User Study and Survey on Material-related Experiments

4CeeD Design Team
University of Illinois at Urbana-Champaign
Website: http://t2c2.csl.illinois.edu

I. Introduction

In this technical report, we present results of our user study and survey on material-related experiments and instruments. The main objective of this study is to understand the user requirements to develop digital tools to support capturing and processing data generated from material-related experiments. In addition, by studying actual usage of scientific instruments, we also obtain important insights from the workload generated from the instruments, which can be vital in designing system to support archiving and processing material-related data.

The technical report is organized as follows. In Section II, we provide some background information on the state-of-the-art methods in capturing, transferring, and managing digital data from material-related experiments. In Section III, we present the methodology and results of our survey to verify the necessity and practicality of digital data platform for materials research. In Section IV, we present our study on the actual usage of scientific instruments in materials research, which shed light on the characteristics of potential workload that a digital data platform needs to handle.

II. Background

To better understand the target environment of material-related experiments, we provide in this section some background information on the materials, semiconductor experiments, and analytical instruments used in Materials Science research. In addition, we present some insights from our user study to shed light on the user requirements and expectations.

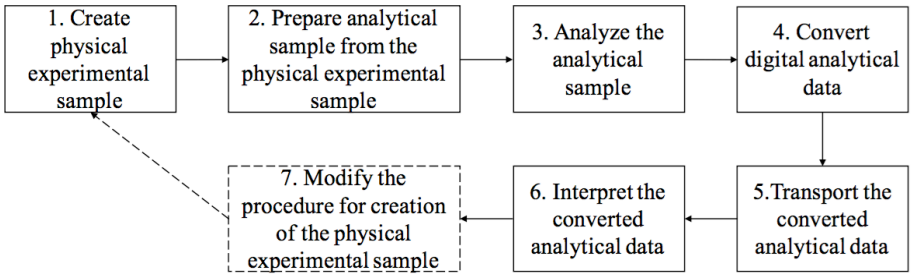
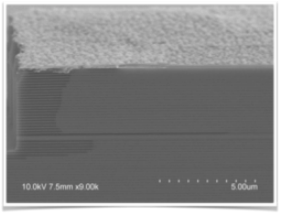


Figure 1 – Typical experimental flow in material research.

Figure 1 shows a typical experiment flow in material research. In the first step, researchers create physical experimental samples, either in their labs or in shared fabrication facilities. These physical samples can range from microelectronic devices, to biological samples, to nanoparticles. Once physical samples are created, they must be prepared for analysis (Step 2). For example, with analysis using Scanning Electron Microscopes (SEM), the preparation usually involves cutting the sample into a size which can be placed under the microscope and attaching it to a SEM sample holder. The result of such preparation is called analytical sample. The actual analysis of analytical sample happens using analytic tools (Step 3), including SEM and other electron, scanning probe, or optical microscopies could be performed, as well as x-ray, ion, electron, and optical scattering experiments.



Result image of 07302013-Oxidation experiment

Experimental setting:

Time 13min
Temp 425 C

(Structured meta data)

Notes:

Oxidation depth is about 12um. Oxidation layer composed of Al(0.98)GaAs with thickness of 30 nm. Furnace in 2111 MNT L, 2" diameter quartz.

(Free text)

Figure 2 – An example of SEM output.

The results of the analysis in Step 3 are the digital footprints of the physical experimental samples. These digital data can vary in format, depending on the type of analytic instrument used. For example, the output of SEM microscopes (Figure 2) consist of: (i) digital images of analytical sample that are stored in standard image format (e.g., .TIF, .GIF., or .JPEG), (ii) instrument specific information and meta-data (e.g., temperature, pressure, accelerating voltage, detector used, etc.) that are stored in a text file, and (iii) unstructured notes by researchers about the experimental or analytical results. On the other hand, output data from the TEM microscopes is in proprietary data format (i.e., DM3) that contains both image data and instrument specific meta-data. In such the case of proprietary data format, it might require another step to convert the results of analytic tools (Step 4). The researchers must then transport the converted files to their personal workstation (Step 5 - which often uses a "sneakernet" of USB thumb drives) for follow-up interpretation (Step 6). If the interpretation result is negative, further modifications might be needed for the procedure to create physical experimental sample (Step 7), which causes repetitions of the process until the desired criteria are satisfied.

While new analytic techniques have allowed for a surge of nanomaterials research publications and related innovative products, the time between discovery of new materials and application in semiconductor fabrication processes is at a relative stagnation, taking several years between an incepted material design and its commercial usage. This slow process can be attributed largely in part to communication of research, or rather the lack-there-of, specifically pertaining to nanomaterial analysis tools. Most often negative results from these nanomaterial analysis tools are not published, the transportation of the collected data is often insecure, and the resulting data files are often propriety causing inherent loss of data through file conversion in order to work up the data for publication quality figures.

In order to accelerate the experimental process, it is necessary to have an expedient mean to capture and process the digital data (i.e., output of Step 3) in real-time and in trusted manner before archiving, further analysis and visualization for more efficient interpretation of the results. Such a distributed real-time and trusted framework would greatly reduce the time, security and data loss risks of the manual efforts involved in the Step 4, 5, and 6 of the experimental process. In addition, a networked platform that provides authorized access to archived experimental data would help close the communication gap between researchers and prevent unnecessary repetitions of the experimental process caused by the lack of information in the literature.

III. User Survey

A. Survey Methodology

To further verify the necessity and practicality of such framework, we undertake a user study by surveying users of Materials and Nano Technology Lab (MNTL) and Materials Research Lab (MRL) in the University of Illinois at Urbana-Champaign.

The survey requests were sent via emails directly to users. Our email list was provided by internal lists from MNTL and MRL. This list is made up of over 1100 undergraduate students, graduate students, staff, and researchers. The survey itself was only accessible through the shared link in the email request and was anonymous. The list of questions asked in the survey can be seen in Table 1. We sent the survey 2 times over a span of 3 months, and received 51 responses in total, mainly from MNTL and MRL (Figure 3).

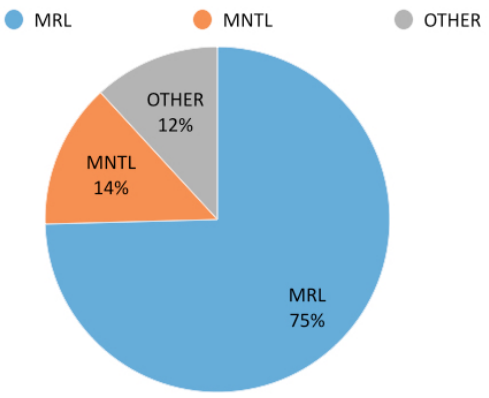


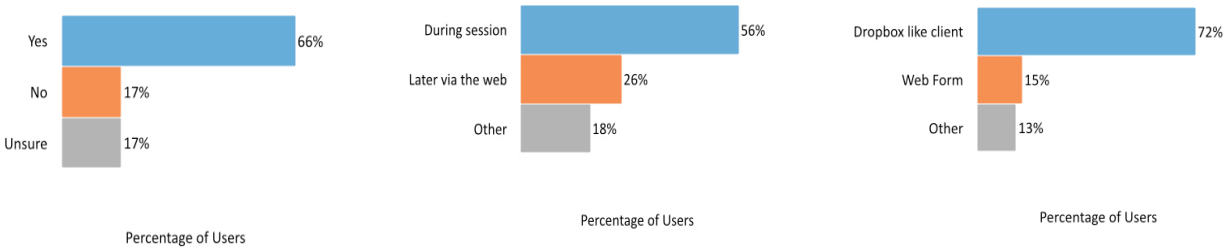
Figure 3 – Demographic of surveyed users by affiliations.

Q1: What facility do you work in?
Q2: What equipment do you use from a selection of TEM and SEM's?
Q3: What other equipment would you like to include?
Q3: What are the typical number and length of lab sessions you do per week?
Q4: What are the image and instrument metadata would you like to collect?
Q5: How much data per session would you want to upload?
Q6: How are you currently transferring data from the lab to your personal machine?
Q7: Do you feel that you have enough time to upload data to a repository during their session?
Q8: How do you organize data on your personal machines? (For example: Folders who have the date for their name.)
Q9: When would you prefer to upload their data (during the session, or later)?
Q10: How would you prefer to upload data to cloud storage (via web form, Dropbox like client)?
Q11: How would you want to search over your data? (keyword, date/time, experimental settings, file types, instruments, experiment type, other)
Q12: If data management system for material-related data is available, are you interested in using it?
Q13: What suggestions would you have for such a scientific data management tool?

Table 1 - Questions asked in the user survey

B. Summary of Survey Results

In terms of the ability to upload experimental data during lab sessions, the results from the survey show that 66% of them feel they have enough time during the session to upload the data if such a data acquisition tool exists (Figure 4a), and that users prefer a tool that is similar to Dropbox client to transfer their data out of the lab (Figure 4c).



(a) Answers to Q7 (b) Answers to Q9 (c) Answers to Q10

Figure 4 – User responses to questions on uploading data directly from the lab sessions.

However, majority of users currently utilizes a "sneakernet" method to transport data from the lab. Specifically, 96% of the users use USB thumb drive to transport data from the experimental session to their office for further analysis (Figure 5a). The survey results also encouragingly show that nearly 80% of users are interested in using such a framework for data acquisition, analysis, and a distributed platform for archiving and sharing data (Figure 5b).

(a) Current way of transferring data out of the lab (b) User interests in using digital tool to upload data from the lab

Figure 5 – User responses to questions on methods to upload data from the lab.

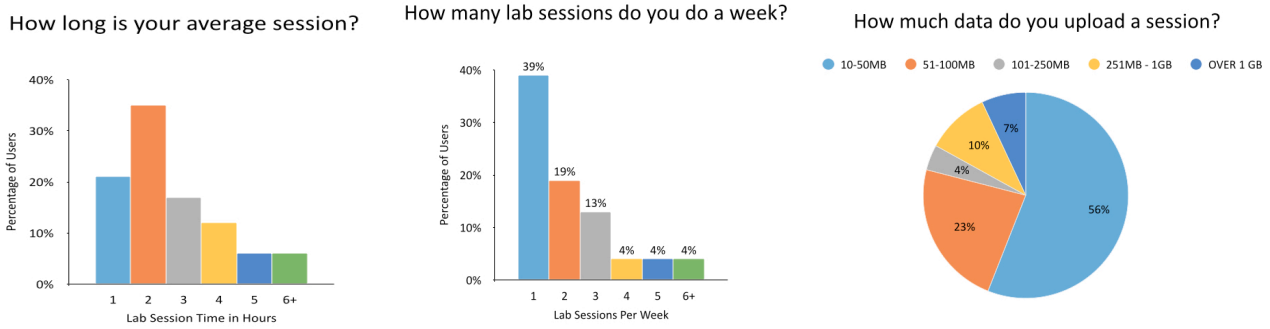


Figure 6 – User responses to questions on their lab session usage.

The survey also points out some challenges in building such a framework. First, the scale of data generated during lab session tends to be different from user to user, as shown in Figure 6. In addition to the large number of users from multiple research labs who might work concurrently during peak hours, the system infrastructure should be scalable and capable of dealing with varying workloads.

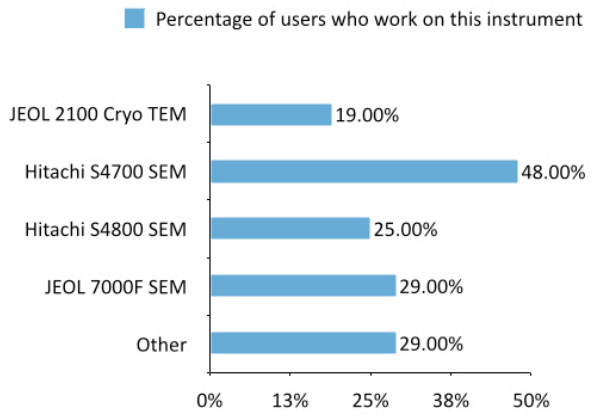


Figure 7 – Instruments used in experiments.

Second, researchers use a wide variety of instruments for their experiments. While Figure 7 shows the most popular instruments (i.e., SEM and TEM), the long tail (i.e., "Other") consists of a very diverse set of instruments. Thus, the framework should be designed to support analyzing heterogeneous types of data generated from different types of instruments. On the other hand, by knowing the most popular types of instruments being used, we can put more focus on those types in designing evaluation and targeting potential users.

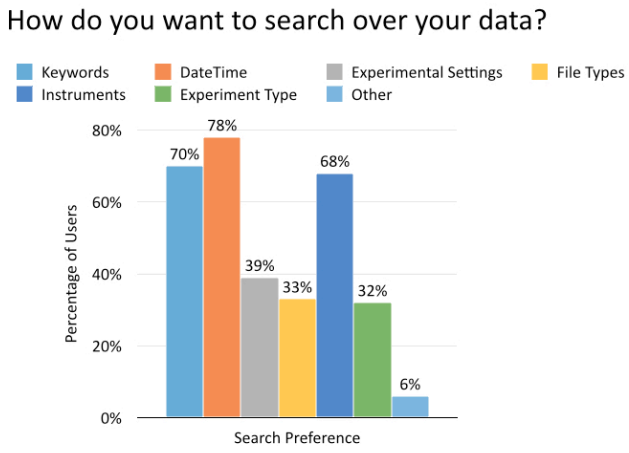
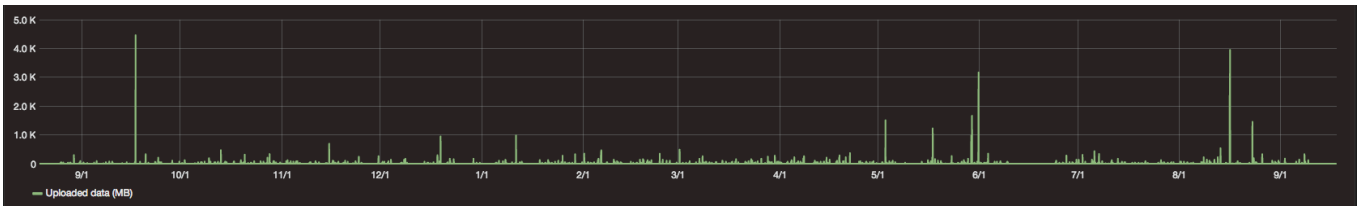


Figure 8 – User search preference.

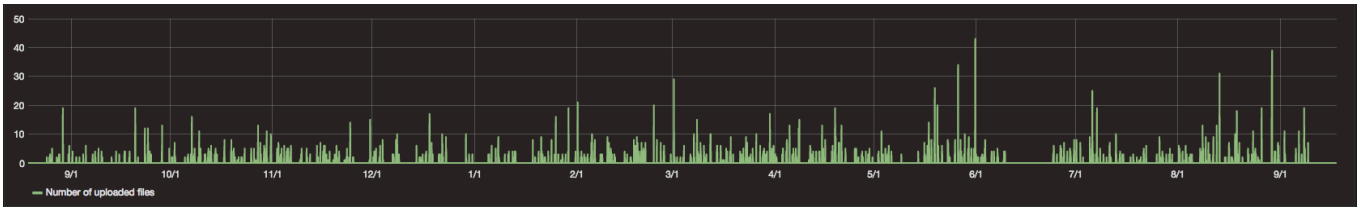
Third, as digital data is collected for wide access and sharing, users might want to perform search over shared repository of experimental data. The objectives can be to update/correct any missing meta-data/setting, erroneous information from user's uploaded experimental data, to learn from others' successful or failed experiments, or simply to look for related experiments for reference purposes. Our survey shows that users want to search over the data using a variety of structured information, such as instrument types, experimental types and settings, beside traditional keyword-based search (Figure 8).

IV. Instrument Usage Study

In addition to understand user requirements of digital tools for collecting and processing material-related data, we also study the actual usage of experimental instruments in material research environment. In particular, we select two of the most popularly used instruments in MRL, namely JOEL and HeliosFIB, and collect information about experimental results created on those instruments, such as created time, size of output file, etc, over a **one-year time period**. This information give us vital information about the actual usage of these instruments and the typical workload generated from those instruments.

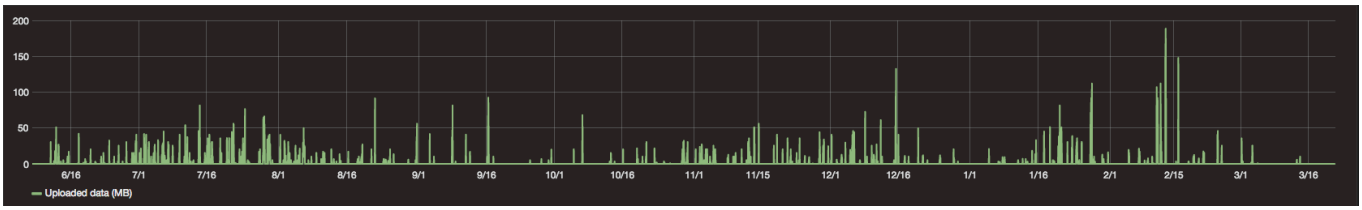


(a) Amount of data uploaded

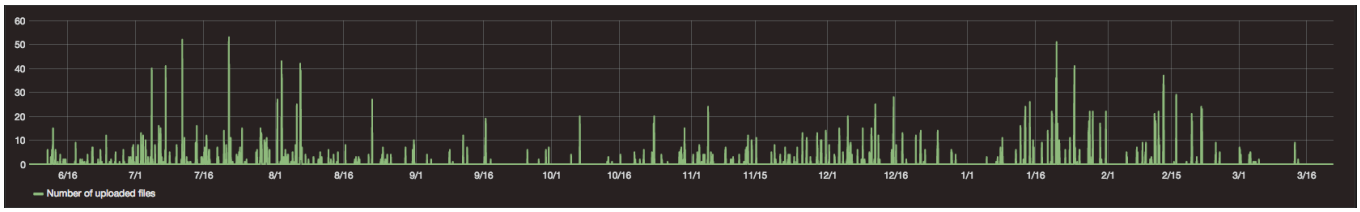


(b) Number of files uploaded

Figure 9 – Data usage on MRL's JOEL instrument from Sep 2015 to Sep 2016



(a) Amount of data uploaded



(b) Number of files uploaded

Figure 10 – Data usage on MRL's HeliosFIB instrument from June 2015 to March 2016

The results on instrument usage are shown in Figure 9 and 10. We can see that, on both instruments, the workloads are highly variable and often bursty. In general, there is a corelation between the number of files and the amount of data uploaded (especially in HeliosFIB case). However, for JOEL instrument, the variability in the amount of uploaded data seems to be more extreme, due to the fact that the files produced by JOEL are generally large files and can vary in sizes. These results suggest that a cloud-based system to capture and process the digital data generated from instruments needs to be scalable and highly adaptive to handle variable and bursty workload.

V. CONCLUSIONS

In conclusions, our user study and survey results show that there are strong interests in using digital platform for capturing, curating, coordinating, correlating, and distributing material-related experimental data. The survey also points out some specific requirements for such a platform.

In terms of user-facing features, in order to save time at microscopes, the platform should require minimal interactions with users during lab session to upload and transfer experimental data. In addition, since the targeted users are non-IT people, the interface should be intuitive and simple to use. The platform should be able to support various types of input data from different types of experiments and instruments and provide an extendable and flexible data model for inputting data to support diverse data types and use cases. To support data discovery, the platform should support users the ability to search through shared data repository by efficient filtering of structured and meta data.

In terms of back-end system, since the uploaded experimental data can be of various types and formats, the platform's back-end should be able to support heterogeneous types of data processing jobs, each job corresponds to a type of uploaded data. In addition, from the results of instrument usage study, we can see that the back-end system needs to be scalable and capable of dealing with the variable and bursty workload generated from instruments in the labs.