

## **A central repository for integrating study progress information from disparate sources**

Ana Sofia Morais

Kiron Open Higher Education

The study journey of a Kiron student is punctuated by one's participation in various study-related offers, including massive open online courses (MOOCs) and language courses, in addition to student support services like study mentoring and university-transfer guidance. By successfully taking part in these offers, students progress in their studies, and develop towards a more advanced state of knowledge and competence. In this paper, we define *study progress information* as data about a student's participation in MOOCs and student support offers. This type of information can be entered manually by Kiron employees, or recorded automatically online, as students use the Kiron online platform or the platforms of MOOC providers.

Study progress information is useful to Kiron for various reasons. First, this data allows for measurements of curricula quality; namely, how attractive they are to students and how successful the students are at completing courses. Second, student progress information enables Kiron to evaluate the effectiveness of its student support offers in increasing student engagement and course completion rates. Third, information about which study and support offers the students already completed can be used to identify which study offers the students may benefit from in the future. Finally, information about the students' study accomplishments can help Kiron counselors gauge the students' needs and possibilities in working toward the recognition of Kiron credit points by a partner university.

Until a while ago, however, technological barriers prevented Kiron from capitalizing on study progress information, namely, the fact that Kiron's study progress data comes from disparate sources and in different formats. Without a central data repository that integrated this information, extensive time was spent gathering all relevant information and carrying out computationally-intensive data pre-processing operations, before the data would be amenable to analysis. In what follows, we describe the different sources of study progress information and the type of data that they provide. Next, we describe the central repository implemented at Kiron in order to integrate study progress information from disparate sources.

## 1. Sources of study progress information

There are three main sources of study progress information. One source of information are the *Kiron students* who can use the Kiron platform to inform the Kiron team about their course enrollments and completions on MOOC platforms such as Coursera and edX. As illustrated in Figure 1A, students inform Kiron about their enrollment in a MOOC by selecting “Enroll now.” But, even doing so, the student does not become automatically enrolled in the course. Instead, the Kiron platform forwards the student to the external platform providing the MOOC, where the student can actually enroll in the course. Once the student has informed Kiron about the course enrollment, the student will see the button “Finished!” (Figure 1B). By selecting “Finished!”, the student informs Kiron that he or she has completed the MOOC on the external platform. The data that the student provides Kiron with—by clicking on “Enroll now” and “Finished!”—is only self-reported and subjective because the actual, objective data about the student’s course enrollments and completions is property of the MOOC providers. As we shall see, the two kinds of course information—objective and subjective—do not inevitably align.

A second source of study progress data are some of the *organizations* that have partnered with Kiron to facilitate access to higher education for refugees. Coursera and Saylor, two of the MOOC platforms providing online courses from the Kiron curricula, provide us with data about the actual, objective course enrollments and completions of Kiron students on their platforms. For Coursera, a dataset is manually downloaded weekly by a Kiron employee. For Saylor, the data is automatically downloaded every day via an application programming interface (API). Since this data reflects the students’ actual activity on Coursera and Saylor, it can be used to evaluate the accuracy of the students’ self-reported enrollments and completions, if only for the subset of courses from the Kiron curricula that are hosted on those platforms.

Another organization that provides Kiron with study progress data is Volunteer Vision, a software company providing digital mentoring solutions. Volunteer Vision has partnered with Kiron to help students be successful in their studies, by providing them with study mentoring and language learning support. Just like for Saylor, the data from Volunteer Vision is automatically downloaded every day via an API. The downloaded datasets include information about the status of the student in their mentoring and language programs, such as the number of sessions completed and the date of their last session.

**MODULE: OBJECT-ORIENTED PROGRAMMING IN JAVA**
A

## Software Construction: Data Abstraction

This course is available.

Enroll now

Free certificate for Kiron students available [How does it work?](#)

Starting Date	self-paced
Finish before	15th August 2019
Workload	6 weeks 10 hours/week
Language(s)	English
Subtitle(s)	English
Institution	University of British Columbia
Provider	edX

Learn powerful data abstraction and decomposition techniques to build large, complex programs.

This is the third course in the Software Development MicroMasters program. You will learn how to build larger and more complex software systems using the Java programming language. The course begins with the topic of data abstraction - from specification to implementation. Particular attention is given to how to write robust tests using JUnit. Then the course expands on these ideas to explore how type hierarchies and polymorphism can be used to decrease redundancy in your code. The course wraps up with a discussion of how to design robust classes. By the end of the course, you will have a solid foundation in designing software in Java, and be ready to move onto Software Construction: Object-Oriented Design, where you will learn more complex design patterns and principles designing object-oriented programs. Learners who enroll in the Verified track will receive staff grading for the course project and increased interaction with the instructors and staff.

**MODULE: OBJECT-ORIENTED PROGRAMMING IN JAVA**
B

## Software Construction: Data Abstraction

You are enrolled in this course.
Not enrolled? [Change here](#)

Go to Course

Finished!

Free certificate for Kiron students available [How does it work?](#)

Starting Date	self-paced
Finish before	15th August 2019
Workload	6 weeks 10 hours/week
Language(s)	English
Subtitle(s)	English
Institution	University of British Columbia
Provider	edX

Learn powerful data abstraction and decomposition techniques to build large, complex programs.

This is the third course in the Software Development MicroMasters program. You will learn how to build larger and more complex software systems using the Java programming language. The course begins with the topic of data abstraction - from specification to implementation. Particular attention is given to how to write robust tests using JUnit. Then the course expands on these ideas to explore how type hierarchies and polymorphism can be used to decrease redundancy in your code. The course wraps up with a discussion of how to design robust classes. By the end of the course, you will have a solid foundation in designing software in Java, and be ready to move onto Software Construction: Object-Oriented Design, where you will learn more complex design patterns and principles designing object-oriented programs. Learners who enroll in the Verified track will receive staff grading for the course project and increased interaction with the instructors and staff.

Figure 1. Screenshots of the Kiron online platform showing an example of a course card. The course “Software Construction: Data Abstraction” is provided by edX and it forms part of the module “Object-oriented Programming in Java” of the Kiron study program in Computer Science. A) Students inform Kiron about their course enrollment by selecting “Enroll now”. B) After that, two other buttons are displayed, “Go to Course” and “Finished!”. The first button forwards the students to the platform providing the course; the second button can be used by the students to inform Kiron that they completed the course.

The last source of study progress information are the *Kiron employees*. For example, Kiron employees who organize German courses occasionally receive information from language schools about the participation of the students in a course, such as whether they have discontinued the course or completed it. Another example is that of student counselors who guide students towards university transfer. Student counselors can provide information about the participation of the student during the guidance process, for example, whether the student is waiting for counseling, already in contact with a counselor, or has stopped attending the counseling sessions. Finally, there is the case of Kiron employees who organize various offline student events and can provide information about which students took part in those events. In the absence of a central repository where they could enter this information, Kiron employees used various spreadsheets, formatted in myriad ways, each updated by a different team.

The information provided by the students, the partner organizations, and the Kiron employees helps draw a useful picture of a student's study journey with Kiron, and in particular, of the status of his or her participation in various study offers. The change in the participation status of a student from "enrolled" to "completed" is indicative that the student developed towards a more advanced competence state, and has therefore progressed in his or her studies with Kiron. Moreover, independent data from multiple data sources helps increase the quality of the data used by Kiron; in particular, data from MOOC platforms like Coursera can be used to evaluate the accuracy of the course enrollments and completions reported by the students.

## **2. A central repository for integrating study progress information**

Since Kiron obtains study progress information from disparate sources, there was the need for the construction of a central data repository that could integrate all information into a single location, making it amenable to analysis and reporting. With the support of the H&M foundation, Kiron implemented a central repository for integrating study progress information, called Participation Tab ("PartiTab", for short). The data is stored in this central repository in a way that is secure, reliable, and easy to retrieve and work with. Access to the data stored in the repository is restricted to selected Kiron employees.

The first step in implementing the PartiTab was to extract the data, which included gathering large amounts of data from multiple sources. Subsequently, we normalized the data in order to keep the naming of similar fields constant across different sources. We then built the capability to store this data as a series of sequential events. Over time, new data is added to the PartiTab, as new information accumulates in the different sources. While adding new data, we

keep a track record of changes with timestamps, so that we know which source reported what at any given time.

The resultant central repository integrates information from multiple sources about the participation of students in courses and student support services. Participation is described by the three information sources introduced above—the student, a partner organization, and a Kiron employee. Each information source describes the student’s participation in a particular course or service by selecting one among a set of possible states. Let us illustrate this with an example of the participation in a course. Each source describes the participation of the student using one of three possible states: “enrolled”, “completed” or “none”, the latter indicating that Kiron has no information from the source about the participation of the student in the course. In case the course is neither from Coursera nor Saylor—so far the only MOOC providers providing Kiron with data about its students— the status indicated under “partner organization” is “none”. With three information sources and three possible states, participation in a course can be described by one of 27 different combinations of states across sources. The same logic applies to the student support services, except the set of possible states will differ in number and meaning.

In case conflicting states are reported by the different sources, the PartiTab indicates which state information should be preferred. The preferred state is the state that the Kiron team believes is more likely to be true, based on existing evidence about the accuracy of the different sources. Take, for example, the case of online course completions. The state reported by a MOOC provider like Coursera is usually preferred to the state reported by the student, as self-reported data is more likely to be biased. The selection of one among conflicting states is particularly useful when using study progress information to recommend new courses and support offers to students via the Kiron online platform.

The underlying architecture of the PartiTab has three key practical advantages. One is that it keeps being efficient and practical when applied to a larger number of study offers, larger inputs of data, or data from other MOOC providers. Second, it is able to track change within a student over time, by keeping a history of the changes in a student’s participation in various study offers. Third, since the data might arrive from the various sources with delay, it allows the Kiron team to retroactively update the participation state of a student.

### **3. Example of a data analysis enabled by the central repository**

In this section, we present the results of a data analysis that would have been more difficult to conduct without the central data repository. The analysis was motivated by the following

research question: How accurate are the students' self-reported course enrollments and completions? To tackle this question, we compared a) the students' self-reported data regarding their participation in Coursera courses and b) the actual, objective data that Coursera shares with Kiron. We take the data from Coursera to represent the "truth" about a student's activity on their platform, and we use it as a benchmark against which the accuracy of the student self-reports can be evaluated.

Students who joined the Kiron program before September 2016 were excluded from the analysis because the objective data from Coursera only shows the course activity of Kiron students after that date. This left us with a sample of 3585 students for the analysis. For each student, we calculated the following two quantities: a) the difference between the student's self-reported number of enrollments in Coursera courses and the number of enrollments reported by Coursera for the same student, and b) the difference between the student's self-reported number of completions of Coursera courses and the number of completions reported by Coursera for the same student. For both a) and b), a zero difference indicates the student reported the same number as Coursera, a positive difference indicates that the student reported a higher number (i.e., overreporting), and a negative difference indicates that the student reported a lower number (i.e., underreporting). Table 1 shows the percentage of students who reported correctly, underreported, or overreported their course enrollments and completions. Note that the percentages in each row add up to 100%.

Table 1  
Percentage of Kiron students who reported correctly, underreported, or overreported their course enrollments and completions

	Percentage of students who reported correctly	Percentage of students who underreported	Percentage of students who overreported	Percentage of students without information
Enrollments	47	18	27	8
Completions	75	5	12	8

Note. Results obtained on the 1<sup>st</sup> of October, 2018.

These results show that the students were overall more accurate at reporting their course completions than their course enrollments. While 75% of the students reported their course completions correctly, only 47% reported accurate course enrollments. Second, the percentage of students who reported inflated course enrollments and completions (27% and 12%,

respectively) was higher than the number of students who underreported (18% and 5%, respectively). These findings advise us not to accept students' self-reports blindly, especially regarding their course enrollments, and raise our awareness of the fact that students' are more prone to overreporting their course activity than they are to underreporting it.

#### **4. Conclusion**

Disparate sources of study progress information led Kiron to implement a central data repository for integrating large multi-source datasets on a single location, making the data easy to retrieve and analyze. As illustrated by means of data analysis, the implemented solution enables Kiron to turn the problem of disparate information sources into an opportunity to evaluate the quality of the self-reported data and better gauge the progress students make in their studies with Kiron.