

Distributome Annual Progress Report

(Year 3, October 2012 – September 2013)

Table of Contents

A. Background.....	2
B. Summary of Progress	2
C. Infrastructure.....	3
D. Computational Tools.....	3
E. Instructional Modules.....	4
F. Education & Training	5
G. Dissemination & Professional Development.....	5
H. Publications.....	6
I. Science Advisory Board.....	7
J. Personnel	7

A. Background

The Probability Distributome is an innovative web-based project utilizing open-source, open content-development and crowdsourcing for developing new Cloud infrastructure for exploring, discovering, navigating, learning and utilizing diverse probability distributions. Probability distributions are functions ubiquitously used in science, engineering, and biomedical applications to model, study and analyze complex phenomena and natural processes. The multi-institutional Distributome project is funded by the National Science Foundation (NSF) under the DUE TUES/CCLI initiative as a Type 2 (Expansion) project (grants 1023115, 1022560, and 1022636). The project aims include:

- **Computational infrastructure:** Develop robust, extensible and sustainable community-based infrastructure for representation of probability distribution meta-data. And introduce a rich and useful distribution meta-data resource navigator that includes distributions as nodes, relations as edges and functors as glyphs.
- **Learning materials:** Develop Wiki/Blog-based wrapper activities that provide concept- and data-driven demonstrations of the utilization and application of various distribution properties, modeling approaches and inference techniques. And design hands-on knowledge assessment instruments for evaluation of students' understanding of distribution properties, selection of distribution as models and relations between distributions.
- **Instructional resources:** Introduce continuing education curricula for high-school and college instructors from various STEM disciplines and provide a web-based Wiki/Blog repository for probability distribution materials, applications, projects, datasets, assessment tasks, and web-services.

SUMMARY OF DISTRIBUTOME DEVELOPMENTS IN 2012-2013

RESOURCE TYPE	EXAMPLES OF DISTRIBUTOME RESOURCES
Infrastructure	<ul style="list-style-type: none"> • Introduced the Distributome Game www.distributome.org/V3/DistributomeGame.html • Developed an HTML Distributome Carousel www.distributome.org/tools.html • Introduced Distributome Preferences distributome.org/V3/data/Distributome.xml.Book.pref
Computational Tools	<ul style="list-style-type: none"> • Expanded the collection of (HTML5) Simulators and Experiments • Significantly expanded the Distributome Navigator www.distributome.org/V3
Instructional Modules	<ul style="list-style-type: none"> • Introduced 4 new instructional modules demonstrating the classroom use of the Distributome resources for a wide spectrum of applications • Developed 50+ examples of problems and corresponding probability distribution models www.distributome.org/V3/data/DistributomeGame_ProblemExamples.csv
Training & Dissemination	<ul style="list-style-type: none"> • Validated the Distributome infrastructure in 5 courses at UAH, OSU and UCLA • Organized a Distributome Continuing Education Workshop at 2013 USCOTS meeting • Organized a 2012 CAUSEWeb Webinar entitled <i>Hands-on Distributome Activities for Teaching Probability</i> • <i>Publications:</i> In 2012-2013, we had 3 peer-reviewed scientific publications reporting on our theoretical, educational and applied R&D efforts

Table 1: Summary of the 2012-2013 Distributome developments by category with examples of specific developments.

B. Summary of Progress

During the third year of funding, 2012-2013, we continued our R&D efforts in expanding the Distributome computational infrastructure, the development of new interactive learning resources and the continuing education training for students and instructors. **Table 1** summarizes our achievements by category and illustrates examples of each type of development.

C. Infrastructure

In 2012-2013, we expanded the infrastructure of Distributome project including the web-page (www.Distributome.org), the new Distributome Blog (<http://www.distributome.org/blog>), and the version 3 (V3) Distributome Navigator (<http://www.distributome.org/V3>). **Figure 1** shows snapshots of the latest versions of these 3 core Distributome web-infrastructure resources. They provide access to all probability distribution meta-data, learning materials and activities. Learners may traverse, discover, and learn about specific distributions, the relations between different distributions, and their practical applications. We also significantly expanded the Distributome XML database which contains the meta-data for over 80 probability distributions. This database is made openly available to the entire community (<http://www.distributome.org/V3/Distributome.xml.html>). We are actively working to provide a graphical user interface enabling the editing of the database (in addition to human and machine consumption of these data which is already available). The transition to HTML5/JavaScript of all of the Distributome resources is critical to enable all modern browsers and most hardware devices and operating systems to access the Distributome resources without the need of any software installation or significant system configurations (only JavaScript browser functionality is required).

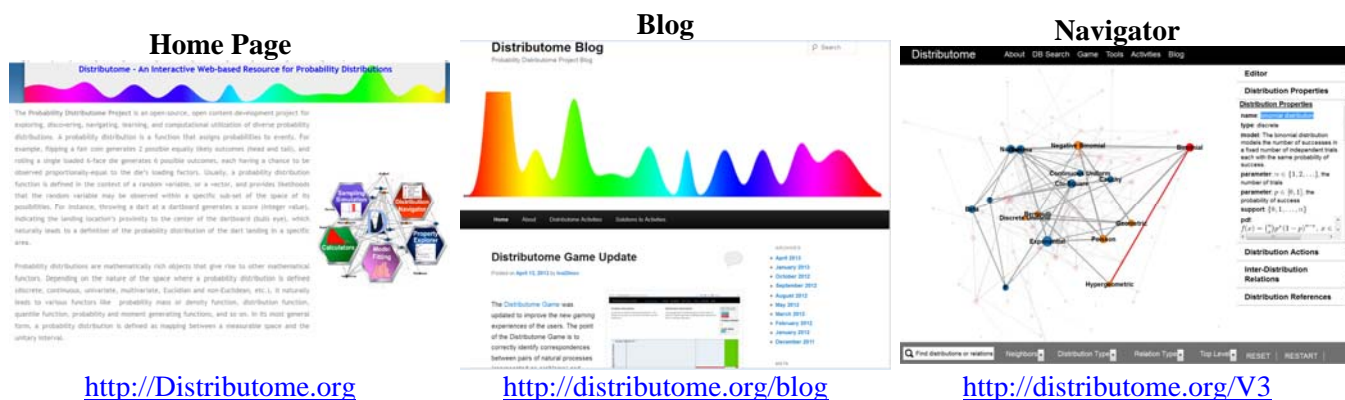


Figure 1: Core web infrastructure of the Distributome project.

The latest version (V3) of the Distributome Navigator includes the Distributome Editor, which provides a user-friendly interface to modify, expand, correct and submit revisions of the core Distributome database. This version uses the new Textbook-based Preference for the hierarchical organization of the levels of distributions. In V.3, all Distributome tools (simulators, calculators, experiments) are linked with the nodes on the Navigator graph. We redesigned our bibliographical references to use the standard math BiBTeX format. All these changes are pushed to Distributome web-site (<http://distributome.org>) and Distributome GoogleCode SVN (https://distributome.googlecode.com/svn/trunk/web_site/V3) and are documented online (<http://distributome.org/blog/>).

D. Computational Tools

We continued to expand the 3 types of computational tool collections that support the needs of instructors and students alike. The first type of Distributome tools are the probability distribution calculators which allow learners and investigators to compute critical and probability values for many distributions included in the database. The second type of tools includes the Distributome interactive simulations. These webapps allow sampling and resampling from each probability distribution. Virtual experiments are the last type of Distributome computational tools and provide examples of applications and use-cases for each distribution in the database. **Figure 2** illustrates the core interface of the Distributome Navigator including the user ability to edit, modify, expand and submit for review new Distributome meta-data.

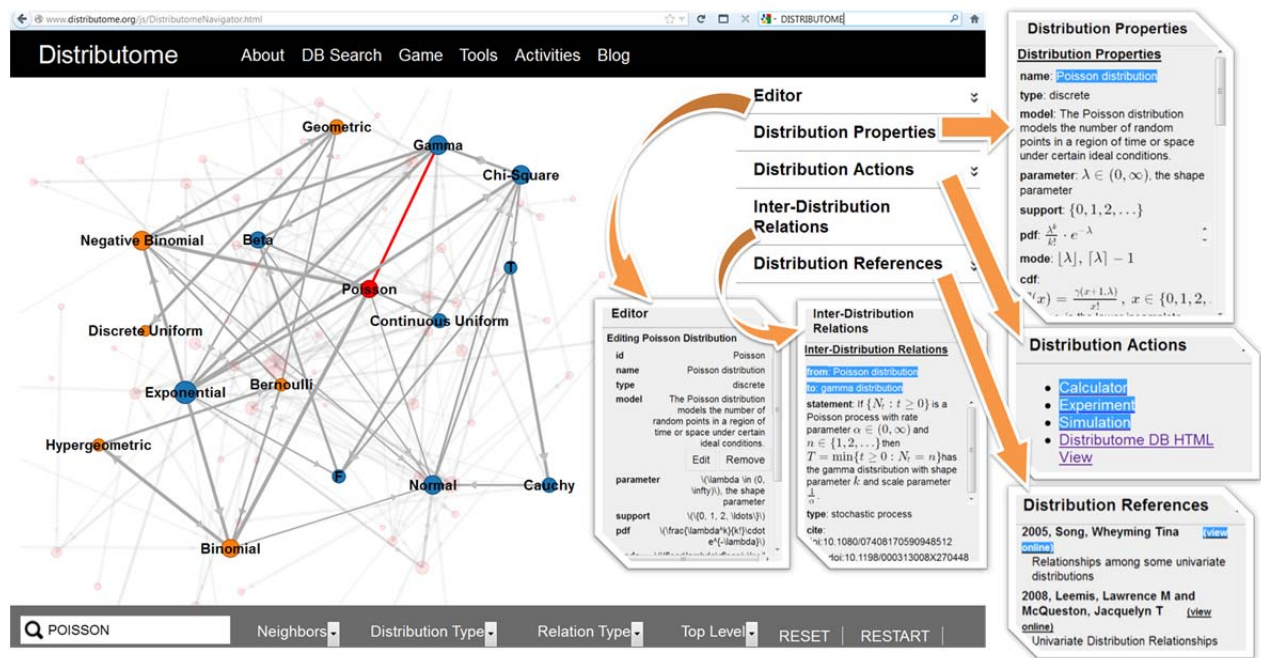


Figure 2: The *Distributome Navigator* provides an interactive web-based interface for traversal, search and exploration of the properties of distributions, as nodes, and their interrelations, as edges in the graph. The Navigator graphical interface is mobile device compatible, software platform agnostic and runs directly in the browser. User can keyword search for distributions, properties or relations, or navigate the graph with the mouse. The top-right corner accordion panels may be expanded to show or edit the appropriate meta data (distribution properties, invoke distribution actions, inter-distributional relations, and scientific publications).

E. Instructional Modules

In the past year, we designed several instructional modules which represent core concepts, methods and applications discussed in undergraduate probability, statistics and science education. Each of these materials includes specific examples relating applied problems and appropriate probability distributions. These modules are integrated with the Distributome computational tools and are intended to be easily foldable into concrete curricular settings. All of the Distributome learning activities and instructional resources are accessible on the web (<http://distributome.org/blog>). **Figure 3** depicts the new Distributome webapp Game which presents learners with a game-like interface asking them to quickly identify correspondences between pairs of natural processes (represented as problems) and probability distributions (as models).

COURSES USING THE DISTRIBUTOME INFRASTRUCTURE IN 2012-2013		
COURSE	INSTITUTION	COURSE TITLE
Stats 13	UCLA	Statistical Methods for the Life and Health Sciences
Stats 199	UCLA	Research in Probability Modeling and Statistical Computing
MA 385	UAH	Introduction to Probability
ST 487	UAH	Introduction to Mathematical Statistics
Stat 4201	OSU	Introduction to Mathematical Statistics I

Table 2: Undergraduate courses utilizing the Distributome resources in 2012-2013.

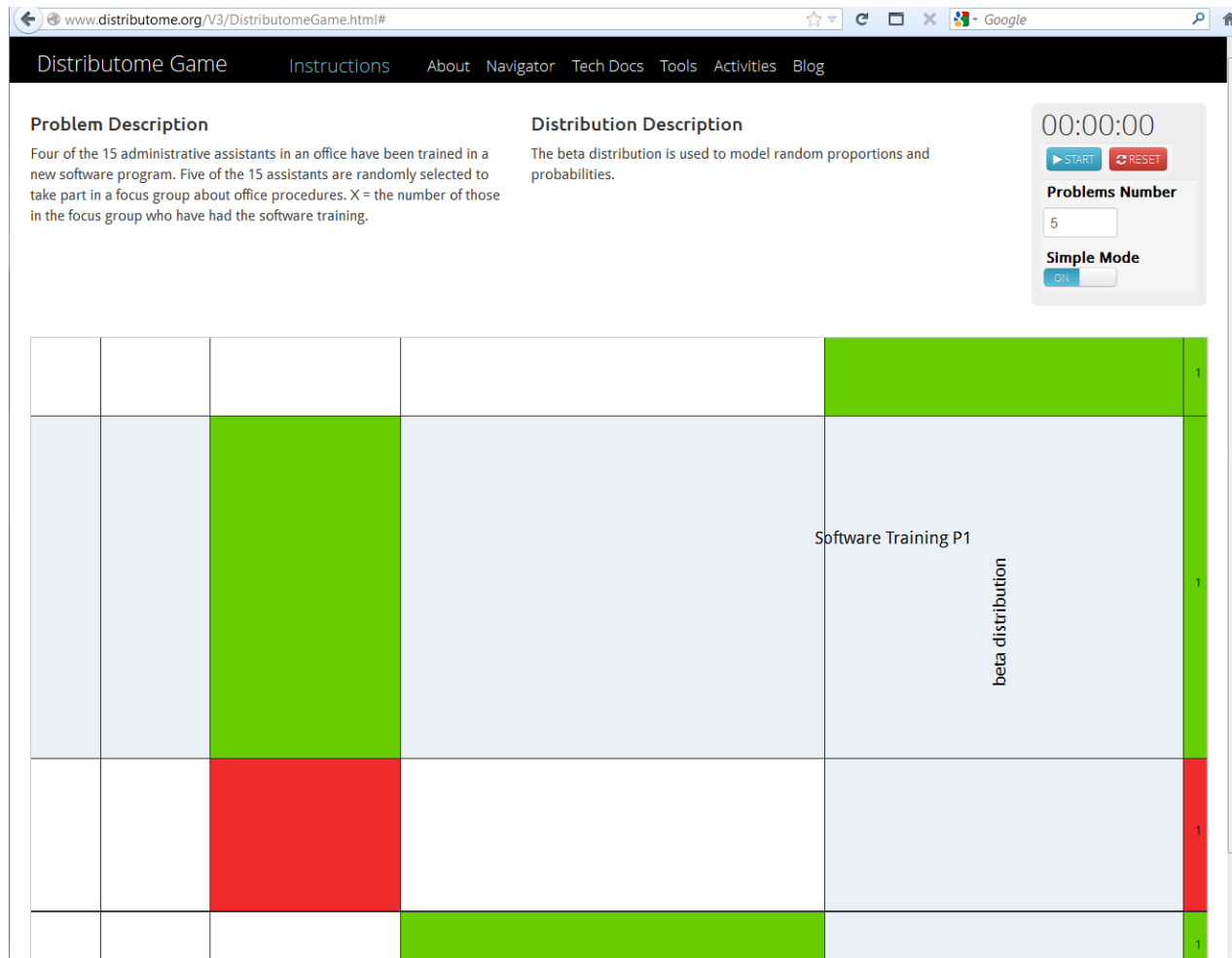


Figure 3: The *Distributome Game* is a game-interface where players aim to quickly identify the correspondences between pairs of processes (represented as problems) and probability distributions (as models). The Cartesian plane represents the game-board where rows and columns show problems/processes and distribution models, respectively. Correct and incorrect matches are green and red colored. Various optional hints and help mechanisms are provided for the players. Green and red cells indicate correct and incorrect pairing of the problems and model distributions, respectively. The last column indicated the number of guesses for each problem (<http://www.distributome.org/V3/DistributomeGame.html>).

F. Education & Training

The Distributome project resources continue to be tested at the participating institutions. The courses taught by the PIs monitor student interactions and experiences with using the Distributome resources and collect feedback that is used to (re)design, expand and improve the Distributome activities and computational infrastructure. **Table 2** contains a list of (known) courses formally utilizing the Distributome infrastructure and we anticipate extending this list in the coming months. In addition, we are training students in the development and validation of such interactive computational resources.

G. Dissemination & Professional Development

In 2012-2013, we continued our training and dissemination efforts and presented the Distributome resources at the following venues:

- May 15-16, 2013: Organized a continuing education workshop entitled “*Interactive Probability Instruction Workshop*”, as a satellite meeting of the 2013 USCOTS, Durham, NC (http://www.distributome.org/meetings/USCOTS_2013). Over 20 instructors and students from 2-year Colleges and 4-year Universities attended the workshop. Post-workshop evaluations were both praising and constructively encouraging. Some summary graphs of these (anonymous) evaluations we shown on **Figure 4**.

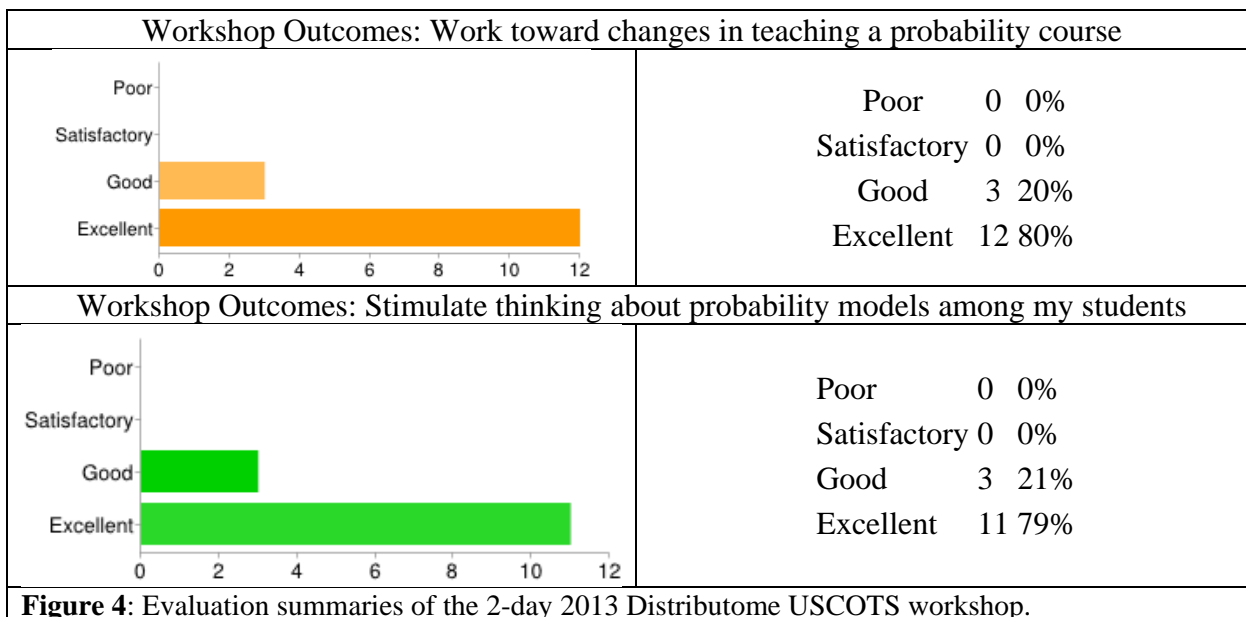


Figure 4: Evaluation summaries of the 2-day 2013 Distributome USCOTS workshop.

- November 27, 2012: CAUSEWeb Webinar entitled Hands-on Distributome Activities for Teaching Probability (http://www.distributome.org/meetings/CAUSEWeb_Webinar2012). Forty-nine attendees participated in this webinar, provided valuable feedback about the available Distributome resources, and made constructive recommendations for enhancements and improvements to increase students’ motivation and their learning experiences. Examples of open-ended comments include:
 - *"Your webinars are very efficient and worth every minute. I learn something from each one. The technology works well for me, and the information is always about the real world. Thank you."*
 - *"Website looks like a fantastic resource. Looking forward to exploring it in greater depth."*
 - *"[A] fine work of software. It is a great resource. I am using Distributome in a graduate course in Biostatistics (in Spanish). My students enjoy Distributome in spite of their limited English skills"*

H. Publications

Table 3 contains a summary of the 2012-2013 peer-reviewed Distributome project publications. These publications reflect our R&D efforts in enhancing probability and statistics educations, improving the open-scientific computing infrastructure and conducting applied biomedical research studies using the Distributome resources.

2012-2013 DISTRIBUTOME PROJECT PEER-REVIEWED PUBLICATIONS	
REFERENCE	URL
Dinov, ID, Kamino, S, Bhakhrani, B, Christou, N. (2013) Technology-enhanced Interactive Teaching of Marginal, Joint and Conditional Probabilities: The Special Case of Bivariate Normal Distribution, Teaching Statistics , <i>in press</i> , DOI: 10.1111/test.12012.	http://onlinelibrary.wiley.com/doi/10.1111/test.12012/abstract
Lam, H-C, Dinov, ID. (2012) Hyperbolic Wheel: A Novel Hyperbolic Space Graph Viewer for Hierarchical Information Content, ISRN Computer Graphics , Volume 2012, Article ID 609234, doi:10.5402/2012/609234.	http://www.hindawi.com/archive/2012/609234/
Torri, F., Dinov, ID, Zamanyan, A, Hobel, S, Genco, A, Petrosyan, P, Clark, AP, Liu, Z, Eggert, P, Pierce, J, Knowles, JA, Ames, J, Kesselman, C, Toga, AW, Potkin, SG, Vawter, MP, Macciardi, F. (2012) Next Generation Sequence Analysis and Computational Genomics Using Graphical Pipeline Workflows, Genes , 3(3):545-575; doi:10.3390/genes3030545.	http://www.mdpi.com/2073-4425/3/3/545

Table 3: Peer-reviewed Distributome project publications (2012-2013).

I. Science Advisory Board

The Probability Distributome Project is guided by a Science Advisory Board including expert educators and scientists who review the project and provide directive to maximum Distributome impact. The SAB includes Dr. Lawrence Moore (Duke), Dr. David Aldous (Berkeley) and Dr. Robert Dobrow (Carleton College). The board regularly reviews the Distributome resources, interacts with the PIs, provides critical feedback on the annual progress report, and directs the core developments and vision for the following year. The next Distributome SAB meeting is in Fall 2013.

J. Personnel

In 2012-2013, there were 2 types of personnel engaged in various Distributome projects at the 3 main sites – University of Alabama, Huntsville (UAH), Ohio State University (OSU) and University of California, Los Angeles (UCLA) – each with different roles and complementary responsibilities – *students* and *faculty*. This list excludes collaborators and educators that have directly or indirectly contributed to the development of Distributome materials and resources. Due to some turn-over in student researchers, unforeseen challenges in managing the *open-science* community of Distributome project users, outside contributors and feature requests, we were not able to complete all the project goals set forth in the original application. Thus, we request a one year no-cost time extension (NCTE) to complete the project aims by September 29, 2014.

- Students: **Table 4** describes the projects of all students engaged in the Distributome project.

DISTRIBUTOME STUDENT TRAINING			
TYPE	INSTITUTION	NAME	PROJECT
Undergraduate	UCLA	Selvam Palanimalai	HTML5 web interface
		Alexandr Kalinin	Distributome Game
		Prarthana Alevoor	Distributome database and web infrastructure
		Bilal Bhakhrani	JavaScript interfaces
Graduate	UAH	Yunzhu He	XML, JavaScript, MathJax coding

Table 4: Distributome student researchers (2012-2013).

- Distributome Faculty/PIs: The **3** Distributome project PIs, Kyle **Siegrist**, Dennis **Pearl** and Ivo **Dinov**, continued to actively participate in managing all Distributome project activities within each site, conducting teaching and dissemination activities, and coordinating the interactions and developments across the participating institutions. The PIs continued to hold monthly interactive conference calls and exchange emailed on a regular basis to ensure the progress on project is according to the aims and timeline.
 - Dr. **Siegrist** led the development of the Distributome HTML5 redesign, the expansion of the Distributome XML database, contributed to the design of instructional activities, and steered the project dissemination and professional development efforts.
 - Dr. **Pearl** led the developments of the instructional resources and learning activities, introduced the new Distributome Game infrastructure, and coordinated the Distributome dissemination events.
 - Dr. **Dinov** led the development of the Distributome HTML5 navigator and editor, contributed to the design of instructional resources and learning activities, and managed the Distributome web-site.