northwestern University Libraries' Repository and Digital Curation Department (NUL RDC), offered a solution for staff at UNT to measure and quantify the DPU's capacity in a way that could create and clearly communicate reasonable expectations to other library staff as well as non-UNT partners. The system has enabled the staff to increase the DPU's efficiency and better manage its resources.

Staff members of UNT's DPU first encountered the RDC Points Process at the 2016 Digital Library Federation Forum in Milwaukee WI.[2] The system's goal and result was to accurately gauge a lab's capacity to digitise materials and, therefore, commit to projects and partners in line with the resources and capacity of that lab. Although UNT's and Northwestern's digitisation labs differ in many ways, the idea of having lab-specific units in which to communicate the DPU's capacity was worth attempting. In the spring of 2017, staff at UNT began replicating the points system.

UNT DPU VERSUS NUL RDC

Upon beginning the replication project, lab staff took note of the differences between UNT's DPU and NUL RDC. Two staff members from UNT were kindly hosted at Northwestern as part of a separate research project and were able to tour Northwestern's digitisation facilities.[3] The authors were provided the opportunity to ask questions directly of NUL RDC staff, learn about RDC's workflows and identify some key differences in the two arrangements.

One of the main differences is that the RDC, at the time of the visit, employed nine full-time staff members, one temporary worker and two part-time students to

complete the digitisation work. UNT's DPU, on the other hand, functions with four full-time staff members and, on average, 20 part-time digitisation and six part-time metadata student assistants. The authors knew that the part-time and often variable attendance of the DPU student workers would make estimating the time of digitisation, and thus capacity, more complex than that of full-time staff members.

Additionally, RDC's projects come from curators within NUL with whom they are able to communicate directly when arranging projects. RDC staff are able to see samples, if not the entire project, before agreeing to do it.[4] The DPU, in contrast, has multiple sources for its materials both inside and outside the university. From within the university libraries, the DPU digitises materials for the Music Library (one of the largest in the USA), Special Collections, Government Documents (UNT is a US federal depository), the Media Library and Scholarly Publishing Office. Typically, a staff member from the DPU is able to coordinate the digitisation of projects with these divisions. However, projects, at times, also come from other colleges at the university. The Repository Librarian, in a separate unit from the DPU, usually coordinates these projects. The bulk of materials that the DPU handles come from outside the institution, and are generally agreed to and coordinated by other library divisions. The DPU has the assistance of an external relations representative who handles communication with its 300-plus external partners and arranges for the import of new projects and export of finished projects from the lab. DPU staff are also responsible for the digitisation of projects that are coordinated by the Division Dean, who works with certain long-standing partners. In essence, the UNT DPU acts much like a vendor for digitisation as it is a typical in-house university digitisation unit. Furthermore, because DPU accepts external projects from outside institutions, staff members must rely on partner-provided estimates and descriptions, and can only visually inspect materials once they arrive. Committing to projects sight unseen, along with having multiple staff, departments and divisions committing the lab to such a large number of projects, has increased the size of the demand for the Points Process and its benefits should DPU staff apply it effectively.

NUL RDC POINTS PROCESS

The NUL RDC steps in developing the Points Process consisted of: determining the capacity of each staff member, calculating the capacity of the entire lab, interviewing prospective project curators, creating a shortlist of proposed projects, reconciling capacity with the shortlist of proposed projects, selecting the final projects and scheduling the projects. A full explanation of the NUL RDC development is available elsewhere.[5]

The RDC decided to use a merging of two systems: a points system based on the Library of Congress and a project selection process from Emory University.[6] The points system sets one point equivalent to one hour of work. RDC measured its capacity by calculating the points for each of its lab's staff members and then reduced that amount by one-fifth to find 80 per cent efficiency.

With the capacity measured, RDC could implement the Points Process by conducting individual interviews with the library curators wishing to have projects done. The projects were compared with the lab's estimated capacity. Priorities were identified, and the ranked list was presented to library special collections and archives administrators. With the administrator's input, a shortlist was developed and further scrutiny given to those projects to assure the accuracy of the original estimate. The final selection was then scheduled, taking into account staff resources and curator deadlines.

UNT DPU POINTS PROCESS REPLICATION

Before trying out the RDC's Point Process, the DPU had attempted some production and statistics collection efforts. The lab currently uses several overlapping, low-cost documentation systems for its projects, including an internal wiki for recording project information, a large whiteboard for overall lab progress and project tracking, and, most recently, the project management software Trello to track project status with more detail and at finer levels of movement. Before the autumn of 2014, using only the internal wiki, staff members took statistics on what was created in the lab on a weekly basis and recorded them in the DPU's weekly report — a snapshot still created today to capture the weekly status of projects currently in the lab. But while this kept a record of the number of items and images created each week in the DPU, no real analysis of the numbers was ever done. In a separate effort, in 2015, lab staff experimented with timing the scanning of individual objects on each of the lab's four types of equipment. The materials scanned were not part of any digitisation project, but rather were fabricated for the purposes of testing. These numbers, while useful, did not take into account the other types of activities conducted by students or their overall pace of work. The idea of measuring their production on a normal, daily basis, with the real materials they handled in partner projects, was much more appealing. In addition to measuring how long it could take to complete a project, the Point Process also promised an indication of how many projects the lab could reasonably plan to finish in the space of one year.

In imitation of NUL RDC, UNT's DPU began by measuring the capacity of its personnel doing the actual imaging work. The method, however, varied from the RDC's in staffing type and work hours. The authors first decided how to track production and time. They created and gave a production log to each student imaging assistant as an Excel spreadsheet. This spreadsheet allowed the students to enter their own production data. In the history of the lab, other statistic-collecting measures, added to the lab's weekly report, were collected by a single lab staff member and required a whole day's effort. This older method was not replicated partly because of how long it took, but also because it provided only a weekly snapshot of imaging progress for the whole lab rather than daily, hourly, or person by person. It was thus decided that the student imaging assistants would collect their own production data, thereby acquiring the whole of the desired information at a reasonable time cost.

Each student production log recorded the date on which an activity took place, the total number of hours for that activity, the name of the project on which the student was working, the type of objects in that project, the particular activity conducted on those objects and the type of equipment used for the activity. The log also captured the number of individual objects the students created or worked on in that time-frame. The DPU holds the standard that one object, such as a book or photo, is placed in individual digital folders and given a unique identifier, making it a significant unit of measure. The approximate number of objects, however, is not a fair judgment of productivity, as one photo would consists of two Tagged Image File Format (TIFF) images (UNT digitises both front and back of all photographs), while one book could contain dozens or several hundred TIFF images. To provide context for the number of objects, the authors also collected the number of TIFFs created or worked on. Lastly, students were given the opportunity to record any relevant details that may have impacted the activity, such as removing staples, handling delicate materials or fragile pages, or leaving a shift early. The staff members advised students to add to their production logs at the end of each shift, but

Table 1: Production log input options

Type of objects	Activity	Equipment
Books	Disbinding	A0 Quartz
Documents — flatbed	Scanning	A1 Quartz
Documents — autofeed	Scanning/Cropping	Bookedge
Documents and photographs — flatbed	Rescanning/Fixing	Copibook
Magazines	Metadata	EPSON large
Mixed materials	ScanTailoring	EPSON small
Negatives	Pre-QC	Fujitsu
Oversized — flat	Re-checking	Metadata
Oversized — bound	Final-QC	Phase 1
Photographic prints	Initial inventory	No equipment
Scrapbooks	Final inventory	
Slides or positives		
Yearbooks		

only required them to have their production logs updated by the end of each Friday.

The first month of data collection revealed some differences from student to student in data input and formatting. Some students lumped all the activities and the different equipment used into a single row in their logs. Some called certain activities or equipment one name, while others working on similar projects called it another. There were words and symbols used in the numerical columns that made the data clean-up extensive and analysis difficult. To remedy the issue, input suggestions and a standardised vocabulary for three of the columns were created. Staff added a production log instruction page to the DPU's inter-departmental wiki (the lab's primary method of documentation) to include standardised lists for type of object(s), activity and equipment (Table 1). The staff additionally added a time conversion chart to the instructions for how to represent minutes in decimal form.

ANALYSIS

The first round of data was gathered from February to mid-May 2017. Having cleaned and sorted the data for each student, the authors calculated the average time per TIFF for each of the main types of scanning: Epson documents, Epson photographic prints, Epson negatives, Fujitsu pages, Bookedge pages, Copibook documents and i2S Quartz materials. Each main type of scanning is determined largely by the piece of equipment used, as each machine varies greatly in its scan time. For collections that included a mixture of these item types (and therefore a mixture of equipment used), the authors conducted a separate calculation, termed 'mixed'. The authors gathered all the mean times and TIFF numbers from each student and calculated the mean number of TIFFs per hour for each type of digitisation. It is the standard of the DPU to digitise the backs and fronts of all objects, except film material. For example, one photograph would include two TIFF images. To simplify future calculations, the authors divided the number of average TIFFs per hour for Epson documents, Epson photographic prints, Fujitsu pages, Copibook pages, Quartz items and mixed items to determine the number of pages or items per hour. The reported number represents one object, or page, consisting of two images, rather than a single

Table 2: Objects able to be digitised by capacity point (1 hour)

Objects capacity estimate	1 Point
Epson documents	7
Epson photographs	6
Epson negatives	8
Fujitsu pages	36
Bookedge pages	50
Copibook documents	12
I2S Quartz materials	13
Mixed items	15

image. Following the NUL RDC Point Process model, one point of capacity equals one hour of work. Table 2 presents the lab data for spring 2017.

Due to differences in staffing, the authors' calculations of the DPU's time capacity varied slightly from that of NUL RDC's Point Process (See Table 3). The number of hours requested and scheduled by the 15 student workers employed at the lab at the time the initial data collection and analysis was 290 hours per week. That number was then reduced to 80 per cent — the amount of requested and scheduled hours the lab requires students to work in order to be in good standing — making the expected number of student hours per week 232. The authors multiplied the expected number of hours per week by the number of working weeks in the spring semester (18) to find the DPU's capacity for that semester: 4,176 points.

The authors then calculated the yearly amount of capacity points using the estimated working weeks (49) in a year multiplied by the number of hours per week (232): 49 weeks × 232 hours = 11,368 points per year.

This number, however, assumes that staff, student workers and hours requested remain constant over each semester; this, of course, is unlikely. For instance, the allowed number of hours students are permitted to work in the summer increases, in some cases, by a factor of two, while some students may choose to return home during summer months. Hiring more students, as the authors did following the end of the spring semester when production data analysis was done, would also affect this capacity number. Thus, this capacity number is, expectedly, tentative and will likely be in flux over the course of the fiscal year. It does however provide a better gauge of the possibilities in terms of the DPU's capacity.

DISCUSSION

The process and results of calculating the UNT DPU's capacity by adopting the NUL RDC's process have enlightened its processes and planning. With more confidence, DPU staff can now provide estimates on the digitisation time for projects to partners when asked about a prospective project. Lab staff can additionally project, when given page and item counts, how long it may take a student to complete a job or fix issues that emerge while digitising. This small beginning has helped to chart next steps in putting the data and resulting capacity points into action throughout the lab.

Although it has provided an accurate estimate for the DPU's imaging capacity,

Table 3: DPU capacity variables, calculations and total

Variable	Value
Student hours requested per week	290
Hours of worked expected per week	232
Weeks in the spring semester	18
Capacity for spring 2017 equation	18 weeks × 232 hours
DPU capacity hours for spring 2017	4,176 points

the DPU's application of the NUL RDC's Points Process can still be applied further within the lab. In its first round of capacity analysis, the DPU measured the capacity of its imaging operation, but not that of its inventory, metadata or quality control steps. An assumption was made at the beginning of data collection that the imaging process was the primary factor of the lab's overall capacity. It was assumed that if the students could digitise the material, then the other staff could conduct processing, perform quality control and create metadata for that material in an equal timeframe without creating a significant backlog queue at each step in the workflow. However, this is an assumption and cannot as yet be supported with data. Thus, lab staff have already begun gathering data on these other areas.

There have also been unforeseen benefits with creating and requiring the production logs from each student assistant. While the UNT DPU tracks projects weekly using a report in its internal wiki site and monitors the status of each project with the large whiteboard and Trello, it is sometimes critical to identify what a student has worked on during a particular shift. The production logs have, thus, been the stop-gap between the weekly reports and project tracking methods by providing daily detailed accounts of individual student activities. The log has also been used to track attendance when other time-accounting software has been faulty or when a student's presence comes into question. Scanning rate can also be a determination in cost setting.[7] The DPU has used it historically for that purpose. The rate of scanning numbers generated from this project has already been used to give quotes for prospective projects. Furthermore, the DPU has for years maintained a set cost for imaging and metadata. The data garnered from this capacity measuring project may help to enlighten and update these set costs to better reflect actual cost spent on digitisation work in the lab. These figures also would be able to vary more easily over time as data continue to be collected, given the increased or decreased speed at which the lab produces images.

LESSONS LEARNED

Having begun the NUL RDC's Point Process thus far, the DPU is able to recommend some areas of attention should others wish to apply the process in another lab. The first recommendation would be to consider lab differences. The UNT DPU knew that the staffing composition (student workers), scheduling of projects (done by external representatives), project origin (inside or outside the library) and operation size would all create differences in how the Points Process would be applied. For the UNT DPU, the number of students in the lab, the speed of the students on various equipment and the variety of materials digitised made recording the average time per material type crucial.

Secondly, data collectors should define the input value and guidelines early on to better capture manageable data. The DPU was able to do this, but needed to refine those guidelines after a pilot period. The sooner one can set defined values and guidelines, the more useable the collected data will be. Thirdly, have student or staff members track their own production. Due to the size of the DPU and the limited time of its full-time staff members, the most efficient way to capture the data was to have those creating the digital files collect the data themselves. This lightened the burden on full-time staff who at one point were asked to collect statistics on all items scanned during the week. It also provided student workers an opportunity to view and benchmark their own production.

Lastly, there is a great deal more information beyond a lab's capacity that can be distilled from the production log data, such as error rate for a project, the number of items digitised in a week, or production costs of a project. It should be noted,

however, that while these other answers are valuable, it is advised that data analysts decide what they would wish to learn from the data before diving in. Many of the possibilities can become distracting if the main target is unclear or unidentified, and data can always be revisited.

NEXT STEPS

Since beginning the production logs in the spring of 2017, UNT's DPU has looked forward to expanding the method in order to provide a clear understanding of the entire lab's process beyond just the imaging. The authors have expanded production logs to metadata, inventory and quality control steps in the lab's workflow. After a full year's worth of data, a better understanding of the lab's capacity for projects as they enter and exit the lab will be more accessible.

Another area the authors have been developing is that of an automated data collection and analysis method. A pre-filled drop-down menu form that DPU students and staff could fill out to capture production numbers would eliminate a lot of irregularity within the data input. The form could also be the front end of a database that would assist the authors in generating reports on the data. The authors have consulted on, and are currently in the process of requesting the creation of such a tool from the local UNT Libraries information technology department. They have begun utilising Google Forms and Sheets as a prototype, but hope for a more refined and customised tool in the future.

Next, the DPU staff will need to get buy-in from other department, division and library members on re-evaluating project measurement and establishing workflows for accepting projects that account for the lab's capacity. As the DPU staff are not the primary staff members at the university either acquiring projects for digitisation or coordinating the digitisation for the majority of projects, it is especially important that all stakeholders are brought into the conversation and understand the importance of respecting the lab's capacity. With all involved personnel on board, the DPU will have the ability to stay within its means, better meet the expectations of its partners and grow its capacity in an intelligent way that understands, based on the data, where and how capacity can best be added. DPU staff members have already had preliminary discussions with library division managers and administrators on what deliverables would be ideal for fully implementing point estimations. It was decided that a succinct time cost sheet, similar to the pricing sheet the DPU has for estimating costs of external projects for partners, would help other departments better estimate the time a given project might take. DPU lab staff have begun crafting such a tool, using data collected from the production logs.

Armed with this capacity data, the next step for the DPU staff will be to use the numbers to measure potential projects and agreed projects *before* they come into the lab. This step, however, can only be achieved by working with DPU partners, both internal and external, to format and provide sufficient information so a capacity points number can be calculated. Unlike the NUL RDC, the bulk of projects digitised in the UNT DPU are not internally held materials from the university, but rather come from institutions across the state of Texas. Therefore, the DPU is often committed to projects that have not been physically seen, processed or counted by any staff member at UNT. Thus, it is important that DPU partners, particularly ones external to the university, provide accurate data and descriptions of the collections to be digitised in a way that makes the calculation of the commitment of that portion of the lab's capacity to that project accurate. DPU staff are in the process of developing project formatting guidelines and inventory requirements to accommodate capacity calculations easily and give accurate estimations of digitisation timelines to partners.

In addition to this, one other major step will be shifting the way in which projects are brought into the lab to better accommodate Point Process application. Currently, there are two methods by which projects enter the lab: with a detailed inventory and without one. Both need to be adjusted. For projects submitted through the DPU's Rescuing Texas History Program (RTH), an itemised inventory is required of all partners.[8] Many of those partners continue after the grant year to send projects to be digitised and follow the same inventory paperwork used for RTH. These forms are extremely helpful to lab staff as the project enters into inventory, digitisation, metadata and out-going inventory steps. The current form, however, does not require a formal count of objects or a detailed material classification type. In order for the capacity number to be most effective, the projects themselves must be described in such a way as to apply the item-type numbers. The second way a project enters the lab is without an inventory at all. This situation makes the inventory process more involved than if there had been an inventory. It also puts the onus of counting objects and/or pages on lab staff. DPU staff have begun conversations and blueprinting on how to change these structures to one that would set boundaries on what format and information about a project is acceptable and required. A draft of the inventory that takes into account the number of objects and/or pages has already been created and is undergoing testing for measuring project size. A proper explanation of how to classify and count would also need to be provided to partners, so that materials would arrive with the desired documentation.

Lastly, while many of the above will require much time and gradual rollout, one step the DPU has already taken is to create project charts. This idea came from a tour taken by one DPU staff member at Stanford University's Digitisation Labs,[9] where the entire time and cost spent on digitisation is tracked for each individual project. The project charts at UNT will record all the production information related to a specific partner project. Thanks to the production log, DPU staff are in many cases able to create these project charts retroactively from the data available in the logs. Going forward, each new project brought into the lab will have a project chart generated for it, informed largely by the numbers gathered from production logs. These charts will provide a much needed verification of the lab's digitisation cost, estimates for digitisation time and offer a way for staff to evaluate and improve upon the lab's current production numbers.

CONCLUSION

While the UNT DPU's effort to replicate the NUL RDC's Points Process focused on measuring the lab's imaging capacity, it has already borne many other fruits. The numbers gathered from collecting data in production logs for each student worker have already enabled DPU staff to estimate imaging times and cost for both internal and external projects. The logs have also provided holistic project tracking and attendance for student workers. Further expansion of the logs to metadata, inventory and quality control will likely result in similar gains for the lab. The plans for developing new processes and technology for production tracking will further enable the refinement and application of the lab's capacity numbers from these invaluable first steps. As the ultimate goal for measuring the lab's capacity is to better regulate and manage the scheduling of digitisation projects, it is hoped the capacity project will be embraced by all involved personnel at UNT, improving the throughput of projects through the lab, and by extension the efficiency and productivity of the UNT DPU. The replication process done at UNT proves the validity and usefulness of the Points Process for measuring capacity and could be equally replicated at other institutions as well.

References

1. University of North Texas Libraries (2017) 'Digital projects unit equipment', available at: https://www.library.unt.edu/digital-projects-unit/equipment (accessed 16th March, 2018).
2. Young, J. and Zellner, D. (2016) 'Snapshot: managing a portfolio of projects', paper presented at the Digital Library Federation Forum, Milwaukee, WI, 7th–9th November, available at: https://web.archive.org/web/20170314224806/https://www.conftool.pro/dlf2016/sessions.php (accessed 16th March, 2018).
3. Willis, S. and McIntosh, M. (2017) 'Cost and feasibility of FADGI compliance for the digital projects unit: final report', available at: digital.library.unt.edu/ark:/67531/metadc1020928/ (accessed 16th March, 2018).
4. Zellner, D. (2017) 'Project portfolio management at Northwestern University Libraries — developing an innovative solution to managing project life cycles', *Journal of Digital Media Management*, Vol. 6, No. 2, pp. 195–204.
5. *Ibid.*
6. *Ibid.*
7. Koelling, J.M. (2004) 'Digital Imaging: A Practical Approach', AltaMira Press, Walnut Creek, CA.
8. McIntosh, M., Mangum, J. and Phillips, M.E. (2017) 'A model for surfacing hidden collections: the rescuing Texas history mini-grant program at the University of North Texas Libraries', *The Reading Room: A Journal of Special Collections*, Vol. 2, No. 2, pp. 39–59.
9. Willis *et al.*, ref. 3 above.