



What History Teaches Us in Designing a Modern Statistical Computing Environment

Caroline Phares | Domino Data Lab

What we can learn from smartphones

- Smartphones have become essential to life as we know it.
- What happened to businesses that were not future-proof?
- What does this even have to do with statistical computing environments?



Statistical Computing Environment (SCE)

What is an SCE, and how can an SCE help future-proof clinical programming?

To answer this, we need to look back at how these have evolved:

- biostatistics
- statistical computing
- clinical data
- regulations in pharmaceutical development

History of Statistics

Statistics “was developed in the early 19th century as the study of populations, economies, and moral actions and later in that century as the mathematical tool for analyzing such numbers.” ¹

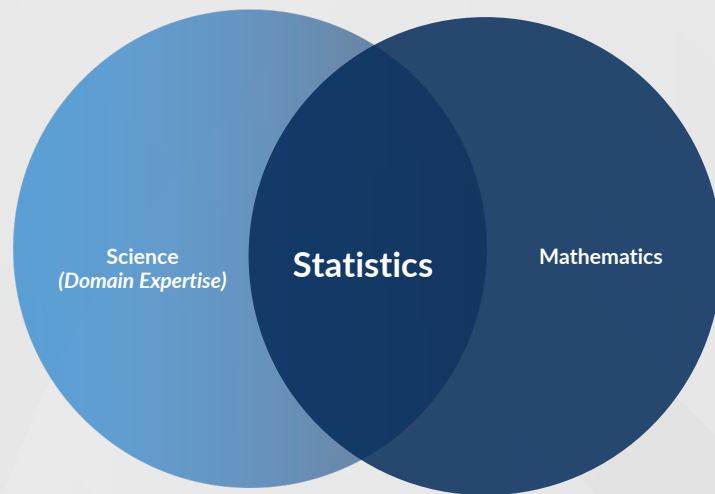
Statistics involves collecting representative data and identifying patterns in the data in order to form predictions and generate models.

Late 1800's - use of generational data to develop hereditary models

1930's - the profession of Statistician emerged:

- Experts in inference
- Defined standards for experimentation (study designs)
- Defined methods of data analysis

¹ <https://www.britannica.com/science/probability>



Required domain expertise and mathematical knowledge

Evolution of Electronic Health Records (EHR)

1960s

- Paper records evolved to include more robust information about a patient, like medical history, rather than just the diagnosis and treatment.
- EHR were so expensive that they were only used by the government in partnership with health organizations.

1970s

- Electronic records were used by the largest hospitals, but only for billing and scheduling.
- Paper records became a means to communicate across healthcare teams, but patient records had to be copied and physically moved.

1980s

- Computers gained presence in physicians' offices and began to be used for record keeping purposes.
- Health Level 7 was founded to address standardization issues as EHR development pushed forward.

1990s

- The internet started being used for recording and transferring prescription histories and other EHR.
- Interoperability became a concern as a growing clinician user base required systems to communicate with each other to effectively coordinate care.

2000s

- With web based software prevalence, EHR became a strategy to: share information privately and securely; improve quality; prevent medical errors; administrative efficiencies.
- EHR implementation nearly doubled from 2007 to 2012.

2010s

- EHR used to provide insights that improve clinical decision-making through predictive analytics.
- Along with the benefits, liability risks, like privacy breaches, errors and miscommunications, were introduced.

2020s

- Precision/personalized medicine, Digital trials, Wearables

Regulations

1960s

Declaration of Helsinki

Protection of vulnerable subjects.

Voluntary consent.

1970s

Declaration of Helsinki revisions

Independent Review of Protocols.

"Informed Consent" always required.

Accuracy in publishing.

Patients' safety and wellbeing always supersedes scientific interests.

1980s

Declaration of Helsinki revisions

Minor subject consent.

Trial is subject to laws and independent review within the country in which it is conducted.

Health Level 7

International data sharing standards.

1990s

ICH

EU, US, & Japan regulatory standards - GCP Guideline to ensure reliability of clinical trial results.

21 CFR Part 11

Computerized system validation.

Electronic records retention.

Audit trails.

Limited authorized access to systems and records.

Electronic signatures

Operational checks.

Education, training, and experience to perform assigned tasks.

2000s

Declaration of Helsinki revision 5

Placebo used where no increased risk of harm.

Patient consent to treatment blindness.

Treatment of subjects at trial conclusion.

Protection of PII.

DoH trumps country laws.

Trials should only be carried out on populations who stand to benefit from the research.

Disclosure.

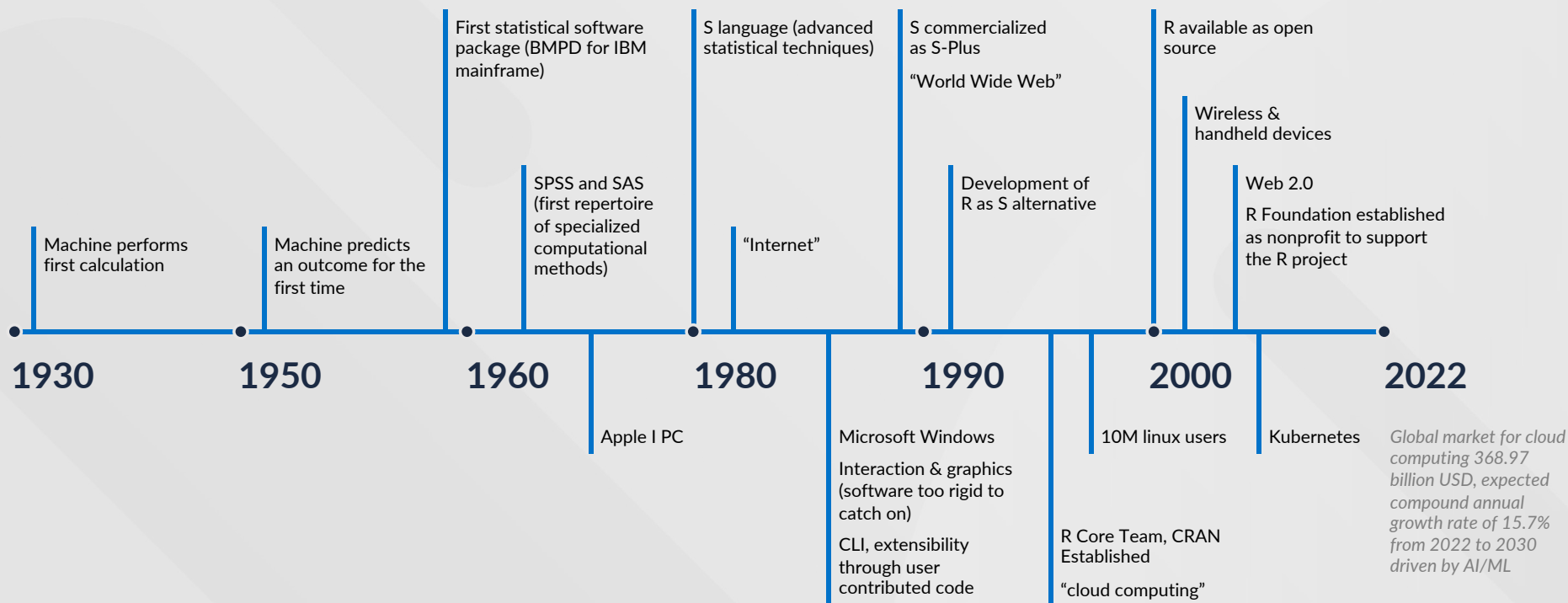
ICH revision 2

Guidelines for more efficient approaches to clinical trial design, conduct, oversight, recording, and reporting.

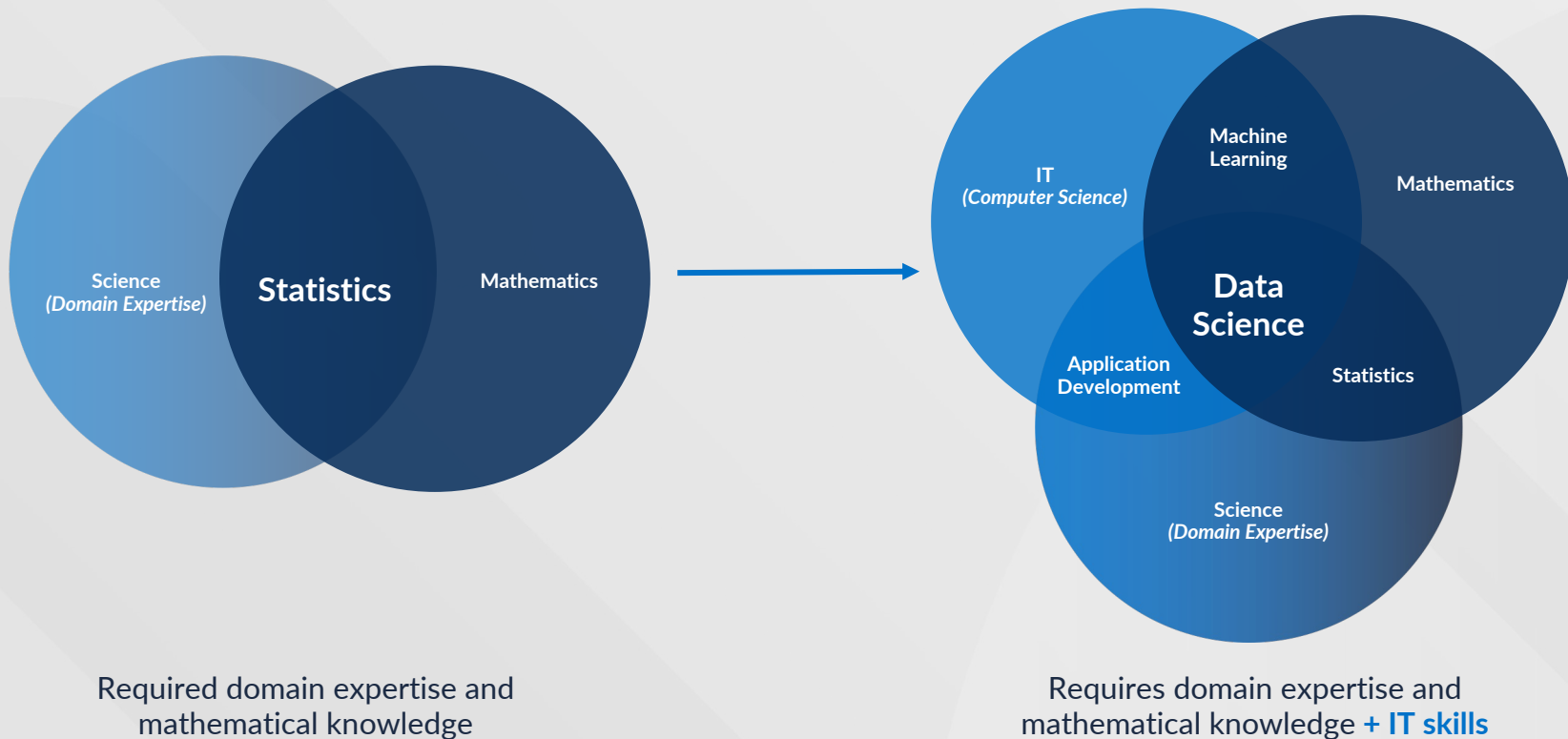
GDPR

Focuses on the core areas of data governance, management, transparency: lawfulness, fairness and transparency; purpose limitation; data minimization; accuracy; storage limitation; integrity and confidentiality; accountability.

90 Years of Statistical Computing



Paradigm Shift from Statistics to Data Science



2021 PHUSE SCE White Paper

“To construct a fully functional SCE that meets evolving business need whilst maintaining regulatory compliance, it must:

- Be an application that is intuitive and easy to use.
- Involve modern and scalable forward focused technology.
- Adopt technology that is reliably used for big data.
- Facilitate multiple programming languages.
- Effectively integrate with a variety of systems and tools to allow seamless access to data.
- Successfully manage programming workflow e.g., version control, traceability, reproducibility.
- Allow effective outsourcing and collaboration.
- Include automated functionality, AI and machine learning to replace repetitive and manual-heavy tasks.
- Include machine learning to facilitate decision making.
- Enable high compute power (high performance computing capability) to enable statistical methods that require complex simulations to be run in minutes.
- Allow the capture and reporting of essential business-related metrics to help track value generation over time, as well as identify possible opportunity areas to drive further value.
- Enable search services and dashboards with drill down functionality.
- Provide capability to anonymize data and documents.
- Seamlessly manage access provisioning and review.
- Easily integrate with sponsor data structure ecosystem.
- Support data and document reproducibility and archiving.
- Use metadata to enable the search and re-use of code.
- Be a high availability cloud-based computing platform.”²

² https://phuse.s3.eu-central-1.amazonaws.com/Events/2021/US+Connect+2021/SCE_White_Paper_Final_15Dec2021.pdf

Future Proofing Data Management and Clinical Programming

- Just like smartphones - data science, AI/ML are here to stay and will only grow in importance.
- Having to catch up is never a good place to be.
- Domino is at the forefront of Statistical Computing, because it was at the forefront of data science ~10 years ago.

**Visit the Domino booth PL3
and Veramed booth 26
to find out more about what
you can do with Domino**

