Paper PD08

From Statistical Programming to AI Automation in Data Management: A Cross-Departmental Tech Journey

Irene Luna, Roche Products Ltd, London, United Kingdom
Marcia Vaz, Roche Products Ltd, London, United Kingdom

ABSTRACT

This paper reflects on our journey through two distinct yet interconnected domains within Roche: Analytical Data Science (ADS) and Data Management AI. Starting with no prior clinical background, we navigated the steep learning curve of programming in SAS and R, clinical trial documentation, and the strict regulatory environment of ADS. At the same time, we explored innovative, AI-driven projects that sought to automate manual processes and improve efficiency in clinical data management. By comparing these structured and exploratory experiences, we highlight the challenges, skills acquired, and the broader value of blending precision and reliability with creativity and experimentation. Ultimately, we argue that this hybrid approach not only enhanced our technical and collaborative skills but also positioned us as adaptable generalists in a specialized industry.

INTRODUCTION

The pharmaceutical industry is undergoing a transformation where data science and artificial intelligence are reshaping traditional workflows. Analytical Data Science (ADS) teams ensure the rigorous programming, documentation, and compliance necessary for health authority submissions, while Data Management AI initiatives experiment with automation and large-scale data integration to reduce inefficiencies.

This paper shares our perspective as newcomers entering Roche without prior clinical backgrounds. We explore two worlds: first, the demanding yet structured environment of ADS, where we learned clinical programming under high stakes, and second, the experimental environment of Data Management AI, where we applied Python, AI, and automation to modernize clinical workflows. Through this dual experience, we uncover the value of both approaches and how they complement one another in shaping adaptable, resilient professionals in life sciences.

ANALYTICAL DATA SCIENCE (ADS) EXPERIENCE

When we joined Roche, we went straight into Analytical Data Science (ADS) without any clinical background. Irene started on a busy phase 3 ophthalmology study, while Marcia began on a busy late-stage oncology study that was moving toward FDA approval. It was an intense way to start. Suddenly, everything was new: learning a new scientific area, adapting to clinical trial processes, and at the same time picking up two programming languages, R and SAS. To make things even more overwhelming, there was a flood of acronyms SDTM, ADaM, TLGs, CSR so it often felt like learning an entirely new language every single day.

CHALLENGES AND LEARNING CURVE

Programming itself came with its own challenges. Coming from Python backgrounds, SAS was especially tough to adapt to. There wasn't much help online for troubleshooting, and concepts like macros were completely unfamiliar at first. SAS could also feel very confusing compared to Python. We discovered some of its real strengths, and saw how it could handle certain tasks very well. Still, the language often felt clunky, and figuring things out could quickly get complicated. R felt more flexible and powerful, especially for analysis, but in practice creativity was limited since we had to use standard packages for consistency. This worked most of the time, but when study-specific outputs were needed, things became more complicated. We often had to go function by function to troubleshoot, especially when using internal packages with limited documentation.

Beyond programming, the way of working itself was a steep learning curve. Documentation was thorough: every step needed to be written down how a variable was created, why a dataset was updated, what outputs were produced. At first, it felt excessive, but we came to understand the reason: everything needed to be traceable, and every dataset or output, even the smaller ones, mattered. They were all used for reporting or decision-making, so nothing was wasted work.

BEYOND CODING

Our time in ADS gave us much more than just technical programming skills. It showed us the bigger picture of how clinical trials operate, how outputs are created, and why every detail matters.

All of this required constant communication with people from other teams, including clinical science, statistics, and data management. The structure of the work was very clear: there was always a straightforward path of what needed to be done, why it needed to be done, and when it was due. The strict deadlines and reporting requirements kept everything moving, while the continuous checks and collaboration ensured the outputs were of high quality.

It was overwhelming at first, but looking back, working on such high-stakes studies gave us no choice but to learn fast. There was no better crash course in programming, acronyms, documentation, and clinical trials than contributing directly to studies heading into health authority submissions.

KEY TAKEAWAYS FROM ADS

Working in ADS taught us that every challenge is an opportunity to grow. The complexity of clinical trials and strict regulatory standards pushed us to adapt quickly, think critically, and deliver with precision. We learned that precision builds trust. It's not just a requirement, but the foundation of reliable analysis and sound decision-making.

Also, collaboration played a vital role in amplifying our impact. Close coordination with statisticians, data managers, and clinical scientists ensured that every output was not only technically correct but also scientifically meaningful.

The structured and straightforward workflow in ADS provided clarity and consistency. Defined tasks, timelines, and processes minimized uncertainty and created an ideal environment for learning and growth especially for new joiners. Through this experience, we also discovered the importance of adaptability, learning new tools, languages, and ways of working to stay effective in a constantly evolving environment.

Finally, ADS reminded us to always look at the bigger picture: understanding how our work supports better clinical decisions, enhances trial quality, and ultimately contributes to improving patient outcomes

EXPLORING TWO WORLDS: AI IN DATA MANAGEMENT

While ADS provided structure and expertise in clinical trials, we also wanted to continue developing our Python skills and explore creative problem-solving. At the same time, AI was becoming increasingly influential in the pharmaceutical industry, opening opportunities to streamline workflows and reduce manual tasks. Within Roche, the Data Management AI team develops use cases aimed at automating time-consuming steps in clinical trials, which aligned perfectly with our goals and curiosity.

The Data Management AI team develops use cases, which are targeted projects designed to tackle specific challenges within Data Management. Each use case follows a structured process that begins with ideation and scoping, progresses through proof of concept and minimum viable product (MVP) development, and continues with beta testing, launch, and monitoring. This systematic approach ensures that every AI solution is both practical and impactful, enabling continuous improvement while driving automation and efficiency in clinical data workflows.

Not all use cases, however, reach full deployment. Some progress successfully through the entire lifecycle and are implemented in production environments, directly supporting ongoing studies and data operations. Others may stop at the proof-of-concept or MVP stage—either because their feasibility, scalability, or impact is still being evaluated, or because priorities shift as new challenges emerge. Regardless of the outcome, each use case contributes valuable learning, refining methods, tools, and approaches that inform future developments and enhance the team's collective knowledge.

USE CASE 9: CURRENT PROCESS

This led us to join one of their use case projects focused on integrating lab ranges into a central database, an area known for its complexity and high risk of errors. Clinical trials rely on data from thousands of local laboratories worldwide, yet processes remain largely manual. Many documents arrive handwritten or in low-quality formats, making data handling inefficient. Today, Local lab ranges are entered into RAVE by data managers, often one site at a time, resubmitted for each new site, and in some cases, are only provided at the very end of a study. Clinical

Research Associates (CRAs) spend time chasing labs for documents and uploading them, which causes delays and often leads to duplicate entries across studies. On top of that, lab ranges can only be cleaned after they have been entered into RAVE, which creates repeated queries and slows things down further. With multiple handovers and manual steps involved, errors are common and the process consumes significant resources.

USE CASE 9: SOLUTION AND IMPACT & CONSIDERATIONS

To address these challenges, we joined the team developing a semi-automated solution using OCR, Generative AI, and Large Language Models. The system takes unstructured lab documents, digitizes them, and processes the data. It then maps local lab ranges to the study-specific ranges, building a centralized database that can cover more than 3,000 labs worldwide. The application also integrates with existing lab range and normal value datasets, while a human-in-the-loop component reviews any low-confidence matches to ensure quality and double checks the automatic matches.

This approach streamlines data management by reducing manual entry and repeated queries, standardizing data, eliminating duplicate ranges, accelerating lab data handling, shortening trial timelines, lowering costs, and improving communication with monitors and labs. As a result, data managers receive more consistent and reliable inputs.

However, some challenges still remain. Manual quality checks are still necessary, the tool currently supports only English documents, and very low-quality scans can cause issues. A careful oversight is therefore essential to ensure compliance and maintain data quality.

EXPERIMENTATION AND LEARNING

ADAPTATION ACROSS WORLDS: TRANSLATING EXPERIENCE INTO GROWTH

ADS provided structure, precision, and the foundation for producing reliable, meaningful outputs. In contrast, the AI projects pushed us into more unstructured territory, where we had to think differently, embrace trial and error, and adapt quickly. While many of these projects may never fully launch, the learning they offer is immense—and if they do succeed, their impact on the organization could be substantial. This balance between structured and exploratory work proved essential: ADS ensures stability and immediate value for the company, while AI initiatives build future capabilities and foster personal growth.

Our collaboration with DSD on UC9 clearly illustrated this balance in action. Through that experience, we observed the entire data lifecycle from collection and curation to its delivery into Data Science. Gaining insight into upstream decisions and constraints deepened our understanding of why data appears as it does when it reaches us. This perspective directly improved our ADS work: when creating ADaMs from lab data, we could better anticipate issues related to units, visit windows, or reconciliation, implement earlier checks, and coordinate more effectively with data managers.

Exposure to new topics such as APIs and data integration further broadened how we think about data flows, even when these were outside our daily responsibilities. Just as importantly, the flexibility and creative thinking developed through the AI projects carried over into our ADS work. In situations like ad hoc submissions where standard processes didn’t always apply and we were able to draw on that experience to problem-solve effectively and find innovative solutions.

Time management emerged as another key challenge. Study deliverables often took precedence, occasionally pushing UC9 to the side. Catching up in bursts was demanding, but it taught the team how to prioritise tasks, manage energy, and maintain momentum under competing pressures.

UC9 also encouraged experimentation by testing new tools, exploring different approaches, and tackling problems from multiple angles. Although roadmaps and deadlines existed, there was no single defined path forward. Trial and error became central to progress, often creating ambiguity about the right direction or next step. Some approaches, such as extensively testing programs, required significant effort before being replaced by better solutions. While not every experiment produced an immediate outcome, each one contributed to our growth by reinforcing problem-solving, creativity, and flexibility.

The contrast in resources and approaches between ADS and UC9.1 further highlighted these differences. In ADS, most knowledge stemmed from internal processes and institutional expertise, whereas UC9.1 opened the door to exploring external libraries, testing new tools, and engaging with the broader tech community. This exposure nurtured curiosity, continuous learning, and skill development beyond what was typically possible in the more regulated ADS environment.

Overall, the challenges and variety within UC9.1 significantly accelerated our growth. By navigating ambiguity, managing competing priorities, and experimenting with innovative tools, the team strengthened its problem-solving abilities, adaptability, and gained a more holistic understanding of clinical data workflows.

CROSS-FUNCTIONAL DEVELOPMENT

Building on our experience navigating both structured and exploratory environments, adapting across these two worlds presented its own unique challenges. The team had to constantly shift between clarity and ambiguity by balancing the strict, regulation-driven demands of clinical studies with the open-ended experimentation of AI projects. What was acceptable, even encouraged, in an exploratory AI setting but rapid iteration, trial and error, and flexible thinking was not always suitable within the rigor of a late-phase clinical trial. Learning to move fluidly between these expectations strengthened our ability to manage diverse priorities and operate effectively across contexts.

This experience also tied closely to the idea of generalists versus specialists, as discussed by David Epstein in his book Range: Why Generalists Triumph in a Specialized World. Epstein argues that in a complex and fast-changing environment, generalists often thrive because their varied experiences help them think creatively, make broader connections, and adapt more effectively to uncertainty. Specialists, on the other hand, provide the depth and precision that ensure reliability and rigor.

This experience also highlighted the importance of avoiding the “Law of the Instrument” which is the idea famously summarized by Abraham Maslow as, “I suppose it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail.” In practice, this meant learning not to apply the same mindset or tools to every problem. While ADS required precision, consistency, and strict adherence to standards, the AI projects demanded experimentation, creative problem-solving, and comfort with ambiguity. Because of this, we became not only stronger programmers but also more effective collaborators.. Recognising when to switch approaches and when not to force one context’s methods onto another became a key aspect of our growth.

Ultimately, the combination of ADS and AI work helped us grow both as specialists in clinical data and as generalists capable of adaptation and cross-domain thinking. We learned to translate between the structured requirements of clinical programming and the exploratory approaches of AI innovation bridging disciplines, facilitating communication, and adding value across teams embodying the adaptability that *Range* celebrates as key to success in complex, interdisciplinary worlds.

CONCLUSION

Looking back, what we have gained from working across ADS and AI projects goes far beyond technical skills. The biggest lesson has been the value of trying new things. You never really know if you will enjoy or be good at something until you give it a go. Curiosity should drive you, because even if you later decide it’s not for you, the experience will still teach you something useful.

We also learned how interconnected different departments really are. When you are focused on your day-to-day work, it’s easy to miss the bigger picture and not notice what new tools or approaches are being developed around you. By stepping outside our usual tasks, we were able to see how much collaboration and cross-team awareness matter.

Another key takeaway is the importance of constantly expanding your knowledge. Whether it’s learning a new programming language, exploring software development practices, or picking up product management skills like Scrum, every new skill keeps your mind sharp and prevents you from getting stuck in repetitive tasks. Each small learning experience builds on the last and gives you new ways of thinking.

These experiences also showed us the danger of narrow thinking. Specialists can sometimes fall into the trap of only seeing problems through the lens of what they already know the “Law of the Instrument”. However, broad

experiences help avoid this by giving you multiple perspectives to draw from. That's why having a "sampling period," where you try different kinds of work before specialising, is so valuable. It helps you discover your real interests and develop a more versatile skill set.

Finally, we came to see how essential AI will be for the future of pharma and beyond. It is not enough to specialise narrowly but it is important to keep building skills across areas so you can adapt as the industry evolves. For us, bridging structured ADS work with exploratory AI projects has not just been a career shift but a personal development journey. These lessons continue to shape how we grow today.

RECOMMENDED READING

Epstein, David. *Range: Why Generalists Triumph in a Specialized World.* New York: Riverhead Books, 2019.

Maslow, Abraham H. *The Psychology of Science: A Reconnaissance.* New York: Harper & Row, 1966.

CONTACT INFORMATION

(In case a reader wants to get in touch with you, please put your contact information at the end of the paper.)
Your comments and questions are valued and encouraged. Contact the author at:
Author Name: Irene Luna
Company: Roche Products LTD
Address: -
Work Phone: 07522235431
Email: irene.luna@roche.com
Website: -

Author Name: Marcia Vaz
Company: Roche Products LTD
Address: -
Work Phone: 07824034915
Email: marcia.vaz@roche.com
Website: -