

From Complexity to Consistency: *Scalable Anonymization Across Clinical Trial Portfolios*



Danielle DiBonaventura
Bayer Corporation



Ahmed Eldafrawy
Medispend (RLS)



RESTRICTED

Introduction



Danielle DiBonaventura
Medical Writing Documentation Expert

Whippany, NJ



Ahmed Eldafrawy
Solutions Manager

Philadelphia, PA





Agenda

- Disclosure Landscape and Overview
- How Anonymization at Scale Works
- Scenario Tradeoffs & Impact
- Scaling Across Trials
- Key Benefits & Takeaways



Disclosure Landscape and Overview

Historical and Upcoming Disclosure Submissions

Approximately 405,000 pages including data for 102,000 trial subjects anonymized and disclosed to date.



Health
Canada

- 17 HC PRCI completed

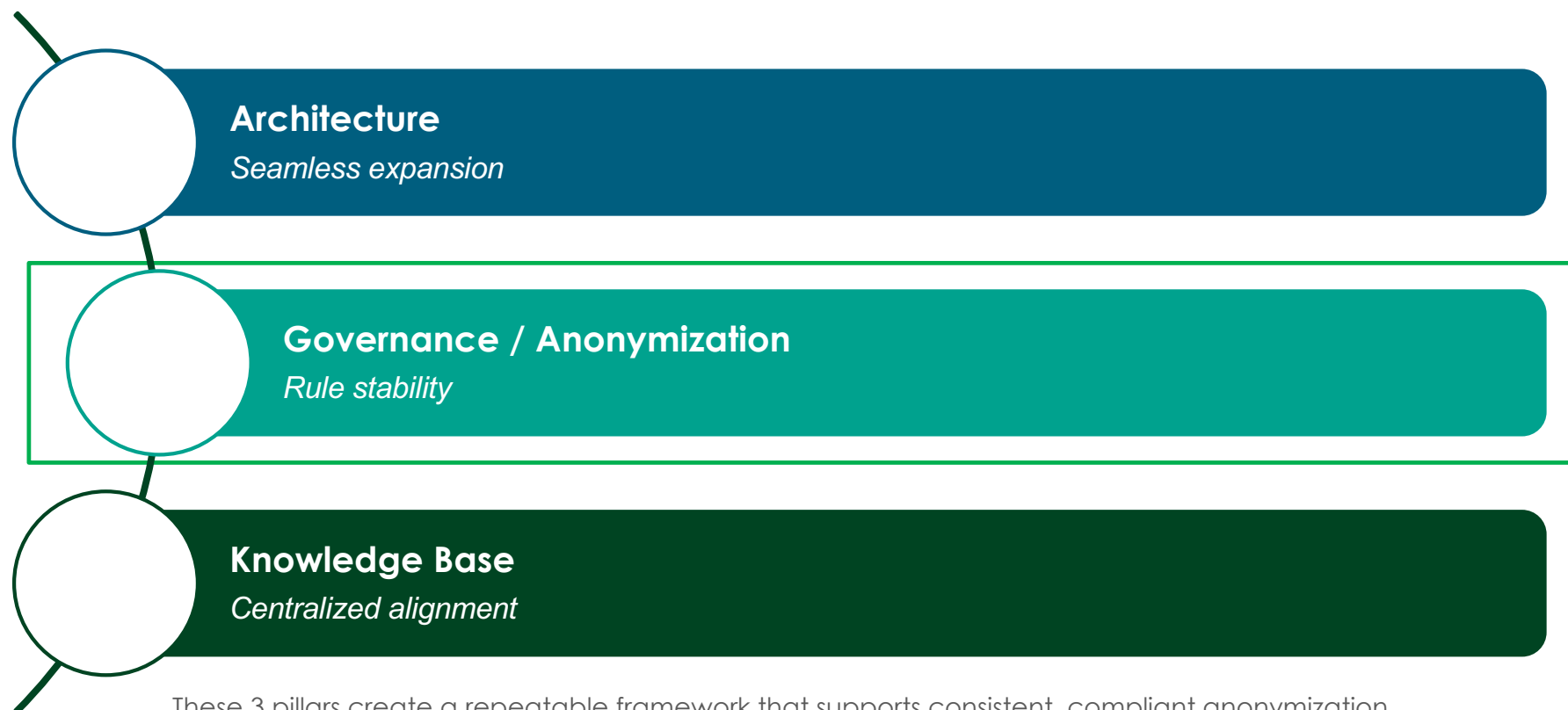


EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

- 4 CDP/Policy 0070 completed

(~3-4 submissions are completed and made
publicly available per year)

Bayer's Scalable Anonymization Framework



These 3 pillars create a repeatable framework that supports consistent, compliant anonymization across the portfolio

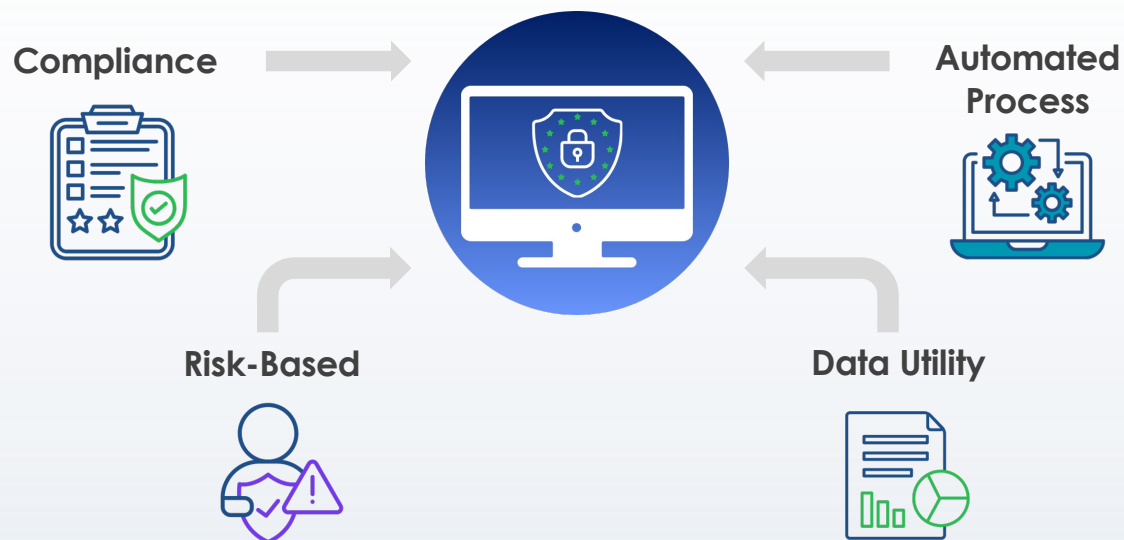


How Anonymization at Scale Works

What Is Anonymization at Scale?

Anonymization at scale is the systematic, repeatable and risk-based anonymization of trial datasets across products or multiple studies ensuring consistency, regulatory compliance and data utility.

- ✓ Lower Risk
- ✓ Compliant Sharing
- ✓ Operational
- ✓ Trusted and Auditable
- ✓ Balance Utility and Privacy
- ✓ Standardized Privacy Controls



What Is The Goal?

Quantitative anonymization applies measurable controls to preserve data utility while adhering to the risk of re-identification.

Risk of Re-Identification

- Framework and threshold defined by you
- Study & population
- Demographics
- Attributes collected



Data Utility

- Anonymization scenario with best clinical utility tailored for your product!
- Granularity of the data preserved
- Linkability/cross-referencing consistency across studies

How to Anonymize at Scale?

Risk of Re-Identification



Data

Structured clinical trial data with relevant personal data and demographics



Technology

Purpose built system/process to help automate the quantitative data analysis



Domain Expertise

Qualified personnel with knowledge of data anonymization and safe data sharing standards

Anonymization Lifecycle

Configure

- Privacy model
- Risk threshold
- Attributes
- Hierarchies
- Reference population

Evaluate

- Anonymization scenarios
- Information loss
- Data utility
- Transformation levels
- Distribution of quasi-identifiers

Assess

- Residual risk of re-identification
- Adversary attacker models
- Anonymization reporting
- Data outputs



Understanding the Context

Data Utility

Product: AXR-102

Description: AXR-102 is focused on clinical trials evaluating Vascimercept for cognitive decline in early Alzheimer's disease by reducing neuroinflammation.

What does that mean for anonymization and clinical utility? Certain data attributes should be prioritized for transparency

Prioritized Attribute	Why it Matters?
Age	Neurodegenerative diseases are age-driven
Sex	Outcomes of dose/admin differ by sex
Race/Ethnicity	Helps supports equality and unbiasedness across study
Geographic Location	Demonstrated equitable representation (if applicable)
Adverse Events	Required from a safety and benefit assessment



Scenario Tradeoffs & Impact

Evaluating the Tradeoffs

Risk Summary: Study AXR 102 – Number of Participants: 298 – Reference Population: 6,907 – Sample k: 3

All Scenarios will adhere to the following

Attributes	Scenario 1
Age	
Sex	
Race	
Ethnicity	
Country	
BMI	
Height_Weight	
Conclusion	Best overall data utility and preferred system option

Study No.		AXR 102		
Risk Threshold		9%		
Reference Population (Similar Trials + Study Population)		6,907		
Sample k		3		
Number of Participants / Study Population (inc Screen Failures)		298		
Identifier	Approach	Scenario 1	Scenario 2	Scenario 3
Participant ID	Pseudonymize	100%		
Site ID	Pseudonymize	100%		
Age	Retain	-	-	83%
	Age GRP (18-39, 40-64, 65-74, 75-84, >=85)	89%	91%	-
	Suppress	11%	9%	17%
Sex/Pronoun	Retain	89%	83%	-
	Suppress	11%	9%	100%
Race	Retain	-	83%	-
	Generalize (Caucasian / Non-caucasian)	89%	-	-
	Suppress	11%	9%	100%
Ethnicity	Retain	89%	-	83%
	Suppress	11%	100%	17%
Country	Retain	89%	-	-
	Generalize (Continent)	-	-	83%
	Suppress	11%	100%	17%
BMI	Generalize (<18.5, 18.5-24.9, 25-29.9, >=30)	-	-	83%
	Suppress	100%	100%	17%
Height	Generalize (<150, 150-160, 160-170, 170-180, >=180)	-	91%	-
	Suppress	100%	9%	100%
Weight	Generalize (<60, 60-70, 70-80, 80-90, 90-100, >=100)	-	91%	-
	Suppress	100%	9%	100%
Participant Date	Randomized Date Interval	100%	100%	100%
High Risk Participants		11% (32 records)	9% (27 records)	17% (51 records)

84, >=85

=30

TRICTED

Impact of Scenario

Risk Output from AXR 102 Data

Selected Scenario per Risk Assessment:
Scenario 1



NLP Recognition within AXR 102 CSR

Subject 01-701-1211, a 76-year-old white woman from Austria with Alzheimer's dementia, receiving xanomeline low dose on 15 November 2012. The subject also had diabetes mellitus, which was considered to be mild in severity. The subject was 166.7 cm and had a BMI of 36.



Anonymized via Risk Output

Subject 01-865-4944, a 75-84-year-old Caucasian woman from Austria with Alzheimer's dementia, receiving xanomeline low dose on Day 45. The subject also had diabetes mellitus, which was considered to be mild in severity. The subject was Height cm and had a BMI of >=30.

Reusing Anonymization Logic Across Trials

Because multiple studies on the same product tend to look alike in terms of population, data structure, and risk, anonymization doesn't need to start from the beginning each time. A consistent approach can be applied and then validated for each study.



What is reused?

- Framework and threshold
- Defining of attributes
- Anonymization approach/scenario
- Transformation logic
- Decision rationale



How is it applied to a new trial?

- Verifying trial characteristics
- Assess need to re-evaluate the risk
 - Additional participants
- Apply established logic
- Adjust generalization or suppression when necessary



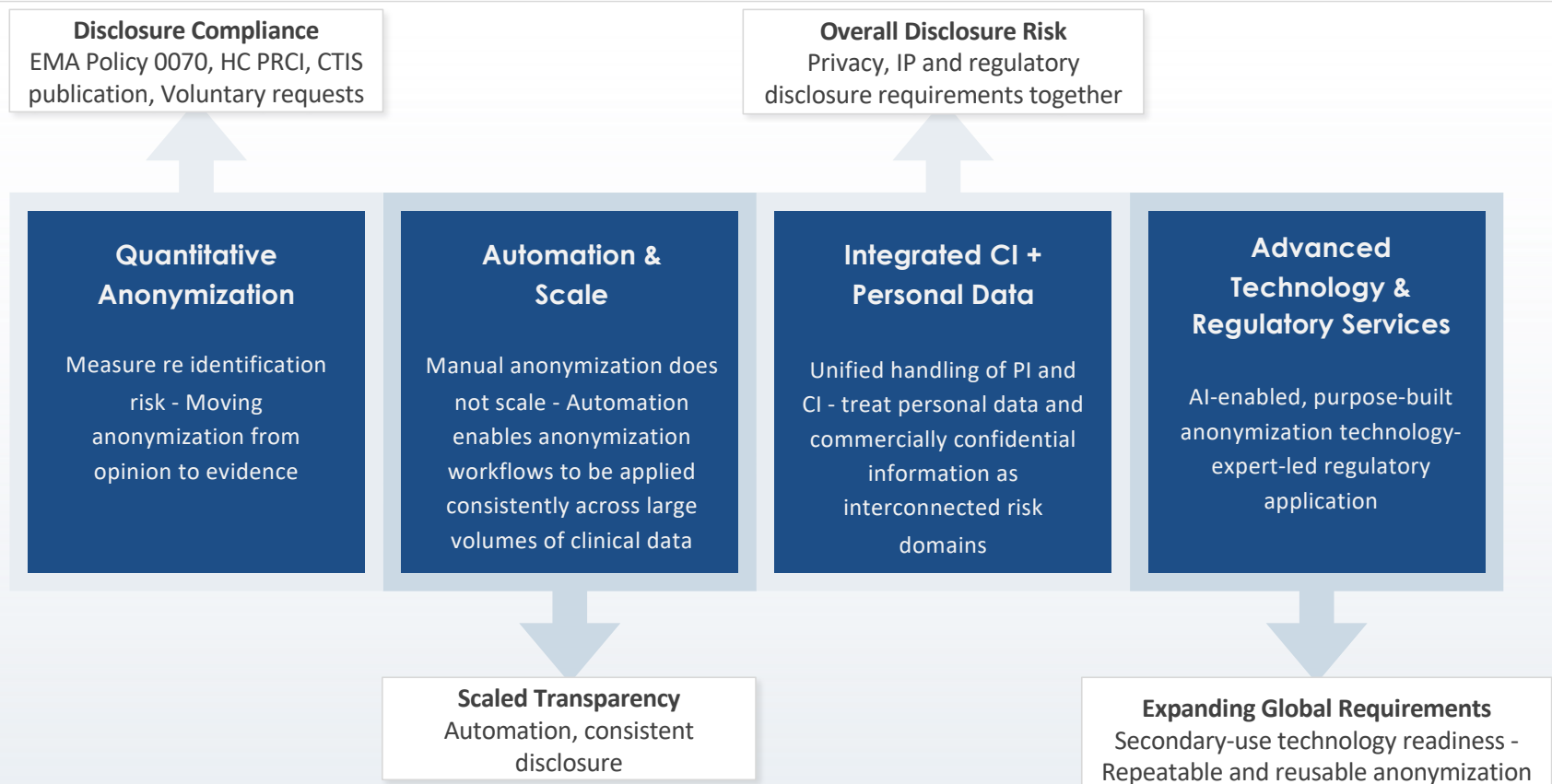
How is consistency maintained?

- Same anonymization approach
- Deviations are limited, intentional and catalogued
- Differences are driven by change in trial characteristics and output of calculation, not subjective judgement



medispend

One Anonymization Strategy





Key Benefits & Takeaways

Why Anonymization at Scale Matters

A Sponsor's Perspective



Program-Level Consistency

Prevents transformation drift, conflicting identifiers and inconsistent thresholds



Reduced Fragmentation

Without scale, submissions that are years apart can lead to inconsistent anonymization and higher privacy risk



Long-Term Strategy

One anonymization approach repeated across the full program lifecycle-enabling reliability and long-term alignment



Regulatory Predictability

A unified anonymization strategy reduces variation and discrepancies and streamlines Health Authority reviews

Scaling anonymization ensures stability, reduces risk, and builds Health Authority trust over time

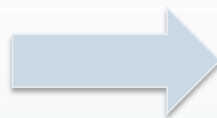
Benefits of a Scaled Anonymization Strategy

What Sponsors Gain from Scaling Anonymization

Without Scale

- ⚠ Each submission is anonymized differently
- ⚠ Anonymization is built for every package
- ⚠ Higher linkability risk across submissions
- ⚠ Will create more health authority questions

Inconsistent & Reactive



With Scale

- ✓ Stable rules
- ✓ Lower cumulative re-identification risk
- ✓ Decreased internal discussions
- ✓ Smoother HA reviews
- ✓ Predictable turnaround & time savings

Stable, Efficient & Trusted

How Anonymization Scales Across a Program



**Original Submission
(Product A):**
Program-level framework
establishes



Subsequent Submissions:
Same rules & risk model used
New Risk Assessment: 3-4 weeks
Reused Risk Assessment: 0 weeks
Modified Existing Risk Assessment: 1 week



Cross-Program Reuse:
Anonymization approach
(i.e., Prosecutor Model) used
for Product B



Outcome:
Faster cycles, fewer
inconsistencies, increased
trust with Health Authority

Impacts of Anonymization at Scale

Ensuring Quality Outputs for Future Use



Patient Centric

Consistent disclosure of trial outcomes and patient perspectives



Safe & Secure

Focused on participant privacy with GDPR compliant anonymization methods and rules



Transparency

Reduced variability across trial outputs



Efficiency

Faster education with ongoing trials & lower rework for everyone



Value to Submission

Defensible & regulator-preferred approach

Health for all, Hunger for none



medispense®

bayer.com

Thank you!



Danielle DiBonaventura
*Medical Writing
Documentation Expert*



Ahmed Eldafrawy
Solutions Manager



RESTRICTED