Patterns

CellPress
OPEN ACCESS

People of Data

Data science enables X-ray vision

David A. Santos,[1,2,4,*] Binbin Lin,[3,*] and Justin L. Andrews[1,2,5,*]
[1]Department of Chemistry, Texas A&M University, College Station, TX 77843-3255, USA
[2]Department of Materials Science and Engineering, Texas A&M University, College Station, TX 77843-3255, USA
[3]Institute of Materials Science, Mechanics of Functional Materials, Technische Universität Darmstadt, Otto-Berndt-Str. 3, 64287 Darmstadt, Germany
[4]Twitter: @DA_Santos_
[5]Twitter: @JLAndrews_Chem
*Correspondence: dasanto2@tamu.edu (D.A.S.), b.lin@mfm.tu-darmstadt.de (B.L.), justinandrews.chemistry@gmail.com (J.L.A.)
https://doi.org/10.1016/j.patter.2022.100660

In their recent publication in *Patterns*, the authors proposed a novel workflow to derive compositional and stress maps for positive electrode materials of Li-ion batteries from hyperspectral X-ray imaging data. They describe their interdisciplinary collaboration, the elements that sustain such collaborations, and their effect on the flourishing of the domain and data science.

What would you like to share about your background (personal and/or professional)?

David A. Santos: My interest and trajectory in science likely began during my childhood exploratory "why" phase, which was marked by many deconstructed radios and dangerously failed experiments in an attempt to prototype hydrogen-powered bottle rockets. I remain driven by such curiosity, although over the course of a few advanced degrees in chemistry, my approach to designing experiments has slowly grown more refined. While my training is formally in chemistry, I have never identified exclusively as a chemist. I believe this mindset has allowed me to approach problems with an open mind and a proclivity to combine approaches from disparate fields in search of a solution. Currently, I am a senior research specialist at Dow Chemical Company, working in the polyurethanes segment alongside our customers to deliver solutions that positively impact society. Dow continuously integrates artificial intelligence (AI) and predictive capabilities to accelerate product development, and I am excited about harnessing this ability to enhance innovation and customer collaboration.

Binbin Lin: I am a PhD candidate with a background in mathematics and applied mechanics. I am currently working with Prof. Dr. Bai-Xiang Xu in the mechanics of functional materials group at the Technical University (TU) of Darmstadt, Germany. Over the last several years, I have used computational tools to investigate the interplay between microstructure and structure-function relationships in a few different materials systems. Currently, I am interested in developing new ways of applying data-driven approaches to materials research, specifically by using high-performance computing techniques to tackle large data problems within the framework of NHR4CES, a Germany-wide high-performance computing center for computational engineering science.

Justin L. Andrews: Before pursuing a career in chemistry, I actually wanted to be a linguist (of all things). I loved learning to read Latin and Greek in high school, so it seemed like a natural fit. I chose to attend a liberal arts college so that I could study philosophy, literature, and linguistics in addition to the sciences. During my first year at university, I fell in love with chemistry and the rest is history. My first experience with data science was during a class project in linear algebra during my undergraduate studies. I teamed up with one of my roommates (who was studying "parks, recreation, and leisure studies") to see if we could mine available data to develop a model that would predict which university successful high school athletes would choose to attend. Our project wasn't very successful (as one might expect for a few undergrads with no real experience in data science), but the experience always stuck in my head and cemented in my mind that data science could be a useful tool at my disposal. Currently, I am a post-doctoral researcher at MIT working on the design of materials for energy storage applications. While I am not currently using data science, I am always looking for an appropriate opportunity to get "back in the game" as I move into the next stage of my independent career.

How did you come to collaborate? What is the fun part of your collaboration?

JLA: The collaboration between Dr. Banerjee's group at Texas A&M University (TAMU) and Dr. Xu's group at TU Darmstadt has been ideal. On a practical level, the collaboration has allowed us to meet our scientific goals by developing new analytical methods for probing fundamental chemical processes and has further allowed us to gain new insight into the coupling of chemistry and mechanics through simulation. We couldn't have met many of our most substantial goals independently of each other. From the perspective of productivity, the collaboration has led to several important publications in which we take great pride.[1–3] Our article in *Patterns*[4] is both a celebration of this work and an elevation of our approaches to new levels through the incorporation of data-driven methods. On a more personal level, it has been immensely rewarding to see the collaboration succeed across several "generations" of graduate students, post-doctoral researchers, and staff scientists. I started working on the collaboration with (now) Prof. Luis R. De Jesus, who was a senior graduate student at the time and an excellent mentor. Toward the end of

my time in Sarbajit's group, I had the privilege of mentoring (now) Dr. David Santos as he picked up the mantel for the Banerjee group. He has been incredibly successful in graduate school and is now working at Dow. Before he left, he mentored another student, Luis Carillo, who I am sure will go on to do great things. From Dr. Xu's group, I had the opportunity to start working with (now) Prof. Ying Zhao followed by another talented researcher and good friend in Dr. Peter Stein. Now Binbin Lin has stepped in and brought new techniques and approaches to the collaboration. Each researcher made their own unique mark on the project. Some collaborations are fleeting, and some are built to stand the test of time, despite group turnover. This collaboration is built to stand the test of time!

BL: The collaboration between our groups has always been driven by the marriage of experiment (TAMU, Banerjee Group) and simulation (TU Darmstadt, Xu Group) to understand materials properties and behavior in a new way. What I love most about our meetings is when Dr. Banerjee starts showing us some new experimental observations and asks if we have any theoretical insight or might be able to model the materials. That is really when our work (and the fun!) begins. For us, the ultimate goal is to better understand how we might understand and optimize materials properties. Between those steps, our groups have many interesting and fruitful discussions that constantly improve almost everything we do. It's really a dream collaboration!

Was there a particular element (paper, collaboration, talk/conference, key experiment, idea, result) that motivated you to start/participate in this project?

JLA: The paper that started it all for our groups was Dr. Luis R. De Jesus's work published in *Nature Communications*.[5] In that paper, Luis showed that scanning transmission X-ray microscopy (STXM) could be used as a powerful tool to map chemical information across individual particles of a battery cathode material. The paper was primarily focused on understanding the electronic structure of the cathode material, but we also observed phase-separation patterns within the particles that were really interesting. We started to consider the implications of interfaces and phase boundaries in individual particles and how they might evolve during the normal cycling of a battery and ultimately impact battery function. Sarbajit correctly identified that the questions we were asking about the chemistry were rooted in mechanics and found the best collaborator we could have hoped for in Dr. Bai-Xiang Xu. Under their guidance, Luis, Ying Zhao, Peter Stein, and I started probing the link between mechanics (strain/stress) and phase separation in V_2O_5 materials.

DAS: Borrowing from an adage popularized by Spider-Man in Marvel comics, I would add here that "with big data, comes big responsibility." When I first began processing high-dimensional hyperspectral imaging data, I retained a healthy level of skepticism toward our existing workflows—frequently looking for areas of improvement. I feel incredibly fortunate that my mentor, Justin Andrews, welcomed these discussions with an open mind. Together, we worked through many thought experiments validating several hypotheses through artificial/model and experimental datasets. The first iteration of this paper was finished in 2019! In many ways, our shared commitment to the development of rigorous workflows fueled many of the key insights we now detail in our *Patterns* article.

Who were the driving forces behind the project?

JLA: This project drew inspiration from so many different people across the years. Dr. Luis De Jesus was a pioneer of applying the STXM technique we employed during our study in *Patterns* to the investigation of cathode materials. Dr. Ying Zhao and Dr. Peter Stein together paved the way with some of the code we employed to understand the mechanics of the materials we were investigating through modeling. David and Binbin have really thought deeply about these systems and the role of data science in our analysis and pushed the limits of the analysis beyond what I thought was originally possible. It has been a true team effort. Most importantly, behind these team members are two driven, creative, and supportive principal investigators in Prof. Dr. Bai-Xiang Xu and Prof. Sarbajit Banerjee. They are responsible for sustaining this collaboration over the years and are incredibly supportive of us as young researchers. They constantly put us in position to succeed.

BL: Dr. Xu and Dr. Banerjee and their will to understand things better really drives the collaboration forward. Cheers to it!

When did you start preparing the manuscript?

JLA: As David mentioned, the first iteration of this paper was finished all the way back in 2019. That's what makes this paper so exciting! Back then, the paper was more focused on the implications of different chemical lithiation techniques on inter- and intraparticle phase homogeneity. What we quickly realized was that this presented a unique opportunity for us to go back and look at a lot of our old data in a very systematic and detailed way. When we realized that we could extract new information from our old data using our newly developed workflow, the paper took on a new shape (and a lot more work) on its journey to becoming the manuscript that we eventually submitted to *Patterns*.

Why did you decide to publish in *Patterns*?

JLA: I've come to see *Patterns* as a forum that juxtaposes data science with interesting concepts. In the October issue alone (volume 3, issue 10), I am drawn to an article on AI-inspired decision making titled "Humans versus machines: Who is perceived to decide fairer?" This is followed by articles on the design of nature-inspired neural networks, data science approaches to reducing carbon emissions by improving fleet scheduling, accurate quantification of biodiversity at the earth's poles, and finally a really interesting study on data science insights into the fundamental limits of high T_c superconductors. I recall last year an interesting *Patterns* paper on using machine learning to reduce the time needed to model battery aging. Very cool! What other journal offers such diversity of topics and features that level of interdisciplinarity? *Patterns* is not only an appropriate home for our manuscript, but I think we all view this as an exciting opportunity to share our research with communities beyond chemistry and materials science.

DAS: I could not agree more with Justin—the diversity of subject matter in *Patterns* is remarkable. One of the most

rewarding outcomes I could envision from disseminating this work in such a unique venue is that a researcher from an entirely different discipline might find utility in the data science approach we have developed for this study. I think that prospect is truly exciting, and I can think of no better place to enable its possibility than *Patterns*.

BL: I've been reading articles in *Patterns* for some time now. The articles are multi-disciplinary and have inspired me to explore disparate domains of science. I am always curious to see how authors from other disciplines use machine-learning and data-driven approaches to tackle their field of problems. It is a great way to incorporate new ways of thinking!

What is the definition of data science in your opinion? What is a data scientist? Do you self-identify as one?

JLA: As a chemist, I spend a good part of each day collecting data. In the case of STXM (the subject of this study) and X-ray ptychography, Sarbajit's group has spent years across several doctoral students collecting what now amounts to mountains of data. As scientists, I think too often we rely on a combination of our education, experience, and training to analyze the data and arrive at insights and conclusions—leaving vast swaths of data uninterrogated. Data science represents another powerful means of combining through these mountains of data that can sometimes further strengthen our conclusions or even point us in the direction of important aspects that we missed the first time around. In this way, data science is responsible science—a way that we can make the most of the data that we spend time and resources to acquire.

DAS: I think that Justin made a good point about how we spend time (which costs money) and resources (which cost money) to collect data. Data are therefore monetarily valuable. The scientific insight we gain from data is intellectually valuable. For me, data science is a field of study that extracts meaningful insights at the intersection of these two value streams. As someone whose formal training is primarily in chemistry, I champion an inclusive definition since it allows for multi-disciplinary teams, such as our own, to come together and develop a data science solution for a problem with immense practical importance. While I do not identify exclusively as a data scientist, it is exciting to know that data science is not beyond my reach. I would encourage other researchers to confidently cross the boundaries of traditional disciplines and subdisciplines in the pursuit of ground-breaking research!

BL: I see data science as a tool that enables data to have the ability to solve different fields of problems. Consequently, a data scientist is someone who crafts such tools. Therefore, I identify myself as a user of these tools with the ability to optimize such tools for my use cases.

What motivated you to become a (data) researcher? Is there anyone/anything that helped guide you on your path?

BL: When I started my doctoral research a few years ago, Dr. Xu and I had an impromptu discussion during a project meeting that would profoundly shape the direction of my studies. Over the course of the discussion, we started to wonder whether we could use simulated data as a basis for data-driven approaches to provide insight into the materials properties of fibrous paper-like materials—which are truly among the most random and chaotic material structures. This deep discussion prompted my first work using machine learning and data-driven research, inspired much of my doctoral research, and has led to several fruitful collaborations on various material systems. A fine example of that is the work published here in *Patterns* in collaboration with the group of Dr. Banerjee.

JLA: Sarbajit Banerjee is a one-of-a-kind advisor and mentor. To say I learned a lot while studying in his group would be a massive understatement. One thing I am constantly amazed by is his drive to incorporate new techniques and expertise into his research program. He is constantly investing in professional relationships outside of chemistry and building connections that frequently blossom into fruitful collaborations. When I first joined the lab, his group was mostly focused on designing nanomaterials for various applications. How did we end up incorporating data science, one might ask? Sarbajit cultivated expertise in data science by recruiting several students through the NSF-funded D^{3}EM fellowship at Texas A&M University. This was a program that recruited graduate students from disciplines across the sciences and taught them to apply data science techniques to their dissertation research projects. Learning from Sarbajit has impressed upon me not only the growing importance of data science but also more broadly the importance of establishing collaborations and searching for solutions from adjacent fields and disciplines.

DAS: Prior to graduate school, I had no exposure to data science. Working in a highly interdisciplinary and collaborative research environment at Texas A&M University allowed me ample opportunities to learn and grow into the role of a (data) researcher. Much like Justin, when I think about what drove me to data science, I would point toward Prof. Sarbajit Banerjee, specifically highlighting his fearless approach to initiating collaborations outside the walls of our immediate discipline. Looking back, perhaps my first major achievement in graduate school was developing an approach to tomographic reconstruction of X-ray microscopy data that allowed us to chemically map a magnesium dendrite.[6] This work, which was in support of ongoing experimental work in the group, ultimately sparked considerable interest (and some much-needed debate) in the broader field of magnesium batteries. As my advisor, Sarbajit encouraged me to pursue this early-career opportunity in data science, which ultimately gave me the confidence to integrate data science approaches into several of my other projects during my doctoral studies.

How can data science help your domain (e.g., science, manufacturing, industry, policy, society), and how can collaborations between academia and your domain be started?

DAS: Having recently transitioned from academia to an industry position, I am reminded how materials innovation drives economic development and is critical to meeting global sustainability targets. In the wake of digitalization, industry, academia, and policymakers alike have begun to emphasize the importance of digitizing and homogenizing data to construct robust materials infrastructures. Directly incorporating data science

tools into these databases can unveil hidden structure-property relationships and propel the scientific discovery of new materials. Today, several partnerships exist between universities, governmental agencies, and industry leaders through a variety of mechanisms. Growing these multi-dimensional relationships between stakeholders through cross-domain alliances is critical to the vitality and longevity of data-driven materials science.

JLA: As a chemist, I belong to a very old scientific tradition that has generated massive amounts of data since the 18th century. Can you imagine if all of these data had been digitized from the outset? OK, I get it—they didn't have computers in the 18th century, but what if, right! I think moving forward, an important advancement will be developing an improved means of digitizing some of these old data (as we did in in a recent review article, Figure 3B)[7] to draw out some new and possibly overlooked conclusions. As academic and industrial researchers begin to collect and organize new data digitally, I think the role of data science will only grow and expand. Imagine all of the unexpected patterns (!) that we might have missed or might start to notice!

BL: As a computational engineer, data science is becoming increasingly critical to enabling effortless knowledge transfer, for example, to material experimentalists using surrogate models based on simulation data, which can usually be produced in less time-intensive and expensive ways, we can study materials from a multi-physical perspective, testing various material microstructures. Furthermore, this allows the optimization of desired properties of the considered material, even under harsh environmental influences, to which experiments have limited access and are extremely difficult to perform.

What is your advice for future data scientists?

BL: Learning from examples has always been helpful for me. I recommend looking at the literature (*Patterns* would be a good start) to see how other researchers have attempted to tackle problems in their respective fields. You never know when you might be inspired by something you read about in a different field.

JLA: Perhaps I can give some advice to chemists and materials scientists on the fence about incorporating data science into their research programs. Sometimes I think that we can be reluctant to step outside of our comfort zone. Data science is a great area to step into and use as a basis for building collaborations since it tends to unite researchers from disparate fields toward a common goal. If you have the data (and I know you do), why not use all of the tools available to you?

How was the experience publishing the first paper from your own team? How did it differ from publishing as first author on your previous papers? How long did this take?

JLA: I'm not a group leader (yet!), but this is my first paper as a corresponding author. It is very humbling, and I consider it an honor. This work represents to me the culmination of several years of thinking about the electronic structure of V_2O_5 cathode materials and working to advance the techniques that we present in the paper. It was an absolute pleasure to work with the other authors. I want to single out David, whom I have worked with closely since my third year of graduate school. I watched as he developed into a world-class researcher and exercised a level of independence that was impressive—so this was a rewarding experience for me on several different levels. I plan to take the lessons I learned about mentoring and publishing and apply them in my own group as I move into my independent career.

What are the challenges of collaboration between teams in different countries?

BL: One challenge is that we have fewer opportunities to talk to each other in person. My collaboration with David and Justin started at the beginning of the global pandemic, so it wasn't easy to meet up, and we still have not met in person since I joined the collaboration. Besides that, I felt pretty comfortable with the file exchanges and meetings in an online format.

JLA: Researchers on both sides of the collaboration are incredibly hard working, efficient at communicating, and passionate/knowledgeable about their science. I was always so impressed by how prepared Prof. Xu's group was for the meetings and amazed at the progress they made between meetings. One thing that helped solidify our international collaboration early on was making the effort to meet in person. Our first trip to Germany to visit TU Darmstadt was in 2018, just a bit over a year after our collaboration started. This trip was one of my favorite experiences *ever* and really cemented the collaboration in my mind. Prof. Xu's group visited College Station, Texas, the next year (though I'm afraid the scenery we had to share with them wasn't quite on par with the lift tour they gave us of the vineyards at Rüdesheim am Rhein). Regardless, the rest is history!

What's next for both of your teams? Can we expect more collaborations?

BL: You can expect many more collaborations between our groups! Now that we have developed the data-driven approaches reported in this article, we plan to apply them to STXM analysis in the future. We might also take another look at older data to see if we can draw out new conclusions and information. We are very excited about these techniques and are eager to apply them moving forward!

REFERENCES

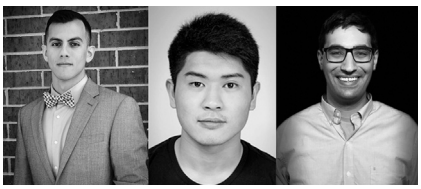
1. Andrews, J.L., Stein, P., Santos, D.A., Chalker, C.J., De Jesus, L.R., Davidson, R.D., Gross, M.A., Pharr, M., Batteas, J.D., Xu, B.X., et al. (2020). Curvature-Induced Modification of Mechano-Electrochemical Coupling and Nucleation Kinetics in a Cathode Material. Matter *3*, 1754–1773. https://doi.org/10.1016/j.matt.2020.08.030.
2. Santos, D.A., Andrews, J.L., Bai, Y., Stein, P., Luo, Y., Zhang, Y., Pharr, M., Xu, B.-X., and Banerjee, S. (2020). Bending good beats breaking bad: phase separation patterns in individual cathode particles upon lithiation and delithiation. Horizons *7*, 3275–3290. https://doi.org/10.1039/d0mh01240h.
3. Zhao, Y., De Jesus, L.R., Stein, P., Horrocks, G.A., Banerjee, S., Xu, B.X., and Xu, B.-X. (2017). Modeling of Phase Separation across Interconnected Electrode Particles in Lithium-Ion Batteries. RSC Adv. *7*, 41254–41264. https://doi.org/10.1039/c7ra07352f.
4. Santos, D.A., Andrews, J.L., Lin, B., De Jesus, L.R., Luo, Y., Pas, S., Gross, M.A., Carillo, L., Stein, P., Ding, Y., et al. (2022). Multivariate hyperspectral data analytics across length scales to probe compositional, phase, and strain heterogeneities in electrode materials. Patterns *3*, 100634. https://doi.org/10.1016/j.patter.2022.100634.
5. D.Jesus, L.R., Horrocks, G.A., Liang, Y., Parija, A., Jaye, C., Wangoh, L., Wang, J., Fischer, D.A., Piper, L.F.J., et al. (2016). Mapping Polaronic States and Lithiation Gradients in

Individual V2O5 Nanowires. Nat. Commun. *7*, 12022. https://doi.org/10.1038/ncomms12022.

6. Davidson, R., Verma, A., Santos, D., Hao, F., Fincher, C., Xiang, S., Van Buskirk, J., Xie, K., Pharr, M., Mukherjee, P.P., and Banerjee, S. (2019). Formation of Magnesium Dendrites during Electrodeposition. ACS Energy Lett. *4*, 375–376. https://doi.org/10.1021/acsenergylett.8b02470.

7. Andrews, J.L., Santos, D.A., Meyyappan, M., Williams, R.S., and Banerjee, S. (2019). Building Brain-inspired Logic Circuits from Dynamically Switchable Transition-Metal Oxides. Trends Chem. *1*, 711–726. https://doi.org/10.1016/j.trechm.2019.07.005.

About the authors



From left to right: Dr. David A. Santos, Binbin Lin, Dr. Justin L. Andrews

David A. Santos is a senior research specialist at Dow Chemical Company in the polyurethane segment. He received his PhD in chemistry from Texas A&M University (@TAMUChemistry) under the advisement Sarbajit Banerjee (@SarbajitBanerj1). While there, he worked as an NSF Graduate Research Fellow to address challenges in energy storage spanning length scales all the way from the level of atoms and bonds to device level performance and systems-level analysis. Prior to that, he received a BS in chemistry from the University of the Incarnate Word where he worked on the chemistry of ruthenium complexes under the advisement of Robert N. Garner.

From left to right: Dr. Justin L. Andrews, Prof. Luis R. De Jesus, Prof. Sarbajit Banerjee, Prof. Ying Zhao (taken at the beginning of our collaboration in Darmstadt, 2018)

Binbin Lin is a PhD student at the Institute of Materials Science at TU Darmstadt (@TUDarmstadt). There he works in the Mechanics of Functional Materials Group under the advisement of Prof. Dr. Bai-Xiang Xu. In addition to studying the implications of phase inhomogeneity in cathode materials during cycling, Binbin is also interested in micro-mechanical modeling and simulations of functional paper-like materials. Prior to that, he received a BSc in applied mechanics at TU Darmstadt. Binbin has also worked as a researcher at the Institut für Statik und Konstruktion, the Frauenhofer LBF, and as an exchange student at INP Grenoble.

From left to right: Dr. David A. Santos, Prof. Dr. Bai-Xiang Xu, Prof. Sarbajit Banerjee

Justin L. Andrews is a post-doctoral researcher at Massachusetts Institute of Technology. There he designs porous framework materials for application as beyond-lithium energy storage vectors in the lab of Prof. Mircea Dincă. Prior to that, he received a PhD in chemistry from Texas A&M University under the advisement of Prof. Sarbajit Banerjee. While there, he worked as a NASA Space Technology Research Fellow designing materials for application in magnesium-ion batteries, photocatalysis, and neuromorphic computing. Before that, Justin obtained his BS in chemistry from Gordon College, where his undergraduate research focused on developing green chemistry experiments for organic chemistry curricula.