STATISTICAL PROGRAMMING MANIFESTO · V1.3 · 2026

Code Without Purpose.

We built one of the most technically sophisticated disciplines in pharmaceutical science. Then we forgot what it was for.

Sascha Ahrweiler · Global Head of Statistical Programming & Analysis Strategy & Operations · Bayer AG
PHUSE Board Member, Director Communication & Brand Strategy
datacaterer.com

> THE QUESTION NOBODY WANTS TO ANSWER
>
> *If you removed SDTM, ADaM, and TFLs entirely from our world tomorrow, what percentage of statistical programming organizations could still meaningfully contribute to a scientific or clinical decision?*
>
> **The honest estimate: less than 20%. And the reason this manifesto exists.**

CORE MANIFESTO · THE DECLARATION

Manifesto for the Data Caterer™

We are uncovering better ways to turn clinical data into decisions — by doing this work and helping others grow in it. Through twenty-nine years of this work, we have come to value:

Understanding the guest	OVER	executing the specification
Scientific judgment	OVER	standard compliance
Enabling decisions	OVER	delivering outputs
Owning the question	OVER	answering the task
Earning your seat at the table	OVER	waiting to be invited

That is, while there is value in the items on the right — specifications must be met, standards must be respected, outputs must be delivered — **we value the items on the left more.**

The patient is the ultimate guest. Every dataset exists to serve them. We refuse to lose sight of that.

Twelve Principles of the Data Caterer

01 **The clinical question comes first.** Before we write a line of code, we ask: what decision does this data need to support, and for whom?

02 **We know the guest before they sit down.** Understanding our audience — the physician, the regulator, the DSMB, the statistician — is not optional. It is the work.

03 **Standards are a floor, not a ceiling.** CDISC exists to enable. A programmer who cannot think beyond SDTM and ADaM has mistaken the plate for the meal.

04 **Correctness without comprehension is a liability.** We once produced a boxplot we could not explain. We will not let that happen again.

05 **The specification is a starting point, not a destination.** A great Data Caterer reads the protocol to understand the science — not just to follow the instructions.

06 **We own our own development.** No employer, certification body, or offshoring model will grow our judgment for us. That work is ours.

07 **SAS versus R is the wrong question.** The right question is: am I using my tools to advance understanding, or to protect my comfort zone?

08 **AI will take the last mile. We must own the mile before it.** The execution layer is automatable. The judgment, the wisdom, the scientific understanding — that is where we build irreplaceability.

09 **Drift is a choice.** The slow, comfortable slide into pure execution happens through a thousand small decisions. We choose differently.

10 **The best data storytelling renders itself invisible.** When a decision-maker acts with confidence, the data behind it has done its job. We serve the decision, not our own craft.

11 **We curate the kitchen so others can cook.** Great Data Caterers design systems, govern assets, and build platforms that multiply the intelligence of everyone who touches the data after them.

12 **The patient is always in the room.** Behind every dataset is a trial. Behind every trial is a patient who trusted the system with their health. We do not forget this.

EXTENDED MANIFESTO · THE ESSAY

01

We Optimized the Wrong Thing.

For decades, statistical programmers did something genuinely impressive. They built macro libraries, standardized workflows, mastered CDISC, and relentlessly engineered the speed and quality of their outputs. SDTM datasets, ADaM structures, TFLs — produced faster, cleaner, more consistently than ever before.

And in doing so, they forgot one question: **what are those outputs actually for?**

The languages changed. SAS gave way to R. Proprietary tools gave way to open-source communities like Pharmaverse. But the underlying logic — receive a specification, produce an output, move on — remained perfectly, stubbornly intact.

Leaders doubled down on *how*, never asking *what for*. Efficiency became the religion. Effectiveness never made it into the creed.

The result: a profession structurally incapable of answering the question that matters most. Not "did we produce the output?" — but "did it help anyone decide anything worth deciding?"

> *We became world-class at producing artifacts. We became strangers to the science those artifacts were meant to serve.*

02

The Moment It Broke.

There was a specific moment when the fracture began. It was not a crisis. It was a job posting.

The moment the role of "SAS Programmer" was formally separated from "Statistician," scientific understanding and operational execution were cleaved apart. Statistics would do the thinking. Programming would do the doing. It seemed efficient. It was the beginning of a slow hollowing out.

CDISC arrived in 1997 and made things worse in a way nobody intended. Standards meant to enable scientific communication became the cognitive boundary of an entire profession. If it fits the standard, it ships. If it doesn't, it gets bent until it does. **Fit-for-purpose gave way to fit-for-compliance.**

Offshoring deepened the wound. The promise: do the same work, cheaper, faster. The reality: more specifications, more oversight, more review cycles. An entire bureaucracy built to describe what to build — instead of helping people understand why they were building it.

Ask a statistical programmer what their output means clinically. Listen for the silence that follows.

03

The Field Stands Indicted.

01 **Efficiency Over Effectiveness**
We optimized delivery speed while the value of what we delivered quietly eroded. Faster outputs that nobody understands serve no one — least of all the patients at the end of every dataset we produce.

02 **Specification Dependency**
We outsourced our thinking to instructions. We built a profession that executes commands rather than exercises judgment. A checklist is not a career. A specification is not understanding.

03 **Scientific Disconnection**
We became uncomfortable talking to clinical colleagues without a statistician in the room. We surrendered the conversation that matters most — and called the distance "specialization."

04 **Standard Worship**
We confused compliance with contribution. CDISC is a means. We made it the mission. The standard describes a container — not the substance it is meant to carry.

04

What I Believe.

I believe statistical programming is one of the most strategically important — and most poorly leveraged — disciplines in pharmaceutical R&D.

I believe the next generation of statistical programmers will not be SAS programmers or R programmers. They will be **data strategists, decision enablers, and clinical translators** — and they will be irreplaceable.

I believe that the function we have built is sustainable only as long as the past holds. The future requires something different. **It requires us to stop being scribes and start being authors.**

05

AI Is Not the Threat. You Are.

AI is not coming for statistical programmers. AI is coming for **execution-only roles**, in every profession, including this one.

If your contribution is "I take a specification and produce an output that meets it," you are competing with a generative model that does this faster, cheaper, and at higher consistency than you can.

The threat is not the tool. The threat is what the tool reveals: **a profession that stopped reaching for the science the moment standards made execution sufficient.**

> *AI will take the last mile. We must own the mile before it.*

06

The Destination: From Line Cook to Data Caterer.

The Data Caterer is a vision of what statistical programming becomes when it remembers what it is for. Not a technician executing a recipe. A practitioner who reads the room, understands the guest, designs the meal, and is held accountable for whether the meal accomplished what it was meant to.

The journey is a menu of courses. Each course is more demanding than the last — not in technical complexity, but in scientific understanding, judgment, and the willingness to own the outcome.

07

The Journey Every Programmer Must Choose.

Every programmer is at one of these courses today, whether or not they have named it. The journey from Line Cook to Data Caterer is not a promotion path. It is a posture shift — a decision to extend your contribution beyond the boundary where the specification ends.

No one moves you forward except you. No tool, no training, no certification, no organisation. You choose. Daily. Deliberately.

08

The Kitchen Has Seven Stations.

Drug development is not a pipeline. It is a kitchen with seven distinct stations — each with native data, its own course mix, and a cross-pollination that defines the discipline.

01	**Non-clinical Research**
02	**Early Clinical Development**
03	**Late Clinical Development**
04	**Regulatory Submission**
05	**Medical Affairs**

06 **HEOR / HTA**

07 **Commercial & Publication**

The value chain is not a pipeline. The Data Caterer is shaped by the station they serve.

09

What We Carry to the Station.

The Data Caterer is not built from a single skill. They carry a comb — five teeth of deep craft, a crossbar of station-agnostic competencies, and a frontier tooth that bends to whatever the station demands.

The Teeth (deep craft): Clinical & Therapeutic Fluency · Regulatory Instinct · Standards Mastery · Statistical Literacy · Data Architecture & Engineering.

The Crossbar (the carrier): Decision Communication · Cross-Functional Operating · Curatorship of Assets & Systems.

The Frontier Tooth: AI literacy and agentic-system curation. Not a replacement for the comb — a bending of it toward the work the station now requires.

> *The skill is not to grow every tooth to the same length. It is to know which teeth your station demands, grow them deliberately, and never mistake a long tooth for a full comb.*

10

What We Demand of Ourselves.

This is not a memo to leadership. This is not a call to management. This is a demand that every statistical programmer makes of themselves — now, before the choice is made for them.

I **Own the Why**
Learn the science behind every dataset you touch. Understand the clinical question. Know why the trial was designed the way it was. A programmer who understands the why cannot be replaced by a prompt.

II **Refuse the Checklist**
Use standards. Do not worship them. When CDISC forces a fit-for-standard answer where fit-for-purpose is needed, say so. Loudly. Own the judgment — not just the output.

III **Enter the Room**
Stop waiting for statistics to translate the clinical question for you. Build relationships with clinical teams. Be present where decisions are made. Bring your expertise without being summoned.

IV **Evolve the Craft**

R is not the answer. Open source is not the answer. New tools in old patterns change nothing. Evolve your thinking first. Let the tools follow. Be a data strategist — not a syntax migrant.

V **Build Your Knowledge System**

The Omakase course — the highest form of Data Catering — requires accumulated wisdom you cannot borrow. Build a personal knowledge system. Read. Reflect. Connect insights across domains. Your PKM is your competitive moat.

VI **Own Your Development**

Do not wait for your organization to define your future. The Data Caterer™ is not a job title your employer grants. It is a posture you choose — daily, deliberately, before it becomes urgent. The menu is in front of you. Start cooking.

AI EXTENSION · COMPANION TO THE MANIFESTO · 2026

VI

The Multiplier Problem.

Artificial intelligence has arrived in statistical programming. The tools are real. The productivity gains are real. The threat — for those who have not yet made the transformation — is also real.

But not for the reason most people assume.

The assumption is that AI will replace statistical programmers. That is the wrong fear. The right fear is subtler: **AI will not replace statistical programmers. It will replace the ones who were already replaceable.**

A multiplier needs something to multiply.

AI	×	Line Cook	=	Faster execution
The tool		Execution only		Accelerated irrelevance

Multiply AI by a programmer whose entire value lives in execution — translating specifications into code, mapping variables to SDTM domains, producing TFLs from a pre-approved shell — and you get faster execution. The output looks the same. It arrives sooner. The programmer becomes more productive at the work that was already at risk of automation.

This is not liberation. **It is acceleration toward irrelevance.**

AI	×	Data Caterer™	=	Irreplaceable at scale
The tool		Judgment + science		Strategic amplification

Now multiply AI by a Data Caterer — someone who brings deep data understanding, regulatory judgment, scientific fluency, and stakeholder awareness to every engagement. The Data Caterer who once managed ten studies manages thirty. Not because they type faster. Because AI carries the execution load while the Caterer holds the judgment. The irreplaceable part — understanding what the data means, knowing when a finding is submission-threatening, reading the room in a DSMB — is not automated. **It is amplified.**

> *The value of AI to any individual is proportional to the value they brought before AI arrived.*

If the answer to "what do you contribute beyond execution?" was unclear before, AI has made that question urgent.

VII

The New Warning: The AI-Assisted Imitation.

The five foundational capabilities of the Data Caterer were always the standard. AI has not changed what the Data Caterer must be. It has changed how visible the gap has become — and introduced a new risk.

NEW RISK · 2026

There is now a figure who did not exist five years ago: the **AI-Assisted Imitation**. The statistical programmer who has learned to operate AI tools fluently, who produces polished outputs, who speaks in the language of strategy, who moves quickly and sounds confident — but who lacks the underlying judgment that the Data Caterer brings.

The outputs are professional. The reasoning is borrowed. The caveat that should have been surfaced was not surfaced — because surfacing it required knowing the regulatory history, understanding the estimand implications, having enough scientific context to sense that something was off. The AI did not know. The programmer did not know. The output went forward.

The Line Cook was always recognisable. The specification arrived, the code was produced, the review was completed. The limits of the role were visible.

The AI-Assisted Imitation is not recognisable — until something goes wrong.

This is not a critique of AI. It is a critique of mistaking fluency with tools for depth of understanding. The Data Caterer framework has always been about the latter. In an AI-augmented world, that distinction has never mattered more.

THE TEST

Remove the AI. What remains?

DATA CATERER™
A practitioner with genuine data understanding, regulatory instinct, and scientific grounding. **The AI was a force multiplier.**

IMITATION
Someone who can no longer perform at the level their outputs suggested. **The AI was a mask.**

The discipline needs to ask this question honestly. About its teams. About its structures. About itself.

VIII

What AI-Augmented Data Catering Actually Looks Like.

This is not a warning document. The point of naming these risks is to make the alternative concrete.

The Data Caterer in an AI-augmented world does not use AI less carefully than the Line Cook. **They use it more deliberately.**

I **Expand capacity, do not substitute judgment.**
Where the Line Cook uses AI to produce output faster, the Data Caterer uses AI to cover more ground — more studies, more data sources, more scenarios — while keeping judgment at the centre of every decision.

II **Stress-test thinking, not just produce output.**
The Data Caterer treats AI as a thinking partner, not a vending machine. Prompt for counter-arguments. Ask what has been missed. Use AI to pressure-test an interpretation before presenting it to a regulator or a DSMB.

III **Translate, do not conclude.**
AI is useful for transforming complexity into clarity — summarising a long narrative, rendering a finding in plain language, structuring an argument. The Data Caterer uses this capability in service of communication. The conclusion was already reached. AI helps express it.

IV **Maintain the ability to perform without it.**
This is the non-negotiable. The Data Caterer who cannot explain a finding, reconstruct a derivation, or defend an analytical choice without AI assistance has allowed the tool to become a dependency. That is the Imitation. The Caterer can always step back from the tool and still be the expert in the room.

V **Develop others through it.**
The Data Caterer in an AI-augmented organisation does not hoard the tools. They use AI augmentation as a teaching instrument — exposing the reasoning behind outputs, asking teams to critique AI-generated work, building the judgment that makes the whole organisation more resilient.

> *The five foundational capabilities remain unchanged. AI has not altered what the Data Caterer must be. It has made becoming one more urgent — and the cost of not becoming one more visible.*

This extension was added to the Statistical Programming Manifesto in 2026, in response to the rapid integration of generative AI into clinical data workflows.

The Checklist Is Over. The Craft Begins.

The profession of statistical programming is at an inflection point it will not revisit. Those who treat this as someone else's problem will find themselves with a very clean, very compliant, very automated replacement arriving within the decade.

Those who act now — who reclaim the science, own the judgment, and build toward the Data Caterer™ — will define what this discipline becomes. Not the Line Cook. Not the specification executor. The strategist at the table where decisions that affect patients are made.

> *The menu is in front of you. Every programmer must choose which course to master next.*

Statistical Programming Manifesto · Version 1.3 · 2026

Declaration · Extended Essay · AI Extension (Sections VI–VIII)

Sascha Ahrweiler · Global Head of Statistical Programming & Analysis Strategy & Operations · Bayer AG
PHUSE Board Member, Director Communication & Brand Strategy
datacaterer.com