# CO-DESIGN OF THE COMMON ONLINE DATA ANALYSIS PLATFORM (CODAP) FOR CROSS-DISCIPLINARY USE AT GRADES 6–14

William Finzer, Senior Scientist and Daniel Damelin, Senior Scientist
Concord Consortium, Concord, MA, USA
wfinzer@concord.org

*The rising importance of data science and the need for data science education at the pre-college level has led to the need for a new generation of data analysis and visualization software appropriate for use in school classrooms. The Common Online Data Analysis Platform (CODAP) represents a step toward answering that need as it evolves through an ongoing co-design process to be a tool for learning and engaging in data science at the school level. Data science education research conducted as part of the development of the software, sheds light on how learners make sense of hierarchical data structures represented in CODAP's tables and graphs, especially as these representations make use of dynamic linking of data selection.*

## ORIGINS

In 1998 one of us (Finzer), along with Tim Erickson of Eeps Media, led the team that developed *Fathom Dynamic Data*, published that year by Key Curriculum Press, also then the publisher of the dynamic geometry software *The Geometer's Sketchpad*. *Fathom* and *Sketchpad* brought the excitement of direct, dynamic manipulation of objects on the screen such that whether you were dragging the vertex of a triangle or a data point in a scatterplot, all dependents of the dragged object changed simultaneously, smoothly, in real time. We felt, and research showed (Goldenberg, 1998), that these computer learning environments brought about important qualitative improvements in the ways students understood constrained systems such as geometric constructions and statistical models.

A decade later, the National Science Foundation funded as a collaborative grant for Finzer, Cliff Konold, and Tim Erickson—"Data Games: Tools and Materials for Learning Data Modeling." The working hypothesis of *Data Games* was that by embedding very simple games in a data analysis environment learners could make use of data they generated during game play to reverse engineer the game and improve their strategies and scores. A critical decision we faced early on was whether to embed the games in *Fathom* and *TinkerPlots*, or design a new more accessible browser-based derivative of them. The latter choice, to design a web application, has led to the Common Online Data Analysis Platform (CODAP, 2018), subsequently funded through its own NSF grant. It is free and open source.

## DESIGN GOALS

### *Adaptable for use in all STEM and social science subject areas*

By the time the CODAP project started in 2014, educators were faced with the rise of data science and the need to figure out how to incorporate data in learning experiences across the curriculum. We were fond of pointing out that data are everywhere, except in the classroom. Increased emphasis on statistical concepts and practices in the Common Core Standards for Mathematics (CCSM, 2010), plus increased attention to working with and learning from data in the Next Generation Science Standards (NGSS, 2013) strengthened our resolve to develop software useful in a broad range of subject areas.

### *An open source platform on which other projects can build*

Many, if not most, developers of STEM online curricular materials wish to incorporate student engagement with data into their learning activities. But meaningful engagement with data is greatly facilitated by appropriate technology, and, at least at the school level, such technology is best if designed for learners rather than practitioners. We were of the opinion that the needed software did not yet exist, that its development was beyond the means of most projects, and that advances in open source software provided the means by which the elusive goal of collaborative educational software development might be met.

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org 

Accordingly we set out to create a *platform* rather than a single-purpose tool or learning environment. This meant that CODAP should be easily adaptable to a broad range of curriculum development settings and modes of use—as an environment for data-generating plugins, as a separate tool that can import data from spreadsheets and databases, as components embedded in web pages, as a library for web applications, or as a vehicle for interacting with large, online datasets.

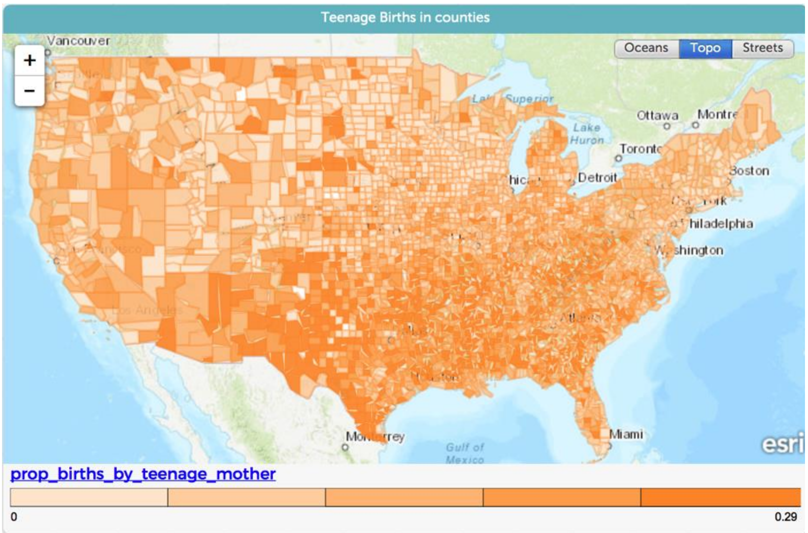


*Figure 1. Co-design with the Terra Populus project led us to boundary maps in CODAP. In this example map, the regions are counties, and the shading corresponds to the proportion of births in that county that were born to teenage mothers.*

*A tool for learning and doing data science at the school level*

As a new field, data science currently lacks an agreed-upon definition. At the college level and beyond courses, majors, and departments are being formed. Least understood is how learners at the pre-college level should engage with data science skills, become conversant in its concepts, and experience its practice. But engage they must—to be better prepared to meet the needs of our increasingly data-driven society. We are convinced that doing data science at the school level involves working with data in ways that lie outside current school practice. It will need to include large data sets, data structures other than flat row-by-column tables, data moves such as grouping and restructuring, and a kind of immersion in data that rarely happens traditionally.

## CO-DESIGN

We understood that work in isolation would be unlikely to yield something that came close to meeting these ambitious design goals. Accordingly, we built into the project close collaboration with three teams from disparate settings: The OceanTracks project at EDC (http://oceantracks.org), the Terra Populus project at the Minnesota Population Center (https://www.terrapop.org), and the InquirySpace project at the Concord Consortium (https://concord.org/our-work/research-projects/inquiryspace/). The result was three strands of co-design, each with its own evolving take on the requirements for CODAP.

What constitutes co-design? For us it involves collaboration between developers, designers, and stakeholders in which, as a group, we look at the evolving ideas and prototypes that are becoming the "product" that students and teachers will use in classrooms. In our case, the CODAP software is an essential element of the product and our central concern, but other materials such as lesson plans, teacher guides, reference materials, and professional development needs easily and appropriately become part of the conversation. Here are some of the critical findings from this process:

- The ability for students to work with geographical (GIS) data in addition to numeric, categorical, and informational data is a necessary part of data-driven investigation in many science and social science settings. Co-design with OceanTracks and Terra Populus led us to provide display of point location data (see Figures 2 and 3) and area data (states, counties, etc.) on maps. (See Figure 1.)
- CODAP's dynamic linking of selection between GIS displays, graphs, and tables (see Figure 3) is not available in most other tools. Having this capability brought about exciting discoveries

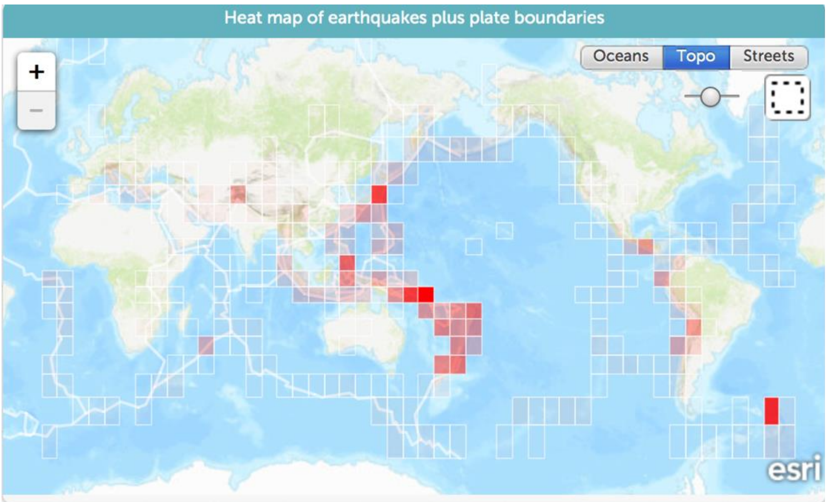


*Figure 2. Four thousand earthquakes from April, 2016 to April, 2017 are plotted as a heat map. This capability came about through co-design with the OceanTracks project.*

during co-design sessions that led us to believe that students would also be highly motivated as they worked with these capabilities.
- The need for integration of data sources with data analysis is a recurring theme, and, for us, proved much more important than capabilities for student data entry from scratch. (In fact, the project went on for a full two years *without* simple data entry and, in the end, it was

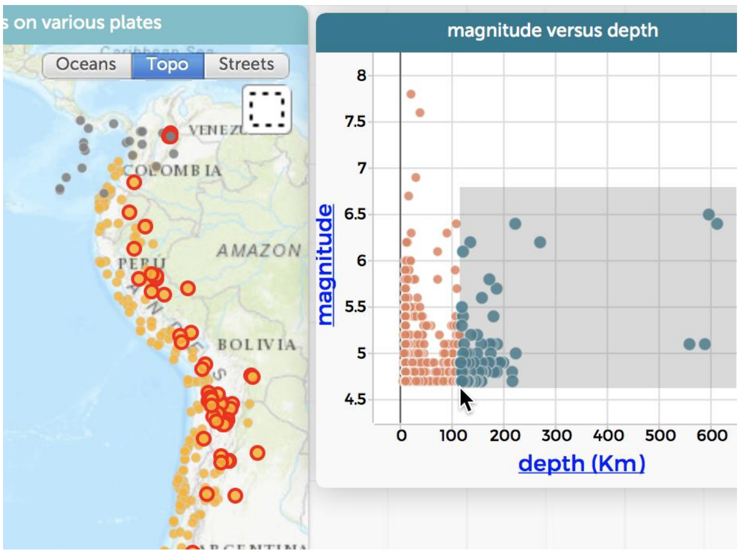


*Figure 3. As the user drag-selects the deepest earthquakes in the graph, the earthquakes plotted on the maps also dynamically select showing the connection between depth and geography.*

implemented at the insistence of the development team rather than to fulfill any desire of collaborators.) All three collaborating projects had data of their own that formed the basis of learning activities.
  - InquirySpace used plugins that streamed data into CODAP from external sensors or integrated simulations. (See Figure 4.)
  - OceanTracks built a plugin with which students could choose GPS tracking data from 65 animals of four different species.
  - Terra Populus devised a way for users to select data using their interactive TerraScope tool and then download CODAP documents for further exploration.
- None of the three collaborating projects had a need for more than the simplest statistical measures. Counts, the mean, and the median were sufficient (see Figure 5). These projects were not looking to involve students with formal inference.

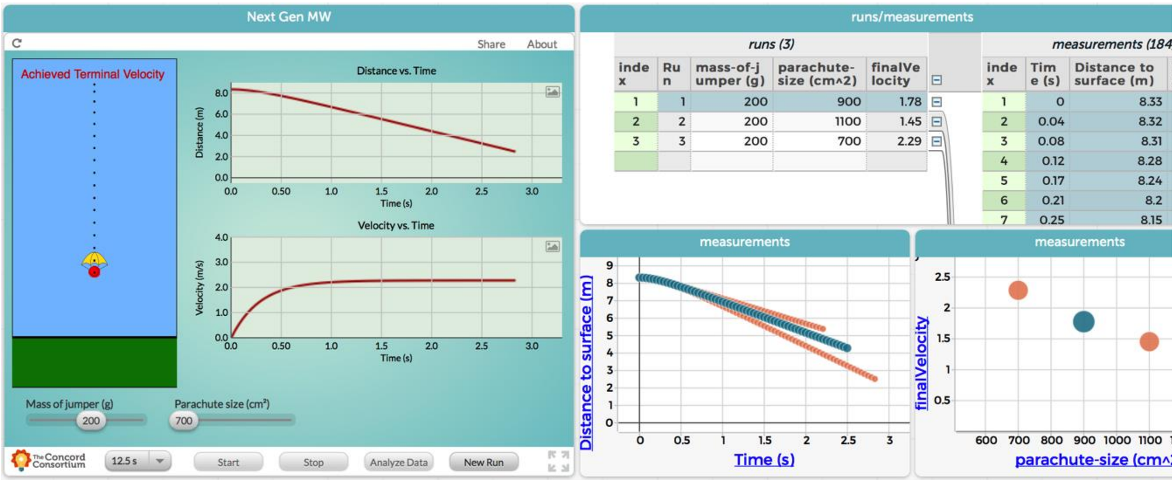


*Figure 4. The InquirySpace project used plugins to bring sensor and simulation data into CODAP. Here data are coming from a simulation of dropping parachutes of various mass and size.*

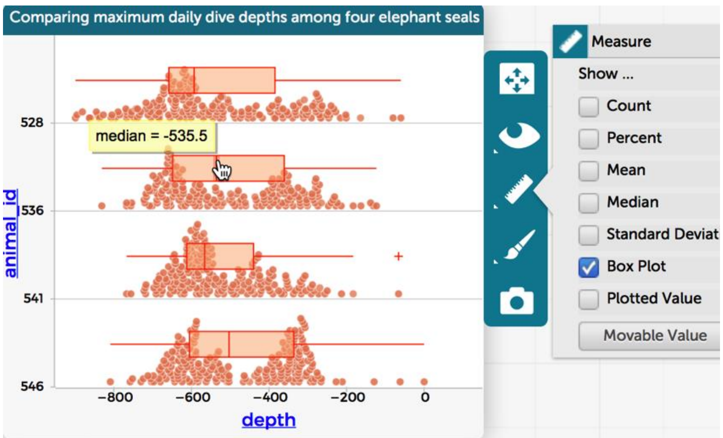


*Figure 5. Statistical modeling needs of science educator collaborators were rather limited.*

- The importance of teachers being able to distribute CODAP documents to students became apparent early. This led to an innovative sharing mechanism by which a master document could be shared through a URL that would open up a *copy* of the master such that modifications to the copy would not alter the master.
- In student investigations in the InquirySpace project a natural hierarchy emerges: Each *run* has a set of *measurements* (see Figure 6). Similarly, in the OceanTracks datasets, each *species* has *tracks* each of which has a set of *measurements*. In co-design with our collaborators we

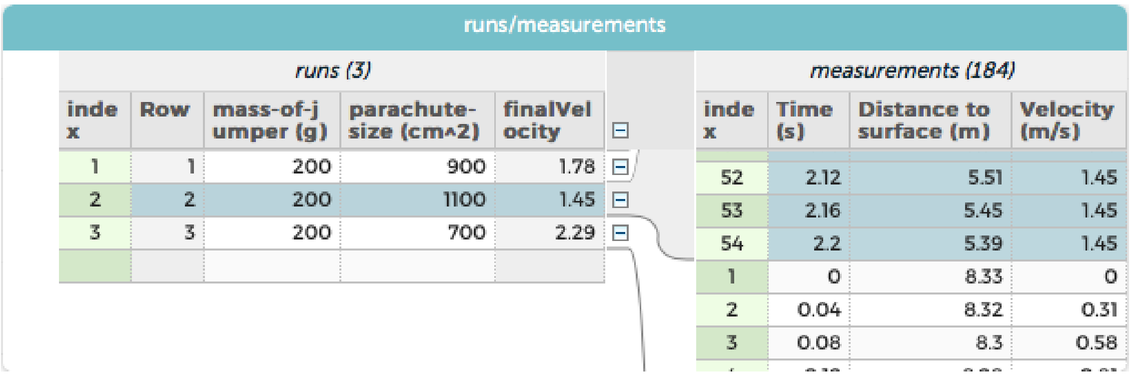


*Figure 6. In an InquirySpace experiment, data are organized hierarchically into runs and measurements.*

explored the ramifications of structuring data hierarchically and found that the case table interface for displaying and interacting with hierarchical structures made sense to users and might even be easier for students to understand than traditional flat, row by column data sets.

## DATA SCIENCE EDUCATION RESEARCH

Data are everywhere, underlying nearly everything we do, knitting the world together, revealing patterns that lead to insightful solutions to difficult problems. Data science has emerged as a new field whose mission is to help us work with data, and data science education is an even newer field focused on bringing about meaningful learning of data science concepts, skills, and practices. Finally, data science education research attempts to reveal how such learning takes place.

### *How Learners Experience Hierarchical Data Structures*

There is so much we don't know about how learners think about data. What is their mental picture of data? Are there structures learners imagine in which data are contained? In learners' minds what kinds of data are there? What connection is there between data and the technology used

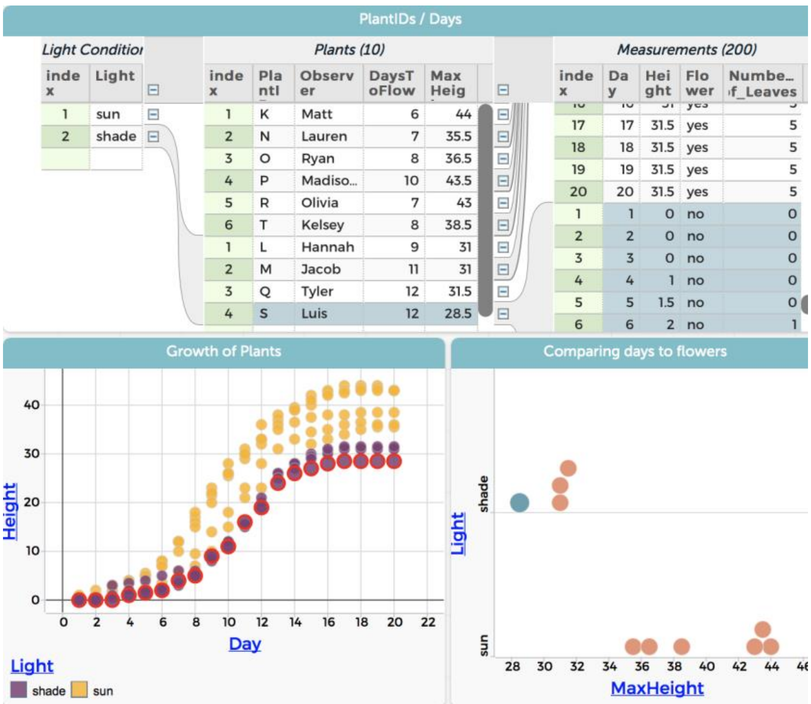


*Figure 7. The 3-level hierarchy used in interviews with middle school students.*

to manipulate data? Most of all, for our purposes, how do learners interact with data hierarchies such as the ones visualized in the tables shown above?

From prior research (Konold, 2017) we know that many learners, as young as 13, when given an opportunity to decide how to set up a way to record data, will spontaneously choose nested or hierarchical structures for the purpose. In the work with CODAP the research team wanted to find out what would happen when middle school students were asked to make sense of existing hierarchical structures as presented to them using CODAP. Researchers conducted semi-structured cognitive interviews with thirty students in grades 6–8 in which they were presented with a CODAP document similar to the one in Figure 8. We were interested in how they make sense of this three-level hierarchy, and what stages they go through as they seek to describe the meaning of points in the graphs and rows in the table.

Somewhat surprisingly, all students were able, in the course of a one-hour interview, to make sense of the data structure with minimal guidance from the interviewer aside from demonstrating how to make graphs and, if not discovered by the student, instructing the student to click on a point or row. Some students initially seemed to think that each point in the plot of Height versus Day represented a different plant, but this confusion cleared up as students discovered multiple linked representations could be used to clarify data structures. As shown in Figure 7, clicking on the row for plant S causes all the measurements for that plant to become selected and for the point representing that plant in the Light versus MaxHeight graph to highlight.

## SUMMARY

One obstacle to effective integration of data-rich learning across multiple subjects at the school level is the lack of appropriate technology for use by students and teachers. Underlying that obstacle lies a dearth of understanding of how learners come to understand how data are structured and the way they can be used in modeling real world situations. As an evolving, browser-based platform designed to support the development of data-driven STEM curricula, CODAP demonstrates that co-design with collaborators from disparate subject areas can lead to capable, flexible, intuitive software with which students can experience data-oriented explorations that expand disciplinary knowledge as well as learning the rudiments of data science.

## ACKNOWLEDGMENT

The work described in this paper was supported by the National Science Foundation under NSF Award Numbers IIP-9400091, DRL-0918735, and DRL-1316728. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect those of the National Science Foundation.

## REFERENCES

*CODAP* [Computer software]. (2018). Concord, MA: The Concord Consortium.

Common Core State Standards Initiative. (2010). Common core state standards for mathematics. Retrieved from *http://www. corestandards. org/assets/CCSSI_Math% 20Standards. pdf*

Goldenberg, E. P., & Cuoco, A. A. (1998). What is dynamic geometry. *Designing learning environments for developing understanding of geometry and space,* 351-368.

Haldar, L.C., Wong, N., Konold, C., & Heller, J. (2018). Students making sense of multilevel data. *Technological Innovations in Statistics Education*, in press.

Konold, C., Finzer, W., & Kreetong, K. (2017). Modeling as a core component of structuring data. *Statistics Education Research Journal, 16*(2), 191-212.

National Research Council. (2013). Next generation science standards: For states, by states. Washington, DC: The National Academies Press.