

HHS Public Access

Author manuscript *J Comput Sci*. Author manuscript; available in PMC 2018 May 01.

Published in final edited form as: *J Comput Sci.* 2017 May ; 20: 205–214. doi:10.1016/j.jocs.2017.03.010.

Biomedical Big Data Training Collaborative (BBDTC): An effort to bridge the talent gap in biomedical science and research

Shweta Purawat¹, Charles Cowart¹, Rommie E. Amaro², and Ilkay Altintas¹

¹San Diego Supercomputer Center, University of California, San Diego, USA

²Department of Chemistry and Biochemistry, University of California, San Diego, USA

Abstract

The BBDTC (https://biobigdata.ucsd.edu) is a community-oriented platform to encourage highquality knowledge dissemination with the aim of growing a well-informed biomedical big data community through collaborative efforts on training and education. The BBDTC is an e-learning platform that empowers the biomedical community to develop, launch and share open training materials. It deploys hands-on software training toolboxes through virtualization technologies such as Amazon EC2 and Virtualbox. The BBDTC facilitates migration of courses across other course management platforms. The framework encourages knowledge sharing and content personalization through the playlist functionality that enables unique learning experiences and accelerates information dissemination to a wider community.

Keywords

e-learning; biomedical; collaborative; big data; education

1 Introduction

Researchers increasingly rely on Big Data, Computational Data Science, and High Performance Computing (HPC) to solve problems within scientific domains and explore them in new ways [10]. However, it is often difficult to effectively apply these approaches in practice. They draw heavily from domains including computer science and applied mathematics, and most researchers are not substantially invested in them [2]. Moreover, petascale and exascale datasets require new data science techniques to process and manage. Many practitioners have not encountered these new techniques in their studies or work [10].

University training is largely offered through a multidisciplinary Computational Science discipline, or as specializations in existing domains such as computational chemistry, computational physics, etc. [2]. These offerings cannot keep up with current demand, or the rapid changes in technologies and practices. Students are not entering the field with enough real-world skills, and there aren't enough programs to rapidly affect a change in the talent gap.

To overcome this challenge, we envision to cultivate an online community focused narrowly on data science and computing in biomedicine, and fostering high quality, well informed, and freely accessible knowledge. Our community, the Biomedical Big Data Training Collaborative (BBDTC), targets the development of technical skills as well as education. Our effort will assist educators in developing and communicating best practices for content development and deployment, along with adaptive learning, assessment metrics, and testing practices.

Our educational material not only includes traditional video and text but all of the data and tools required to participate in our courses or learn independently. Tools come preinstalled and preconfigured on virtual machines that can be run on a local workstation, or in Amazon EC2 cloud. Users can adapt existing material to their needs, by assembling parts of multiple courses into their playlists. Our training material is accessible to a broad community of biomedical researchers and students, including those without access to high performance computing facilities.

By focusing on technical training as well as education, providing real-world datasets, and presenting packaged toolkits that are simple enough to learn yet powerful enough to perform real work, BBDTC differentiates itself from other online learning providers.

Unlike university courses, BBDTC's online presence makes it available to more content providers as well as more students. Being a community-driven resource allows it to evolve quickly and be more responsive to user needs.

The rest of this paper is organized as follows: In section 2, we describe the software architectures of our framework, its user interface, our repository, its playlist builder, and how it integrates with other important software platforms. In section 3, we define our user scenarios. In section 4, we cover our current usage statistics. In section 5, we outline related work. In section 6 we conclude, and discuss future work.

2 Approach: BBDTC Framework

BBDTC framework utilizes and extends HUBzero [8], an open-source software platform that facilitates course creation and content management. BBDTC utilizes HUBzero's LAMP stack, a web-portal comprising the Joomla content management system, Apache web server, and MySQL database. BBDTC extends HUBzero modules whenever required to meet demands unique to the biomedical big data research community. The platform is extended to Amazon EC2 Cloud infrastructure to enable scalable big data experiments. To facilitate offline learning, users can download complete packages as Virtualbox images from the BBDTC platform. The HUBzero platform is further enhanced through the implementation of cross platform integration module. Cross Platform Integration enables instructors to rapidly build content by leveraging their already existing courses on other platforms such as edX in a seamless manner.

The BBDTC framework comprises three main components: (*i*) BBDTC User Interface; (*ii*) BBDTC VM Toolbox; and (*iii*) Cross platform integration module. The components are explained in further detail in successive sub-sections.

2.1 BBDTC User Interface

BBDTC User Interface is designed with simplicity and usability at the center. It consists of four major sub-components: (*i*) course building interface for instructors and course managers; (*ii*) course viewing interface for learners; (*iii*) social media interface for creating an involved learning atmosphere; and (*iv*) a repository with playlist functionality for every BBDTC contributor.

Course Building Interface—The platform offers several features to create an interactive online class through a simple and intuitive interface. A course can be tailored to create multiple 'offerings' with a subset of the content changed or updated. An offering is a collection of 'units'. A unit can be further broken down into one or more 'lessons'. Instructors can upload materials, assignments, assessments, tools using all or any existing options for embedding a video, uploading files, including a wiki page and attaching weblinks. Videos are hosted on BBDTC's YouTube channel¹ and streamed on a BBDTC course page on user's computer. Other existing YouTube videos can also be streamed via the BBDTC id the instructors can track the progress of each learner through course progress bubbles, the real-time updates on quizzes taken, and the breakdown of scores and grades. Figure 2: BBDTC User Interface Screenshots demonstrations a course outline page, an embedded video lecture, an online quiz, learner's grade reporting and progress tracking dashboard.

Course Viewer Interface—Course viewer interface is for learners. Learners can only view the course contents but cannot upload or edit materials. They can track their personal progress and grades.

Social Media Interface—Learners can utilize functionalities such as forums, discussion boards, blogs and groups to enhance their knowledge through community involvement. A forum provides an online discussion area where people can exchange queries, difficulties, concepts and solutions in the form of messages. People with common interests can form peer groups. The group can share contents and exchange conversations privately or publicly. Individuals in a group can create new forums for discussions. Research has shown that collaborative techniques such as Pair Programming [11] result in improved learning outcomes. BBDTC promotes a connected learning atmosphere by providing multiple channels to collaborate and enhance the learning experience through peer engagement.

The BBDTC repository and playlist module is discussed in the next sub-section.

2.2 BBDTC Repository and Playlist Builder

Each learner is unique and needs a course structure that matches her personal learning requirements. This uniqueness is characterized by attributes such as diverse learning pace, prior knowledge of a subject, etc. The playlist feature enables a personalized learning experience through content customization that meets an individual's requirements. BBDTC

¹The BBDTC YouTube Channel: https://www.youtube.com/channel/UCj9gpCafVV23WcP_b66Wkdw

J Comput Sci. Author manuscript; available in PMC 2018 May 01.

offers a central repository where users contribute by uploading course modules, tools, and datasets. BBDTC allows researchers to create customized catalogue for personalized education by selectively adding content from the repository. The playlist feature enables users to add materials from the repository and/or upload new content and arrange them in an order that aids their personal learning requirements.

Resources—Users can contribute new knowledge content for the biomedical big data community training and education using our resources framework, also known as the BBDTC repository. BBDTC offers users the flexibility to upload content in the form of documents, datasets, tools, teaching materials, workshops, seminars, courses, and playlists. On placing the cursor over the *Discover* tab at the BBDTC home page, a drop-down menu with the word 'Resources' will appear. The BBDTC users can navigate to Resources web page and select the appropriate category to add their content. These categories are shown in Figure 3: Resource categories in the BBDTC.

The BBDTC allows users to add a description, multiple authors, and associated tags, during the process of publishing resources. Users can distribute resources privately or publicly.

Playlist Builder—The BBDTC enables users to create playlists through the playlist builder interface. Users can mix and match content from YouTube, the BBDTC Resources, and the BBDTC Courses, in addition to uploading their videos to create a customized course that matches their learning preference. The Playlist component has three major segments: 'My Channel', 'My Playlist' and 'My Videos'. Each user is provided a personal channel that is similar to a workbench: their personal playground for experiencing new knowledge.

One of the critical goals of BBDTC is to accelerate community driven learning and crosssharing of knowledge. To this end, users can subscribe to other BBDTC user channels from My Channel page. BBDTC users can create and manage custom playlists through My Playlist webpage shown in Figure 4: My Playlist Interface. My Playlist exhibits all the playlists created by a user organized in a tabular format. Additionally, information such as playlist name, description, and number of videos for each playlist is displayed. With 'Add Playlist' button, users can create a new playlist. Users can play, view, edit and delete existing playlists through the interface.

The BBDTC users can add videos to a playlist from the My Videos page shown in Figure 5: Add Video. Users can select to add videos from Youtube and or upload videos from their computer. In future, we plan to extend this feature by integrating with Courses and Resources. Users can index their videos using tags and categories. Privacy is paramount to any online service. Users have full control over each video's audience through public and private attributes.

The Playlist component manages viewership of videos and displays popular videos among the BBDTC users based on the number of views. The most viewed videos among the BBDTC community populate this list as shown in Figure 6: Popular and Featured Videos. Popular videos give each user a snapshot of currently interesting learning topics based on community viewership. To boost dissemination of new knowledge, administrators have the

privilege to upgrade videos to *featured* videos, which are presented to users separately. The playlist feature was introduced using the Joomla HD Video share component [6].

2.3 Toolbox Integration: Cloud and Local

In line with our goal of maximizing users' time for learning rather than installing tools, we introduce tailored pre-installed software environments that enhance the learning experience. Users can access these toolboxes over the cloud through Amazon EC2 and locally through Oracle Virtualbox virtualization technologies.

Users with access to the internet can remotely connect to pre-built Amazon EC2 Virtual Machines for accessing biomedical big data tools and perform scalable big data experiments. We also support local execution in equally simple style via downloadable Virtualbox Images. Every downloadable VM contains a standardized software tool environment with all required software packages and test data. The BBDTC provides an efficient software distribution mechanism using ready-to-go Virtual Machine Toolboxes. These toolboxes free the user of installation and configuration hassles. Users can directly get to the task of problem solving using specially customized virtual machines. The notable features of the two categories is as follows:

- **Downloadable VM toolbox on Local Machine:** In this category, the software tools are shared as a downloadable Virtualbox Image. The VM toolbox allows users to get a standardized software tool environment with their required software dependencies and the necessary test data on their personal machines.
- Amazon EC2 VM Toolbox on the Cloud: This category provides a ready-to-go solution for scalable big data experiments, saving users from setting up a local machine with necessary hardware specifications. Instead, the users can directly access an Amazon EC2 Virtual Machine through the BBDTC website using a light-client machine. The system utilizes noVNC [7], an HTML5 based VNC client. The noVNC web-client communicates with a remote VNC server on Amazon EC2 Cloud infrastructure through websockets. Figure 7: Amazon EC2 VM Toolbox Integration Architecture demonstrates Amazon EC2 toolbox integration with BBDTC platform.

We have verified and tested the features of the BBDTC course-building interface. The toolbox framework is live and linked to an existing online course on "*Scalable Bioinformatics Bootcamp*" that is available to learners through the BBDTC portal.

2.4 Cross Platform Integration

Organizing text, video, and other prepared materials into a new online course can be a significant effort. Once the new course has been designed, the metadata capturing this organization cannot always be migrated outside of the online application. The time taken to develop a course online discourages researchers and educators from redeveloping the same course across several MOOCs, encouraging 'vendor lock-in' to a particular provider.

The MOOC provider edX has taken steps to remedy this by defining a specification for *XBlocks*, hierarchical components of text, video, assessments, etc., that can be used like

building blocks to construct a new course. Courses constructed using edX's *edX Studio* can be exported as an XBlocks blueprint to an author's local computer, merged and modified by third-party tools, and imported back into edX.

BBDTC has developed a 'back-end' tool to process this exported data from edX Studio, and import it into BBDTC as new courses and resources. This tool has provided biomedical big data educators and researchers, who made early investments in edX, the capability to also provide same content in a centralized framework that is BBDTC. Similarly, BBDTC is currently working on a tool to perform the reverse operation, and export courses and resources from BBDTC as XBlocks blueprint; this will enable courses created using BBDTC's course building interface to imported into edX as new courses, or as a library of content to make new courses from.

3 User Scenarios

User scenarios have proven very useful for defining the features of BBDTC for a diverse audience. Using scenarios, we modeled the behavior of the BBDTC for users with different goals, in clear, nontechnical terms. This technique allows us to repeatedly review our user scenarios against the design and implementation of our site and determine if users can easily accomplish those goals. The Table 1: User Scenarios describes four major user types and scenarios for BBDTC ranging from academic students and instructors to freelance audiences and content providers.

The flowchart shown in Figure 9: User Scenario Flowchart describes how a learner can successfully complete a course material on the BBDTC. The flowchart illustrates the stepby-step procedure to navigate through the portal to complete a course. Consider a user scenario where Alex, an undergraduate student, first creates an account on the BBDTC platform. He enrolls in the required course and gets started with the first unit. He watches all the video lectures, reads all the material associated with the unit. He posts and participates in a discussion forum for any questions or difficulties with the unit. He completes the unit upon getting the required score in the assessment. He repeats the same procedure for all units. After completing all the units, Alex takes the final assessment for the course. BBDTC allows instructors to set the maximum number of times a student can take an assessment. Alex successfully completes the course by exceeding the passing grades established by the instructor.

4 Usage Statistics

In the first year of the development for BBDTC, we offered three complete and open biomedical training courses during the alpha release of the framework. We received an excellent response with a total of 162 users registered on the BBDTC website in the year 2015. During the second year, the BBDTC user-base increased to 484 registered users.

The National Biomedical Computation Resource (NBCR)² launched two biomedical training courses on the BBDTC platform in 2015: NBCR Summer Training Program - "Data

²National Biomedical Computation Resource (NBCR) website: http://nbcr.ucsd.edu

J Comput Sci. Author manuscript; available in PMC 2018 May 01.

to Structural Models" and NBCR & TCBG Joint Training Program - "Simulation-Based Drug Discovery". These week-long intensive training programs introduced new principles, methods, and NBCR tools to the biomedical community. The primary goals of launching the courses online via BBDTC were to outreach a broader user base of the biomedical research community. It also enabled us to test the platform capabilities and get feedback from the active users for further improvements. Following are the data analysis reports from the two courses:

Course 1, NBCR Summer Training Program - "Data to Structural Models"

We received a total of 31 enrollments for this course. BBDTC provided an interface to track learners course progress and grades. We observed 6 students accessed full course material and 9 completed the quiz. Figure 10: NBCR Summer Training Program - "Data to Structural Models" represents user-profile distribution according to Career level, Location and Continent.

Course 2, NBCR & TCBG Joint Training Program - "Simulation-Based Drug Discovery"

We received a total of 42 enrollments for this course. Figure 11: NBCR & TCBG Joint Training Program - "Simulation Based Drug Discovery" represents user-profile distribution according to Career level, Location and Continent.

Course statistics information delivered by the BBDTC enables content creators to understand our learners and provide a learning experience that goes beyond their expectations. It promotes course providers to experiment with new learning techniques and recognize how the learners respond. This structured feedback mechanism of BBDTC accelerates high-quality content creation by iterative upgradation of techniques.

In the year 2016, we additionally launched an offering for the course NBCR Summer Training Program - "Data to Structural Models" and a new course on "Introduction to Biostatistics."

User and Content Authentication

The BBDTC user-ship has grown steadily over the past two years. The online registration process is very similar to other sites. Once a user has supplied basic information about themselves, an email confirmation will be sent, and their account will await approval by an administrator.

At times, BBDTC will experience bursts of traffic where many new user accounts are created, some of which have proved to be fraudulent; the purpose of these accounts is to spam advertising. BBDTC takes advantage of standard practices, including ReCaptcha, and other authentication methods to filter out many such users. From experience, we've found that most fraudulent users are created using email accounts from free email services, while most actual users are associated with known educational institutions. We have written additional software modules to automatically approve new users from specific domains, reducing administrative workload and turnaround time of the approval process for the majority of our users.

User authentication is an ongoing process, and user-submitted content is scanned for inappropriate content.

5 Related Work

A series of short tutorial videos, of the kind commonly used in Coursera and other MOOCs, was recently used to teach two graduate-level biostatistics courses at the Public Health Program at the University of New Mexico [5]. Students met twice a week in a computer lab, with access to the instructor as well as required software. The instructor also held office hours twice a week. In this setting, students' overall perception of videos was positive, with 87.5% of the students stating that the videos were advantageous to their studying. Specifically, students found that shorter videos were easier to 're-watch' and 'skim'. "The ability to use the videos in a self-paced and repetitive manner was consistently cited as a major advantage by the students [5]."

Students didn't view the videos as a replacement for in-class instruction, but as a supplement; instructors found that the availability of the videos allowed for more time in class to be spent on a discussion. Most importantly, it was found that using videos promoted social support between classmates [5]. Although students would watch videos individually, they would reference and recommend specific videos when communicating with other classmates.

In another study, conducted across six different public institutions, students took the same introductory course on statistics [1]. At each institution, a control group took the course in a classroom-based setting, while a second group took a *hybrid course*. This hybrid course used software developed at Carnegie Mellon University to guide students' mode of instruction, with one face-to-face meeting with an instructor, per week. Results from this study showed that students in the hybrid course performed as well as students in the classroom-based setting, in terms of course completion, course grades, and performance on standardized tests.

Recently, a group of researchers conducted a series of surveys with over five hundred students taking a select group of online courses, over a period of three years [3]. Students found their experience with online instruction only moderately satisfactory, and hybrid courses were rated with higher satisfaction over purely online courses; lack of interaction with others was the most often reason cited for this dissatisfaction. The above research suggests that engagement with an instructor and/or a community is key to a student's overall satisfaction and performance when taking a course. Compared with older online learning paradigms, modern MOOCs are augmented with social networks that provide community and support through message boards and discussion groups [4].

Today's MOOCs are primarily targeted at serving individuals, rather than communities [4]. BBDTC instead focuses on cultivating a community, by offering multiple social networking tools and focusing on the biomedical big data community's particular needs. Rather than supplanting traditional education, the BBDTC expects to be successful by supplementing it.

6 Conclusion and Future work

Biomedical Big Data Science is a thriving and an exponentially growing community. Scientists need a quick-access gateway to high-quality knowledge content on topics that interest them. BBDTC is a community-oriented platform with an aim to promote knowledge sharing and accelerate information transfer rate from the field experts to an inquisitive learner. Smooth information dissemination will encourage new approaches and enable community members to combine tools from diverse domains to solve critical problems.

The collaborative provides essential tools to facilitate content development and deployment of biomedical big data training and education. The simple and intuitive platform enables the community to contribute at many different levels such as by adding course modules and playlists, adding new tools or working with the development team to introduce customized VM packages. BBDTC's fact based progress tracking dashboard and statistical reports enable content creators to understand learners and to provide a learning experience that goes beyond their expectations.

There are several areas in which we plan to expand our current work. To give an experiential view of a scientific facility, we plan to add Virtual Field Trips to BBDTC. Using this feature, community members can gain fundamental understanding of each other's strengths to better collaborate. This feature will enable labs to achieve higher visibility among their target community members.

To further facilitate easy dissemination of read-to-use tools, we plan to automate the VM toolbox generation process. In addition, we will enhance BBDTC's integration with Amazon AWS, by launching virtual machines using AMI id information from the BBDTC website.

For fast exchange of MOOC course structure, we plan to extend the ability to import and export feature to additional platforms that are gaining traction in the community. A userinterface to allow educators to individually import and export data will also be explored in the future. In the playlist interface, we plan to include access to course repository and resources from the same interface.

Finally, a tool is only as good as its utility to the stakeholders. We will build incentive based feedback mechanisms for biomedical researchers to accelerate knowledge sharing. We will aggressively increase community engagement with the key stockholders, such as, R25 awardees, BD2K Centers of Excellence, NIH P41 BTRRs by serving their training and educational material through BBDTC, and acting on feedback at every step.

Acknowledgments

This work is supported by NIH R25 GM114821-01 for BBDTC and P41 GM103426 for NBCR.

References

1. Bowen, WG., Chingos, MM., Lack, KA., Nygren, TI. Interactive Learning Online at Public Universities: Evidence from Randomized Trials. Ithaka S+R; 2012.

- Chapman, B., Calandra, H., Crivelli, S., Dongarra, J., Hittenger, J., Lathrop, S., et al. DOE Advanced Scientific Advisory Committee (ASCAC): Workforce Subcommittee Letter. Washington, DC: USDOE Office of Science; 2014.
- 3. Cole MT, Shelleu DJ, Swartz LB. Online Instruction, E-Learning, and Student Satisfaction: A Three Year Study. The International Review of Research in Open and Distance Learning. 2014; 15(6): 111–131.
- 4. Duderstadt JJ, Colloquium IG. The Impact of Technology on Discovery and Learning in Research Universities. Ninth Glion Colloquium. 2013 Jun.:5–9.
- 5. Hund L, Getrich C. A Pilot Study of Short Computing Video Tutorials in a Graduate Public Health Biostatistics Course. Journal of Statistics Education. 2015; 23(2)
- Joomla HD Video Sharing Plugin. 2016 Sep. Retrieved from Apptha. (n.d.).: https:// www.apptha.com/category/extension/Joomla/HD-Video-Share
- 7. Martin, J. noVNC project website. 2011. Retrieved from https://kanaka.github.io/noVNC/
- McLennan, Ma. HUBzero: a platform for dissemination and collaboration in computational science and engineering. Computing in Science & Engineering. 2010:48–53.
- Purawat S, Cowart C, Amaro RE, Altintas I. Biomedical Big Data Training Collaborative (BBDTC): An Effort to Bridge the Talent Gap in Biomedical Science and Research. Procedia Computer Science. 2016
- Ryabinin, K., Chuprina, S. Using Scientific Visualization Tools to Bridge the Talent Gap. ICCS 2015 International Conference On Computational Science; Amsterdam: Elsevier; 2015. p. 1734-1741.
- 11. Williams, La. In support of pair programming in the introductory computer science course. Computer Science Education. 2002:197–212.

Biographies



Shweta Purawat is Data Science Workflow Specialist for the WorDS Center of Excellence at the San Diego Supercomputer Center, UCSD. She develops compute and data intensive workflows and actors for distributed platforms that involve cloud and grid environments. She is excited about Cloud Computing, Data Science Workflows, Scientific Workflow design, Large Scale Data Intensive Computing, and Predictive Analytics. Her interest lies in designing core technology tools and applications, that act as a catalyst for innovation. She holds M. Tech degree from IIT Bombay, India. Prior to joining SDSC, Shweta was a Design Engineer at Intel Corporation.



Charles Cowart is a data scientist at the San Diego Supercomputer Center, UCSD. He currently develops data models, workflows, and other software for the WIFIRE and DISCOSci projects. Charles has extensive experience in data management and software development, including terms as the LAS Systems Lead for OpenTopography.org, Telepresence Systems Lead at NEESit, and Software Architect for NSDL's Persistent Archive Service. Charles was also a developer of SDSC's Storage Resource Broker, and is co-holder of a software copyright from the University of California. He received an MS in Computer Science from California State University, San Marcos.



Rommie E. Amaro is Professor and Shuler Scholar in the Department of Chemistry and Biochemistry at the University of California, San Diego. She received her B.S. (Chemical Engineering) and Ph.D. (Chemistry) from the University of Illinois at Urbana-Champaign. She was an NIH postdoctoral fellow with Andy McCammon (UCSD). Rommie has received the NIH New Innovator Award, the Presidential Early Career Award for Scientists and Engineers, the ACS COMP OpenEye Outstanding Junior Faculty Award, and the ACS Kavli Foundation Emerging Leader in Chemistry Award. She directs the NIH National Biomedical Computation Resource and co-directs the NIH Drug Design Data Resource.



Ilkay Altintas is the Chief Data Science Officer at the San Diego Supercomputer Center, UCSD. She is the founder and director for the Workflows for Data Science Center of Excellence. She has worked on different aspects of dataflow-based computing and workflows as a PI and in leadership roles across a wide range of NSF, DOE, NIH, and Moore Foundation projects. Ilkay is a co-initiator of the open-source Kepler Scientific Workflow System. Ilkay is the inaugural recipient of the SDSC Pi Person of the Year (2014) and the IEEE TCSC Award for Excellence in Scalable Computing for Early Career Researchers (2015).

Highlights

- **1.** A framework for open knowledge dissemination in the biomedical community is presented
- 2. The BBDTC enables course content personalization through the playlist feature
- **3.** It supports cross-platform operability to reduce effort duplication and encourage innovation
- **4.** The BBDTC exploits virtualization technologies to enable smooth user experience
- **5.** The framework tracks content consumption and usability through usage statistics and user scenarios





Purawat et al.

Ruild Outline		~	Done B Deleted Assets	BBDTC August	8 Day Schension	Vorkflows		UCS	an Diezo
V Day 1, Monday August 17: Introduction to NBCR image analysis and segmentation tools. Beginnenting large electron microscopic image volumes. ESESIONS			Programmable and Reproducible Scalability - Real-Time Hazards Management			- (
Session 1-1: Introduction to 3D EM and Methods for Image Segmentation and Analysis e Image Is Lacture, Sildes, 1-1 e Image Is Video recording for the Session 1-1		Published 📝	Drag files here to upload or	witre.ucsd.ed	WOrD	S	bioKepler.org	NE	3CR
Hands on 1-1: The Minut and Astimated Segmentation of Organeties In EM Data • • • Institution: matchines, buy regards, box • •		Published ? Published ? Published ? Published ?	Drag files here to upload or 0 (2) (2) (2)	WIFIRE	Access and query d Scale computationa Increase reuse Save time, energy a Formalize and stan	ata al analysis and money dardize	bio Keple		
Conversing Conceptioning Conception Conception Sension 1-2: Conversing Conception Conception Sension 1-2: Conversing Conception 1-2: Conception Conception		Published 🖉	Drag files here to upload	kepler-project.org	able Automated Molecular D	ynamics and Drug Disco	worDS.sdsc.edu		<u>Å.</u>
Quiz		=	\$ Name			Unit Pr	ogress		0
			Student 1	-0		-0-	-0-		-0-
NBCR UC San Die	ego	.at 4	Student 2	-6			-0	-0-	-0-
NBCR Summer Training 2015: Modeling			Student 3	-0			-0	-0-	-0-
and Computing using Kepler Workflows	B + C		Student 4	-C	—O—	-0-	-0	-0-	-0-
Quiz	Student 1	Quiz 1 80.00	Student 5	-0		-0-	-0	-0-	-0-
Q1) Which of the following is TRUE about a scientific workflow: can be run on a distributed platform can be used to perform data-driven experiments that can be is a series of connected actions through which data flows i all of the above	Student 2 Student 3	100.00	Student 6	-6	——————————————————————————————————————	-0-	-0	-0-	-0-
	Student 4 Student 6	100.00	Student 7	-6)	-0-	-0		-0-
Q2) Kepler is	Student 6	80.00	Student 8	-0		-0-	-0	-0-	-0-
 Just a rapid prototyping tool an execution engine but cannot run on clusters 	Student 7	100.00	Student 9	-0				-0-	-0-
 an execution engine that runs only on your laptop a reproducible execution engine that lets you run programs on multiple execution platforms including Local and Distributed 	Student 9	100.00	Student 10	-6	<u> </u>	_0_	_0_	- <u>Õ</u> -	<u> </u>
Platforms.	Student 10	80.00				\sim	<u> </u>	9	\sim

Figure 2. BBDTC User Interface Screenshots

Categories	Courses	Datasets	Documents
	BBDTC Course is an online interactive class. The platform offers several features to create a course through a simple and the several set of several severa	Spreadsheets and other data related lists that are available for download but don't fi into other categories.	Articles, technical reports, theses, and t other documents, usually in PDF or DOC format.
	intuitive interface.	Browse >	Browse
	Browse		
	Playlists	🖵 Seminars	🖵 Teaching Materials
	Create playlists by mixing and matching resources, course modules, and additional contents to suit your personal learning requirements.	A lecture of some sort, usually recorded with voice or video. It may be a graduate o undergraduate level seminar, a lecture for class, or a tutorial presentation	Supplementary materials (study notes, guides, etc.) that don't quite fit into any o a the other categories.
	Browse	Browse	Browse
	🕸 Tools	Im Workshops	
	Ready-to-go Virtual Machines with a standardized software tool environment with the required software dependencies and the necessary test data. Browse :	BBDTC Workshop Courses are collections of lectures, seminars, and materials that were presented at a workshop. Browse :	

Figure 3. Resource categories in the BBDTC

J Comput Sci. Author manuscript; available in PMC 2018 May 01.

Author Manuscript

My Playl	list		My Char	nnel My Playlist My	/ Videos Logout + Add Playlist
Playlist Image	Playlist Name	Description	Total Videos Count	Actions	
Hard - Incomparing the state of the second s	Playlist1	Videos from Data to Structural Training Program	3	Play View D	elete Edit
	Playlist2	Videos from NBCR Summer Training Program 2016	4	Play View D	elete Edit
Torona and a second sec	Playlist3	Videos on Introduction to Geometry Modeling and Meshing with HexBlender & Continuity 6 3	1	Play View D	elete Edit

Figure 4. My Playlist Interface

J Comput Sci. Author manuscript; available in PMC 2018 May 01.

Author Manuscript

Video Url*		
Title*		
Description		
		"
Tags		Separate tags by comma
Select Category	Playlists, Seminar, Training, Welcome, Workshops,	
Category*		Reset Category
Privoov	Public Private	

Figure 5. Add Video



Figure 6. Popular and Featured Videos





Amazon EC2 VM Toolbox Integration Architecture



Figure 8. Importing exported edX Studio data



Figure 9. User Scenario Flowchart



Figure 10.

NBCR Summer Training Program - "Data to Structural Models"





NBCR & TCBG Joint Training Program - "Simulation Based Drug Discovery"

Table 1

User Scenarios

User Type	Scenarios
Student/Learner:	Alex is an undergraduate student. He has been directed to register with BBDTC, and enroll in the introductory course on using R, as part of a course he is enrolled in. Alex will dutifully follow the course to its completion, but will not exceed the boundaries of what is required of him. Alex will make ample use of social media features associated with his online course, such as discussion groups.
Instructor:	Aditya is an educator who wishes to create a course to address a specific need in the community, and something they want to teach. He creates an account on BBDTC and requests permission to create a new course. His request and his credentials are reviewed and accepted. A new course is created for Aditya, and he constructs the outline for the online course in 2–3 hours using BBDTC's Course Builder feature. During this time, Aditya is uploading pre-recorded lectures, as well as supplementary course materials; these materials annotate the course outline, and give it substance. Along with a course outline, Aditya elects to set a specific begin and end date for the course. He makes plans to teach the course during the fall quarter. Aditya will be available to manage the course, answer questions, take part in online discussions, and track the overall performance of every student.
Teaching Assistant:	Sungha is a graduate student working with Aditya. She registers with BBDTC and identifies herself as an additional instructor for Aditya's course. Sungha is granted Instructor access for Aditya's course and takes part during the course's term by creating and uploading quizzes, and answering questions in the course's online discussions board.
Freelance Learner:	John is a lifelong learner. He is curious and decides to visit BBDTC after reading about it from an online article. He registers himself as a user and enrolls in the 'Scalable Bioinformatics Bootcamp' course. John creates a profile for himself, where he describes his background and interests to other users. As he becomes aware of playlist feature, John decides to create playlists of videos to suit his individual learning requirements.