

ML In Statistical Production Processes III

IT Infrastructure group

...or the "Everything *but* the
actual ML use case"-group

Report from IT Infrastructure Group

SCB

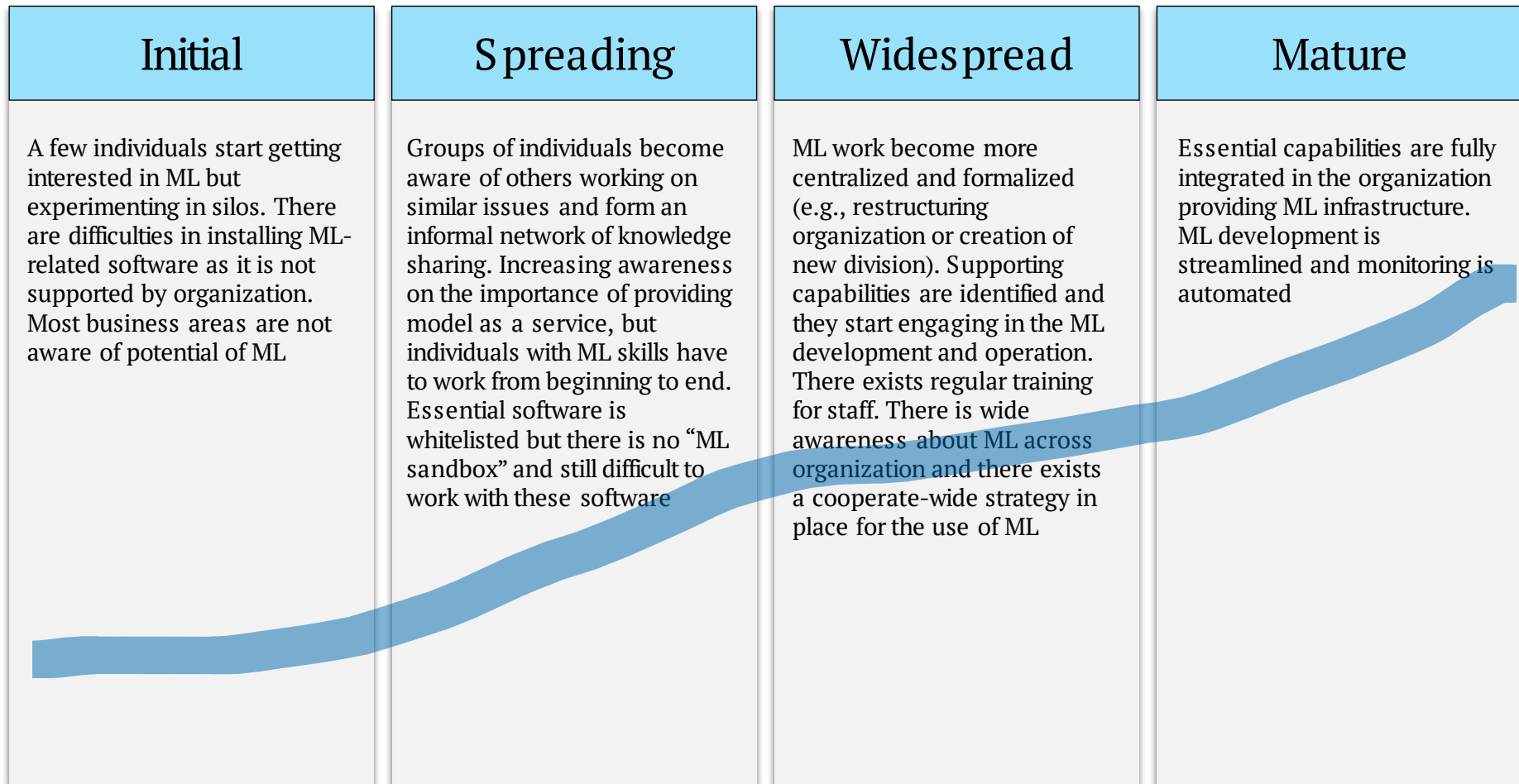
Building capacity for machine learning

"Hidden Technical Debt in Machine Learning Systems," Google NIPS 2015

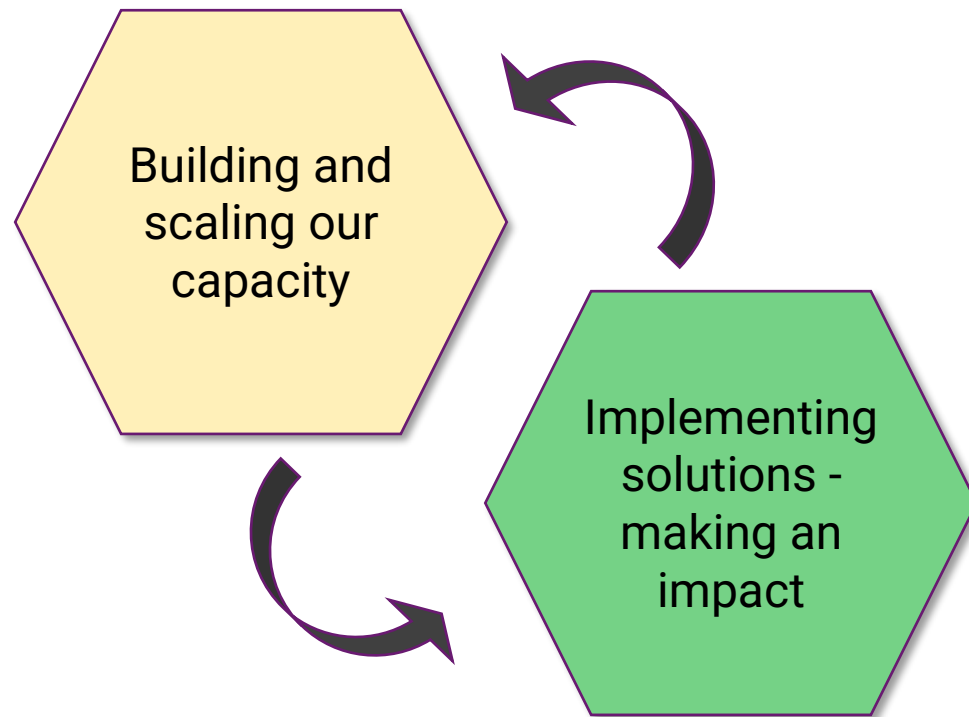


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small green box in the middle. The required surrounding infrastructure is vast and complex.

ML Maturity Model



Building and using ML capacity

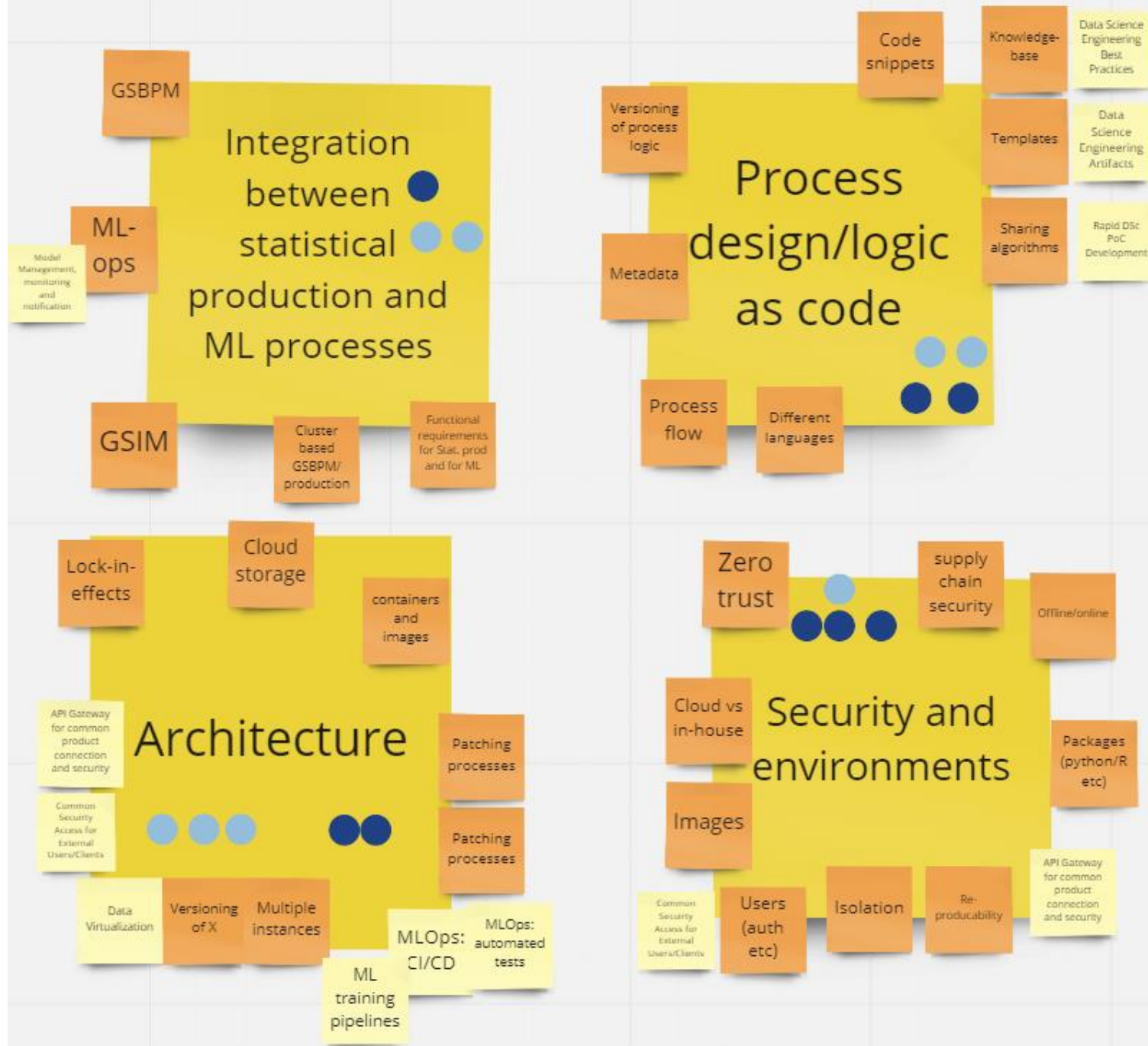


SCB



Topics for our group

- Recurring themes in ML and ML-ops
 - Architecture, Technology, Infrastructure
 - Roles and capabilities
 - Connection between statistical production and ML processes

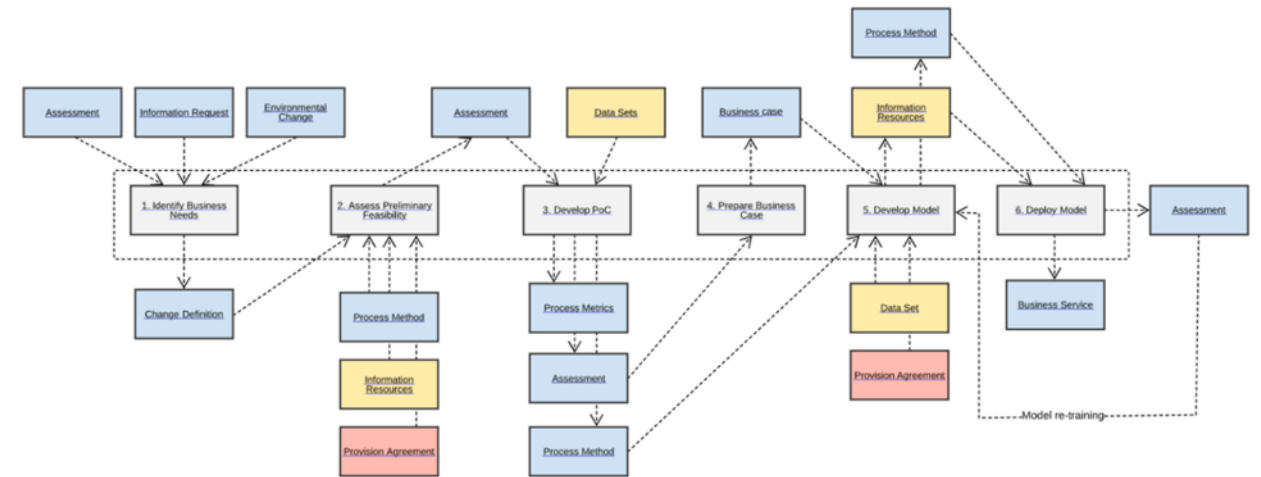


Connection between processes

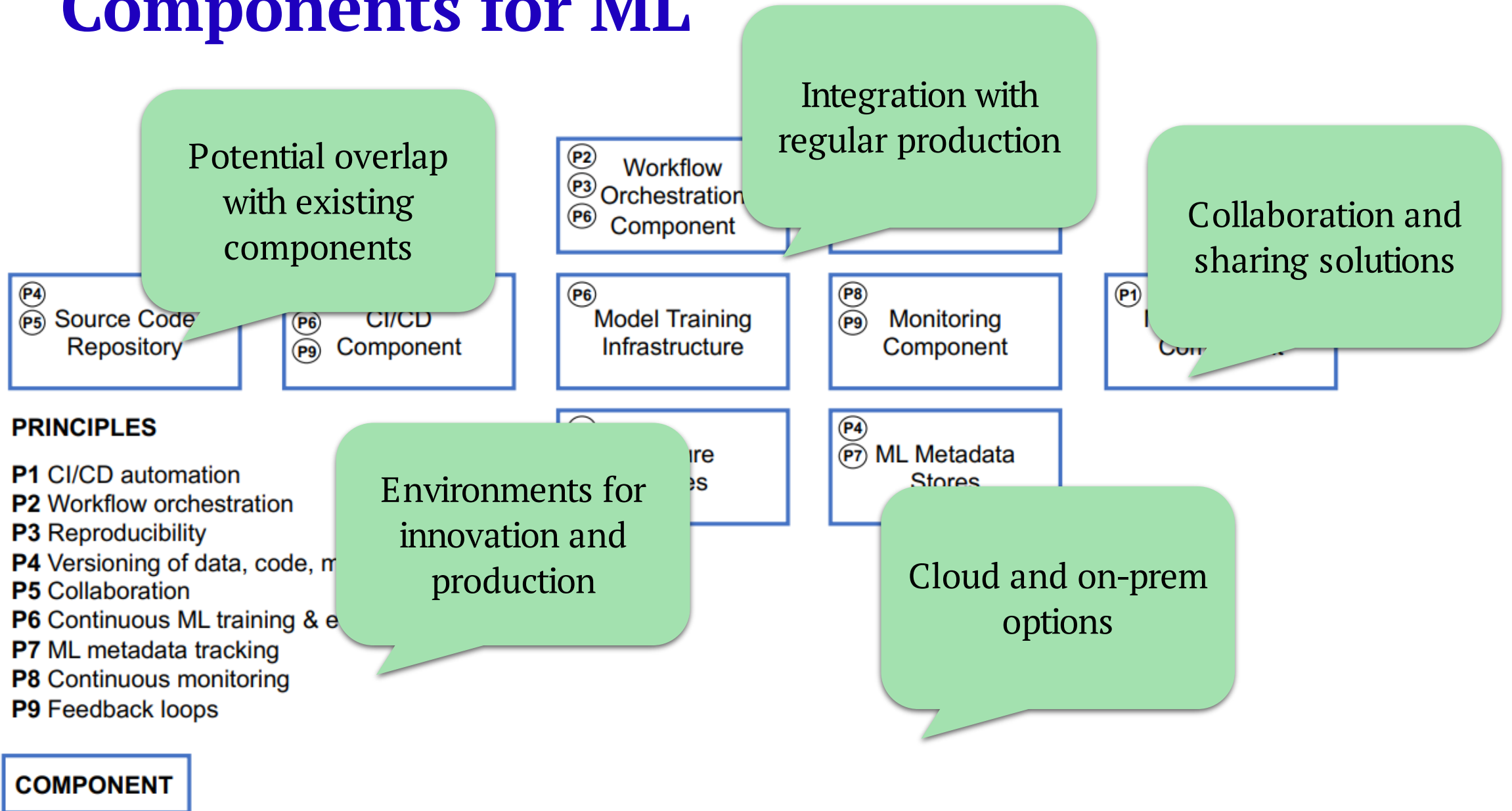
Statistical production

Overarching Processes							
Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Reuse or build collection instruments	4.1 Create frame and select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult and confirm needs	2.2 Design variable descriptions	3.2 Reuse or build processing and analysis components	4.2 Set up collection	5.2 Classify and code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design collection	3.3 Reuse or build dissemination components	4.3 Run collection	5.3 Review and validate	6.3 Interpret and explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame and sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit and impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing and analysis	3.5 Test production systems		5.5 Derive new variables and units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare and submit business case	2.6 Design production systems and workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production systems		5.7 Calculate aggregates			
				5.8 Finalise data files			

ML project/process

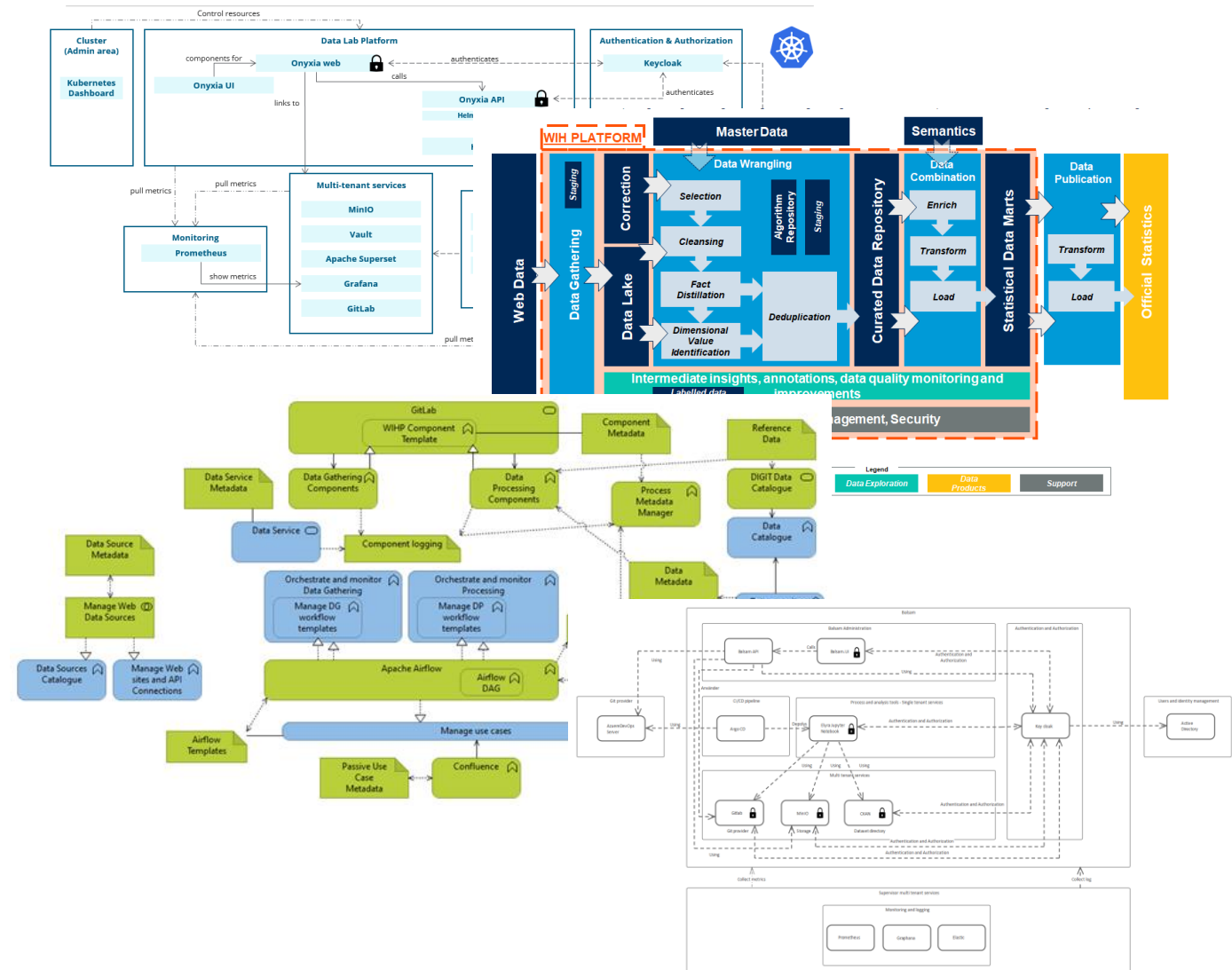


Components for ML



Collaboration around platforms for ML

- High degree of open source in ML
 - Easier to share platforms
 - Examples of platforms being developed and shared
 - Lots of similarities between shared solutions
- On-prem and Cloud
 - Use of container based solutions
- Most solutions aimed at supporting ML-processes



Different focus



How much
statisticians
care

Data

Production logic

Metadata

Versioning of data

Versioning of code

Security

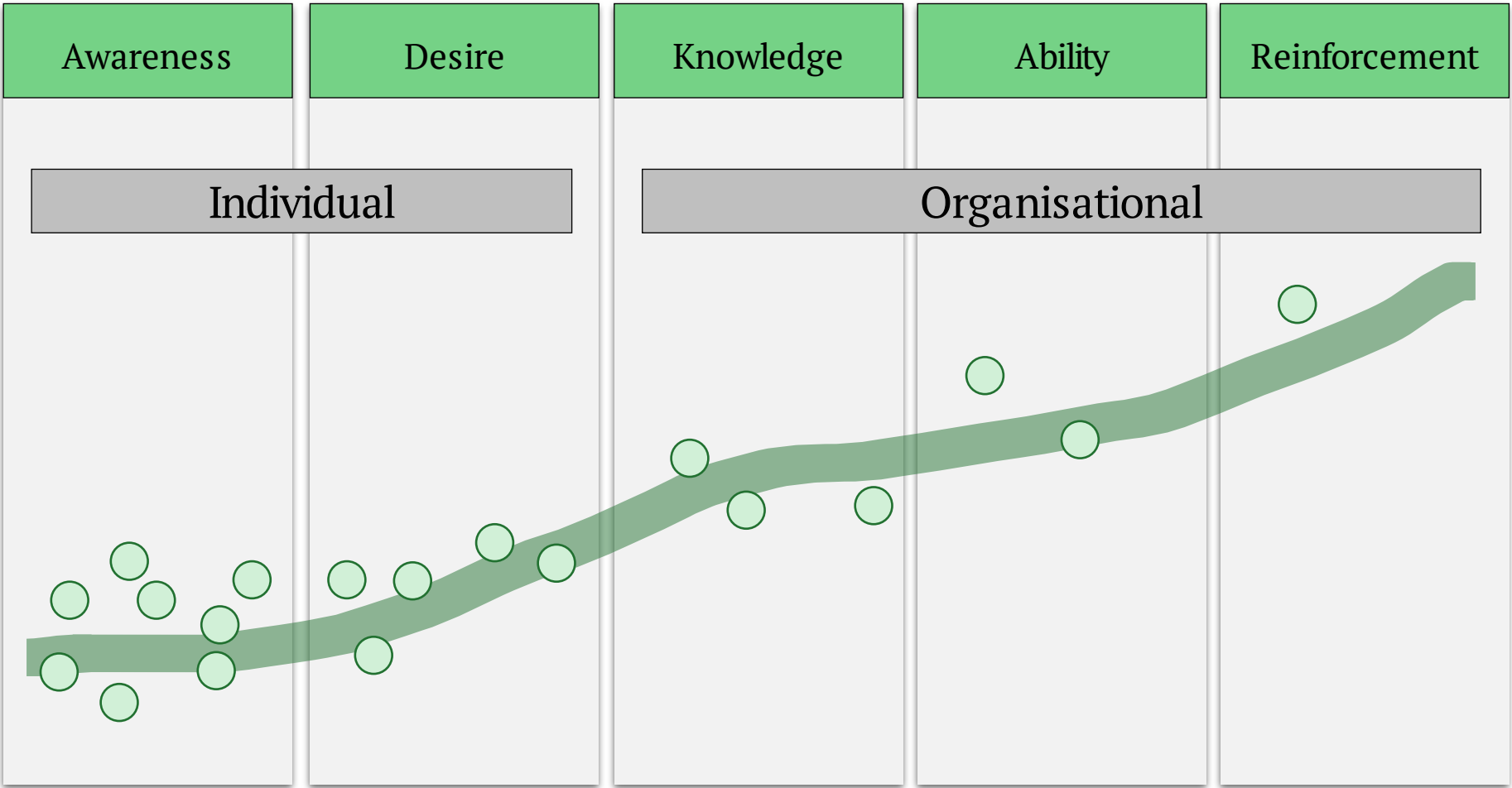
Automated production

Compute resources



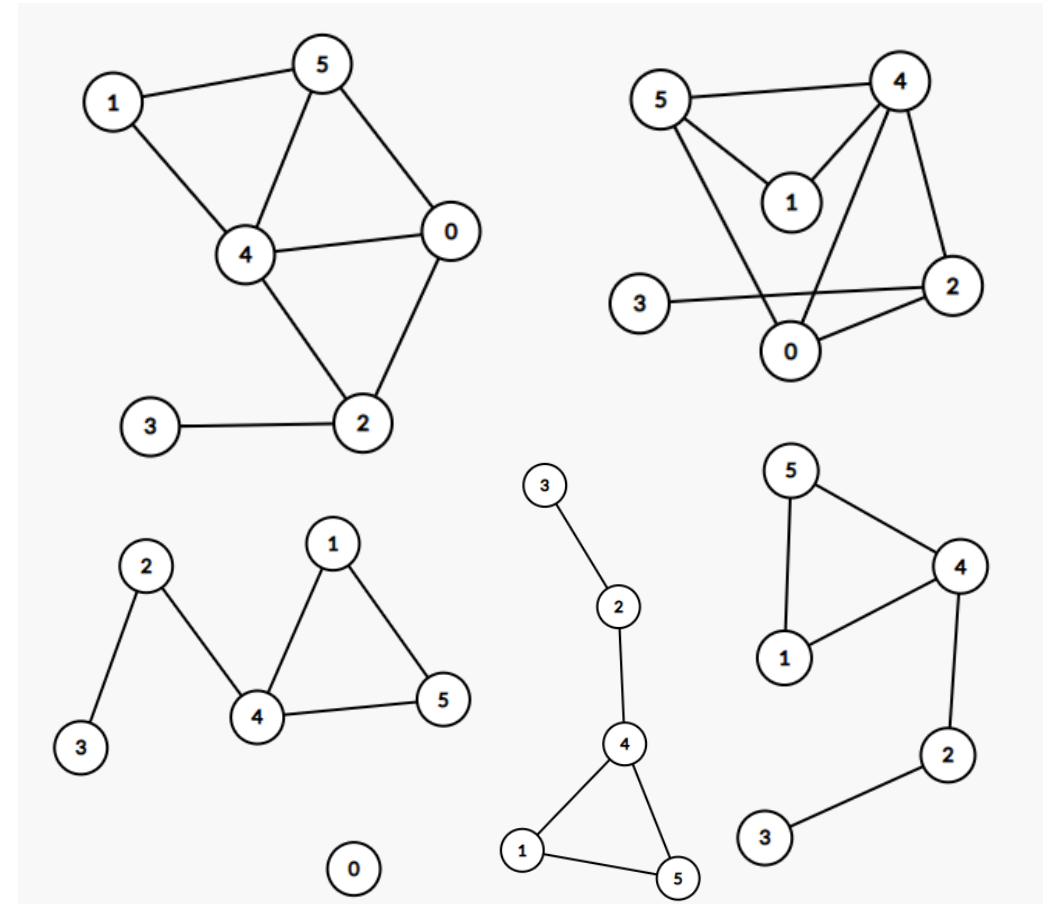
How much
IT care

Building competence



Organisational structure

- Different operating structures
 - Hub & Spoke
 - CoE
 - Communities of practice
 - Informal networks
- Deliverables differ in structure
 - ML-solutions
 - Platforms
 - Skills building capacity



Roles and capabilities

- A combination of new and existing roles
- Cross-cutting topic
 - Business, IT, Methodology plus new roles
 - New forms of collaboration between roles are explored
- Often difficult to recruit and keep

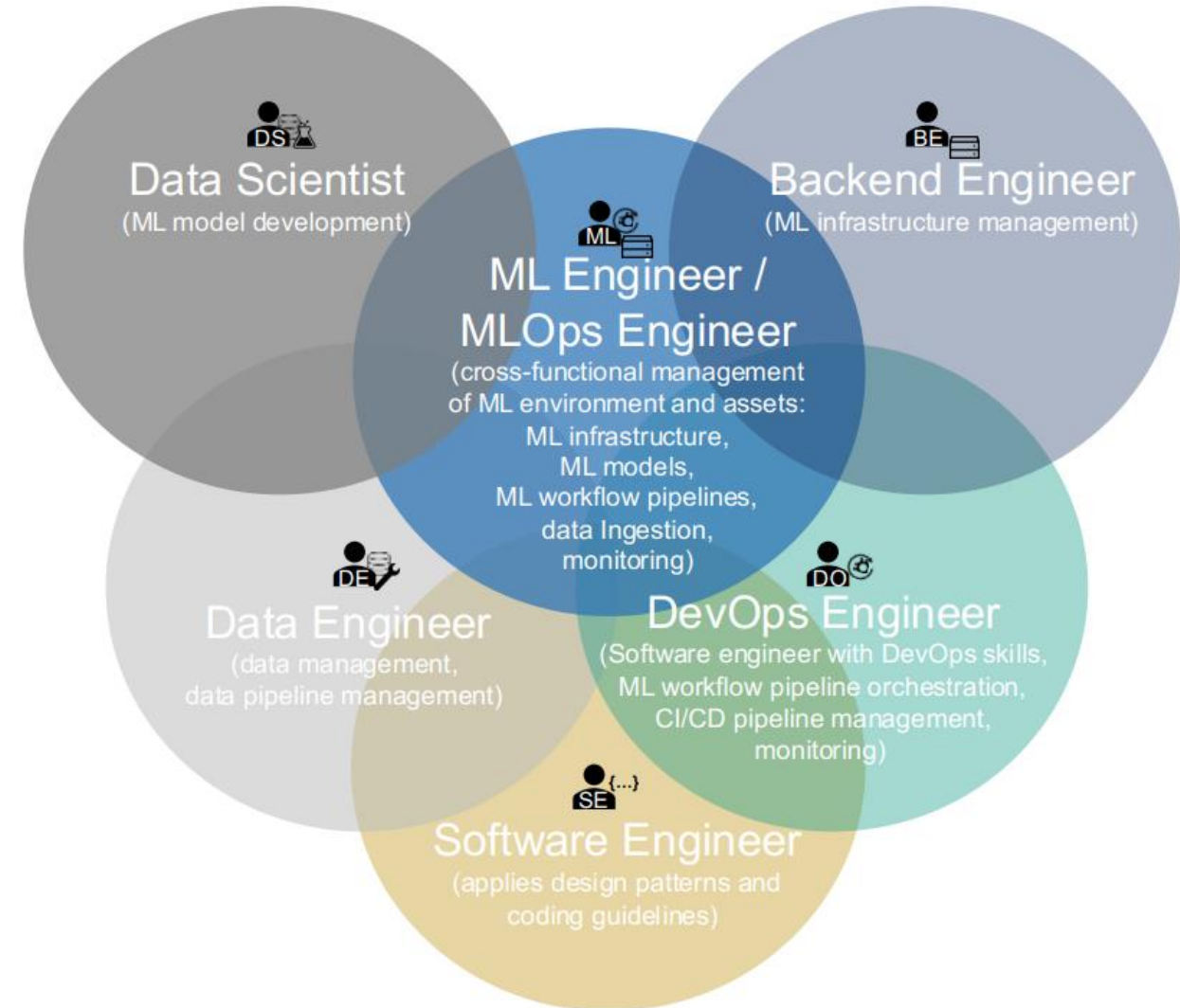


Figure 3. Roles and their intersections contributing to the MLOps paradigm

**Thank
you!**



jakob.engdahl@scb.se

