



Funded by the  
Nordic Council  
of Ministers

# **Lessons Learned from Multi-Site Federated Analysis in VALO NSCLC pilot study**

Report 2.0

Value from Nordic health data – VALO project

9 December 2025





## Document info

### 0.1 Authors

Author	Partner
Atif Adam, IQVIA Senior Epidemiologist	IQVIA
<b>Collaborator / Buyer Representative</b>	
Markus Kalliola	Finnish Innovation Fund Sitra
<b>Principal Investigator</b>	
Åslaug Helland, Main Principal Investigator	Oslo University Hospital
<b>IQVIA project responsible parties</b>	
Susanna Flaherty, Pilot Project Director	IQVIA
Elina Winblad, Pilot Project Manager	IQVIA
James Brash, Pilot Senior Data Scientist	IQVIA

### 0.2 Keywords

Keywords	Health data, Nordic, EHDS
----------	---------------------------

### Disclaimer

The content of this report represents the views of the author(s) and do not necessarily reflect the views of the experts interviewed.

### Copyright Notice

Copyright © 2025 Sitra and the Nordic Council of Ministers. All rights reserved. For more information on the project, please see the [project website](#).



## Contents

<b>Executive Summary</b> .....	<b>5</b>
<b>Tiivistelmä</b> .....	<b>6</b>
<b>Sammanfattning</b> .....	<b>7</b>
<b>1 Abbreviations</b> .....	<b>8</b>
<b>2 Introduction</b> .....	<b>9</b>
2.1 Study Overview .....	9
2.2 Key Findings .....	9
2.2.1 Implementation Timeline and Complexity .....	9
2.2.2 Critical Data Gaps and Readiness Hierarchy.....	10
2.3 Principal Recommendations .....	10
2.3.1 Critical Success Factors.....	10
2.3.2 Risk Mitigation.....	10
<b>3 Introduction</b> .....	<b>11</b>
3.1 Background and Rationale.....	11
3.2 Objectives of Lessons Learned Exercise .....	11
3.3 Scope and Limitations .....	12
3.3.1 Scope .....	12
3.3.2 Data Sources .....	12
3.3.3 Limitations .....	12
<b>4 Methods</b> .....	<b>13</b>
4.1 Survey Design and Development.....	13
4.2 Data Collection Process.....	13
4.3 Respondent Characteristics .....	13
4.4 Data Analysis Approach.....	13
4.4.1 Quantitative Analysis.....	13
4.4.2 Qualitative Analysis.....	13
4.4.3 Cross-Site Comparison .....	14
4.5 Methodological Considerations .....	14
4.5.1 Response Interpretation.....	14
<b>5 Results (Enhanced and Comprehensive)</b> .....	<b>15</b>
5.1 Background Information .....	15
5.2 Data Governance .....	15
5.3 Data Life Cycle .....	15
5.3.1 Data Availability Matrix Analysis (3, 4).....	15
5.3.2 Coding Systems .....	17
5.4 Data Processing.....	17
5.4.1 Variable Capture Completeness .....	17
5.4.2 Implementation Challenge Gradients.....	18
5.5 VALO Pilot Study Process Steps .....	19
5.6 Future Investments and Improvements.....	19
5.6.1 Summary of Investment Priorities .....	20
5.7 IQVIA Execution Experience.....	20
5.7.1 Project Management and Coordination Experience.....	21
5.7.2 Lessons for Future Network Project Management.....	22



Lessons learned from multi-site federated analysis – VALO Value from Nordic Health Data 3

5.7.3 Patient Cohort Assembly Reality.....	23
5.7.4 Technical Execution Challenges.....	23
5.7.5 Successful Adaptations and Innovations.....	24
5.7.6 Genomic Vocabulary Implementation.....	24
<b>6 Discussion.....</b>	<b>25</b>
6.1 Principal Findings and Interpretation.....	25
6.2 Data Readiness and Systematic Gaps.....	26
6.3 Governance and Regulatory Uncertainty.....	27
6.4 The Learning Curve and Capability Development.....	28
6.5 Network Coordination Challenges.....	28
6.6 From Descriptive to Inferential Analytics.....	29
6.7 Infrastructure Optimization as Success Factor.....	29
6.8 Implications for Future Network Studies.....	30
6.8.1 Feasibility Assessment Must Be Data-Tier Specific.....	30
6.8.2 Front-Load Design Investment.....	30
6.8.3 Bridge Technical and Governance Domains.....	30
6.8.4 Plan for Three-Year Capability Development.....	30
6.9 Strengths and Limitations of This Assessment.....	30
6.10 Future Directions.....	30
<b>7 Recommendations.....</b>	<b>32</b>
7.1 Technical Infrastructure Recommendations.....	32
7.1.1 Pre-Implementation Infrastructure Assessment.....	32
7.1.2 Infrastructure Standardization.....	32
7.2 Governance Framework Recommendations.....	32
7.2.1 Integrated Governance Structure.....	32
7.2.2 Regulatory Pathway Clarification.....	33
7.3 Data Readiness and Quality Recommendations.....	33
7.3.1 Tiered Feasibility Assessment Protocol.....	33
7.3.2 Two-Phase Study Design Framework.....	35
7.3.3 Strategic Data Enhancement Initiatives.....	35
7.4 Capacity Building Recommendations.....	36
7.4.1 Comprehensive Three-Year Capability Development Framework.....	36
7.4.2 Comprehensive Knowledge Management Systems.....	38
7.5 Network Coordination Recommendations.....	39
7.5.1 Comprehensive Vocabulary Governance Framework.....	39
7.5.2 Enhanced Communication and Collaboration Structures.....	40
7.6 Methodological Advancement Recommendations.....	41
7.6.1 Comprehensive Framework for Advanced Federated Analytics.....	41
7.6.2 Scientific Infrastructure Development.....	43
7.7 Implementation Prioritization Strategy.....	44
7.7.1 Immediate Priorities (0-3 months).....	44
7.7.2 Short-term Priorities (3-12 months).....	44
7.7.3 Long-term Investments (12+ months).....	45
<b>8 Conclusion.....</b>	<b>47</b>
8.1 Synthesis of Key Findings.....	47
8.2 Transformative Insights for Network Development.....	47



Lessons learned from multi-site federated analysis – VALO Value from Nordic Health Data 4

8.2.1 Strategic Implications for Healthcare Research .....	47
8.2.2 Contributions to the Field .....	48
8.2.3 Limitations and Future Directions.....	48
8.2.4 The Path Forward .....	49
8.3 Final Reflections.....	49
<b>9 References.....</b>	<b>51</b>
<b>Appendices.....</b>	<b>52</b>
9.1 Appendix A: Survey Structure and Content .....	52
9.2 Appendix B: Comprehensive Respondent Characteristics (Questions 1-5).....	52
9.3 Appendix C: Challenge Rating Standardization .....	53
9.4 Appendix D: Ethics and Regulatory Approval Landscape (Questions 7-9).....	53
9.5 Appendix E: Data Source Architecture by Country (Question 14) .....	54
9.5.1 Data Linkage Infrastructure Assessment (Questions 15-16) .....	54
9.5.2 Data Availability Matrix Analysis (Question 17).....	55
9.5.3 Clinical Variable Capture Success (Question 21) .....	56
9.6 Appendix F1: Implementation Challenge Gradients (Question 23).....	57
9.7 Appendix F2: VALO Study Process Steps (Questions 25-32) .....	58
9.7.1 F2.1 Design of OMOP ETL Specification (Question 25).....	58
9.7.2 F2.2 Implementation of OMOP ETL Specification (Question 26).....	58
9.7.3 F2.3 QA/QC of OMOP Instance (Question 27) .....	59
9.7.4 F2.4 Study Variable Survey and Package Execution (Questions 28-32).....	59
9.8 Appendix G: Future Investments and Improvements (Questions 33-36) .....	59
9.8.1 G.1 Data Infrastructure Requirements (Question 33) .....	59
9.8.2 G.2 Data Conversion Needs (Question 34) .....	60
9.8.3 G.3 Current OMOP Capabilities Assessment (Question 35) .....	60
9.8.4 G.4 Recommendations for Future Improvements (Question 36) .....	60
9.9 Appendix H: IQVIA Execution Experience Details .....	61
9.9.1 H.1 Communication Strategy Evolution .....	61
9.9.2 H.2 Stakeholder Management Requirements .....	61
9.9.3 H.3 Resource Coordination Challenges.....	61
9.9.4 H.4 Lessons for Future Network Project Management .....	62



## Executive Summary

The VALO Nordic OMOP Implementation Lessons Learned exercise represents the first systematic assessment of OMOP CDM implementation experiences across multiple Nordic healthcare systems.

Conducted in September 2025 following the VALO NSCLC pilot study, this assessment captured experiences from ten respondents across five countries: Norway, Finland, and Denmark as active data partners, and Sweden and Iceland as observers.

The assessment employed a structured survey instrument with six thematic sections, complemented by operational insights from the IQVIA coordinating team. Response rates ranged from 69% (Finland) to 95% (Sweden). This mixed-methods approach combined quantitative challenge assessments with qualitative experiential narratives to document technical challenges, data availability gaps, governance requirements, and resource timelines for future network implementations.

### Key Takeaways:

- OMOP implementation is a multi-year organizational transformation, not a short-term project (Norway reports ~3 years)
- Data readiness follows three tiers:
  - Tier 1 (diagnoses, medications) is OMOP-ready
  - Tier 2 (labs) requires mapping
  - Tier 3 (biomarkers, PROs) is absent
- Technical infrastructure compatibility—not OMOP conceptual understanding—is the primary implementation challenge
- Final analytical cohorts represent only 10-20% of initial population estimates after applying eligibility criteria
- Governance knowledge is fragmented across roles; cross-functional teams are essential from project inception
- Vocabulary versioning requires network-level coordination to ensure semantic consistency across federated sites
- Timeline estimates should be multiplied by 1.5-2.0× to account for multi-site coordination complexity



## Tiivistelmä

VALO Nordic OMOP -toteutuksen kokemuksista saatujen oppien kooste edustaa ensimmäistä systemaattista arviota OMOP CDM -mallin käyttöönoton kokemuksista useissa pohjoismaisissa terveydenhuoltojärjestelmissä.

Syyskuussa 2025 VALO NSCLC -pilottitutkimuksen jälkeen toteutettu arviointi keräsi kokemuksia kymmeneltä vastaajalta viidestä maasta: Norjasta, Suomesta ja Tanskasta aktiivisina datapartnereina sekä Ruotsista ja Islannista tarkkailijoina.

Arvioinnissa käytettiin rakenteista kyselylomaketta, jossa oli kuusi teemakokonaisuutta, ja sitä täydennettiin IQVIAN koordinaatiotiimin operatiivisilla havainnoilla. Vastausprosentit vaihtelivat 69 %:sta (Suomi) 95 %:iin (Ruotsi). Tämä monimenetelmällinen lähestymistapa yhdisti määrällisen haastearvioinnin ja laadulliset kokemukset teknisten haasteiden, data saatavuusaukkojen, hallinnollisten vaatimusten ja resurssitarpeiden dokumentoimiseksi tulevia verkostototeutuksia varten.

### Keskeiset havainnot:

- OMOP-toteutus on useita vuosia kestävä organisaatiomuutos, ei lyhytaikainen projekti (Norja raportoi tämän vaatineen noin 3 vuotta).
- Datan valmiustaso jakautuu kolmeen tasoon:
  - Taso 1 (diagnoosit, lääkitykset) on OMOP-valmis
  - Taso 2 (laboratoriotulokset) vaatii kartoitusta
  - Taso 3 (biomarkkerit, PRO-mittarit) puuttuu
- Teknisen infrastruktuurin yhteensopivuus – ei OMOPin käsitteellinen ymmärrys – on keskeisin haaste.
- Lopulliset analyysikohortit muodostavat vain 10–20 % alkuperäisestä populaatioarviosta kelpoisuusehtojen soveltamisen jälkeen.
- Hallinnollinen osaaminen on pirstaloitunut eri rooleihin; poikkitoiminnalliset tiimit ovat välttämättömiä projektin alusta lähtien.
- Sanastojen versioinnin koordinointi tulee tehdä verkostotasolla, jotta semanttinen yhdenmukaisuus säilyy hajautetuissa toteutuksissa.
- Aikatauluarvioihin tulisi lisätä 1,5–2-kertainen kerroin, jotta monipaikkaisen koordinaation monimutkaisuus huomioidaan.



## Sammanfattning

VALO Nordic OMOP Implementation Lessons Learned-utvärderingen representerar den första systematiska bedömningen av erfarenheter av OMOP CDM -implementering inom flera nordiska hälsosystem.

Genomförd i september 2025 efter VALO NSCLC-pilotstudien samlade denna utvärdering in erfarenheter från tio respondenter i fem länder: Norge, Finland och Danmark som aktiva datapartner samt Sverige och Island som observatörer.

Utvärderingen använde ett strukturerat enkätinstrument med sex tematiska avsnitt, kompletterat med operativa insikter från IQVIAs koordinationsteam. Svarsfrekvenserna varierade från 69 % (Finland) till 95 % (Sverige). Denna mixed-methods-ansats kombinerade kvantitativa bedömningar av utmaningar med kvalitativa erfarenhetsbeskrivningar för att dokumentera tekniska hinder, brister i datatillgänglighet, styrningskrav och resursbehov inför framtida nätverksimplementeringar.

### Centrala slutsatser:

- OMOP-implementering är en flerårig organisatorisk omställning, inte ett kortsiktigt projekt (Norge rapporterar cirka 3 år).
- Datamognad följer tre nivåer:
  - Nivå 1 (diagnoser, läkemedel) är OMOP-redo
  - Nivå 2 (laboratorievärden) kräver mappning
  - Nivå 3 (biomarkörer, PRO-mått) saknas
- Teknisk infrastrukturkompatibilitet – inte konceptuell förståelse av OMOP – är den främsta implementeringsutmaningen.
- Slutliga analytiska kohorter utgör endast 10–20 % av den ursprungliga populationsuppskattningen efter att inklusionskriterier har tillämpats.
- Kunskap om styrningsfrågor är fragmenterad mellan olika roller; tvärfunktionella team är avgörande redan från projektstart.
- Versionering av vokabulärer kräver koordination på nätverksnivå för att säkerställa semantisk konsekvens mellan federerade miljöer.
- Tidsuppskattningar bör multipliceras med 1,5–2,0× för att ta hänsyn till komplexiteten i koordination mellan flera platser.



## 1 Abbreviations

Abbreviation	Full term
BMI	Body Mass Index
CDM	Common Data Model
DARWIN EU	Data Analysis and Real-World Interrogation Network in the European Union
ECOG	Eastern Cooperative Oncology Group
EHDEN	European Health Data & Evidence Network
EHDS	European Health Data Space
EHR	Electronic Health Record
EMR	Electronic Medical Record
ETL	Extract, Transform, Load
EU	European Union
GLMM	Generalized Linear Mixed Model
HCPCS	Healthcare Common Procedure Coding System
HUS	Helsinki University Hospital
ICD-10	International Classification of Diseases, 10th Revision
ICI	Immune Checkpoint Inhibitor
KI	Karolinska Institutet
KUH	Karolinska University Hospital
LOINC	Logical Observation Identifiers Names and Codes
mNSCLC	Metastatic Non-Small Cell Lung Cancer
NA	Not Applicable
NSCLC	Non-Small Cell Lung Cancer
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
PM	Project Manager
PRO	Patient-Reported Outcome
PS	Performance Status
PD-L1	Programmed Cell Death Ligand 1
QA/QC	Quality Assurance/Quality Control
QC	Quality Control
RACI	Responsible, Accountable, Consulted, and Informed
ROI	Return on Investment
RWE	Real-World Evidence
SNOMED	Systematized Nomenclature of Medicine
SNOMED CT	Systematized Nomenclature of Medicine Clinical Terms
VALO	Value from Nordic Health Data

## 2 Introduction

### 2.1 Study Overview

The VALO Nordic OMOP Implementation Lessons Learned exercise represents the first systematic assessment of OMOP CDM implementation experiences across multiple Nordic healthcare systems. Conducted in September 2025 following the VALO NSCLC pilot study, this assessment captured experiences from ten respondents across five countries: Norway, Finland, and Denmark as active data partners, and Sweden and Iceland as observers. The assessment employed a structured survey instrument with six thematic sections, complemented by operational insights from the IQVIA coordinating team. Response rates ranged from 69% (Finland) to 95% (Sweden-KUH). This mixed-methods approach combined quantitative challenge assessments with qualitative experiential narratives to document technical challenges, data availability gaps, governance requirements, and resource timelines for future network implementations.

### 2.2 Key Findings

The assessment revealed six principal findings:

1. Implementation challenges follow a clear gradient: ETL design is moderately challenging, implementation slightly challenging, and QA/QC becomes routine with OHDSI tools.
2. Data readiness hierarchy: Tier 1 (diagnoses, medications) shows >60% OMOP readiness; Tier 2 (labs, procedures) requires mapping; Tier 3 (biomarkers, PROs) is systematically absent.
3. Technical infrastructure compatibility—not OMOP conceptual understanding—represents the primary implementation challenge; Norway's documentation of solutions creates reusable knowledge.
4. Governance knowledge is fragmented across organizational roles, requiring cross-functional teams from project inception.
5. Vocabulary versioning requires consortium-level coordination; sites independently updating vocabularies creates potential semantic drift.
6. IQVIA operational team documented that final analytical cohorts represented 10-20% of initial population estimates; federated analyses completed within three months post-infrastructure operationalization.

#### 2.2.1 Implementation Timeline and Complexity

The assessment revealed that OMOP implementation represents a multi-year organizational capability development journey (Norway's experience suggests approximately three years, though timelines vary based on baseline infrastructure and institutional capacity) rather than a short-term technical project (1, 2). Norway's empirical report that "OMOP skills have been established in the data warehouse team over the last 3 years" provides the first quantified learning curve in the literature. Implementation



challenges follow a clear gradient: Extract, Transform, Load (ETL) design emerged as "moderately challenging" requiring significant effort, implementation proved "slightly challenging" with minor adjustments needed, while Quality Assurance/Quality Control (QA/QC) became "routine process" facilitated by Observational Health Data Sciences and Informatics (OHDSI) tools. This progression demonstrates that front-loaded investment in design yields downstream efficiencies, as confirmed by Norway's successful reuse of previous work.

## 2.2.2 Critical Data Gaps and Readiness Hierarchy

For purposes of this assessment, we developed a study-specific three-tier data readiness classification (detailed in 6 Discussion). A systematic three-tier data readiness hierarchy emerged from the assessment. Tier 1 data (diagnoses, prescribed medications, demographics) showed >60% OMOP readiness with universal availability. Tier 2 data (laboratory results, procedures, administered drugs) existed in source systems but required substantial mapping effort, with laboratory results showing 0% OMOP readiness despite universal availability. Tier 3 data (biomarkers, performance status, patient-reported outcomes [PROs]) were systematically absent, with all Norwegian respondents reporting biomarker absence and universal PRO unavailability. These gaps fundamentally constrain feasibility for modern oncology research and cannot be resolved through technical mapping alone.

## 2.3 Principal Recommendations

The following recommendations are organized into three time horizons: **Immediate Actions (0-3 months)**, **Short-term Priorities (3-12 months)**, and **Strategic Investments (12+ months)**. They are derived from survey findings and operational experience, and are applicable to data-providing sites, coordinating centres, and network governance bodies. Critical success factors and risk mitigation strategies are also outlined.

### 2.3.1 Critical Success Factors

Success requires simultaneous advancement across six dimensions: (1) Technical infrastructure properly configured for OMOP requirements, not just available infrastructure; (2) Organizational commitment to three-year capability development, not short-term project thinking; (3) Cross-functional integration bridging technical, clinical, and regulatory domains; (4) Network coordination ensuring semantic consistency and knowledge sharing; (5) Methodological innovation enabling causal inference in federated settings; (6) Clinical engagement improving documentation completeness for Tier 3 data.

### 2.3.2 Risk Mitigation

Key risks identified include continued reliance on descriptive statistics limiting scientific impact; perpetuation of Tier 3 data gaps constraining modern clinical research; vocabulary divergence undermining network validity; loss of experiential knowledge through inadequate documentation; and unrealistic timeline expectations causing stakeholder dissatisfaction. These risks require proactive management through the recommendations provided.



## 3 Introduction

### 3.1 Background and Rationale

This document presents a systematic evaluation of the VALO NSCLC pilot study implementation methodology, distinct from the clinical and epidemiological findings of the primary VALO investigation. Whereas the VALO study examined treatment patterns, survival outcomes, and real-world effectiveness of immune checkpoint inhibitors in patients with metastatic non-small cell lung cancer, the present lessons learned analysis focuses exclusively on the operational, technical, and governance challenges encountered during federated OMOP Common Data Model implementation across Nordic healthcare infrastructures, within VALO pilot. The overall objective is to provide evidence-based guidance for future multi-site OMOP implementations rather than contribute to the clinical understanding of mNSCLC treatment.

The implementation of OMOP CDM across diverse Nordic healthcare infrastructures presented unique opportunities and challenges. While Nordic countries share similar universal healthcare principles and have robust health data systems, each nation maintains distinct data governance frameworks, technical infrastructures, and clinical documentation practices. The harmonization of these heterogeneous systems into a common data model for federated analysis required substantial coordination and technical innovation.

This lessons learned exercise was initiated to systematically capture implementation experiences, technical challenges, and operational insights from all participating institutions. The documentation of these experiences is critical for several reasons: (1) to inform future Nordic collaborative studies utilizing OMOP CDM, (2) to guide new sites considering OMOP implementation, (3) to identify common challenges requiring network-level solutions, and (4) to establish best practices for federated RWE studies in the Nordic context.

### 3.2 Objectives of Lessons Learned Exercise

The primary objective of this exercise was to comprehensively document the implementation experiences across all VALO participating sites through structured feedback collection and analysis.

Specific objectives included:

1. To assess the technical challenges encountered during OMOP CDM implementation and study execution across different phases: ETL development, quality assurance, and analytical package deployment.
2. To evaluate data governance processes, including ethical approvals, data access agreements, and compliance with national and EU regulations across participating countries.
3. To identify data availability gaps and infrastructure limitations that affected study objectives and analytical capabilities.
4. To document resource requirements, including personnel expertise, time investments, and technical infrastructure needed for successful OMOP implementation.
5. To gather forward-looking recommendations from sites regarding future investments, capability development, and network coordination improvements.



Funded by the  
Nordic Council  
of Ministers

Lessons learned from multi-site federated analysis – VALO Value from Nordic Health Data 12

6. To establish a knowledge base that can accelerate future OMOP implementations and reduce redundant efforts across the Nordic network.

### 3.3 Scope and Limitations

#### 3.3.1 Scope

This assessment encompasses feedback from ten respondents representing five Nordic countries (detailed in [Section 2](#)). Three active data-contributing sites (Norway, Finland, Denmark) completed OMOP CDM implementation and provided patient-level data for federated analysis. Two observer sites (Sweden, Iceland) participated in consortium meetings and contributed implementation planning perspectives without executing OMOP transformation.

This assessment encompasses feedback from ten respondents representing five Nordic countries participating in or observing the VALO pilot. All five sites participated in VALO Pilot Consortium, which was led by IQVIA project manager.

The analysis covers the complete project lifecycle from initial contracting through final results delivery, spanning from November 2024 until the end of October 2025.

#### 3.3.2 Data Sources

Primary data sources include: (1) Structured Survey Instrument administered September 2025, comprising 36 questions across six thematic domains, capturing both quantitative assessments and qualitative experiential narratives; and (2) IQVIA Operational Team Documentation of execution experiences including feasibility assessment refinements, technical challenges, site coordination, and timeline realities. This dual-source approach enables comprehensive assessment from both strategic readiness and tactical implementation viewpoints.

#### 3.3.3 Limitations

Several limitations should be considered: (1) Response Variability—completion rates ranged from 69% to 95% across sites, with missing responses potentially reflecting questions outside respondents' remits; (2) Temporal Factors—responses collected at a single time point may underrepresent early challenges due to recall bias; (3) Observer Perspective—Sweden and Iceland insights reflect pre-implementation perspectives; (4) Individual vs. Institutional Views—multiple respondents from single countries provided individual perspectives that may not represent institutional consensus; (5) Scope Boundaries—this analysis focuses on implementation challenges, with clinical findings reported separately.



## 4 Methods

### 4.1 Survey Design and Development

A structured survey instrument was developed to systematically capture implementation experiences across all VALO participating sites. The survey comprised 36 questions organized into six thematic sections (Background Information, Data Governance, Data Life Cycle, Data Processing, VALO Study Process Steps, Future Investments) designed to assess different aspects of the OMOP implementation lifecycle. The complete survey structure is provided in Appendix A.

The survey employed multiple question formats: binary responses (Yes/No/Not known) for factual assessments, multiple choice checkboxes for data availability, descriptive rating scales for challenge assessment (ranging from "Not challenging" to "Extremely challenging"), and open text fields for detailed explanations and recommendations. Question 17 utilized a matrix format to assess data availability across four categories (OMOP-ready, available but not mapped, not available, not in remit).

### 4.2 Data Collection Process

The survey was distributed electronically to designated representatives at all VALO consortium sites in September 2025. Sites were encouraged to have multiple stakeholders complete the survey to capture diverse perspectives across technical, clinical, and administrative domains. Respondents were informed that some questions applied only to data partner sites, that observer sites should share current status insights, and that questions outside individual remits could be marked accordingly. Responses were collected electronically and exported to Excel format for analysis.

### 4.3 Respondent Characteristics

Ten respondents representing five countries participated in the assessment. Respondent distribution by country and role is provided in Appendix B. Experience with OMOP/RWE studies ranged from 2.5 years to extensive experience across respondents. Five respondents indicated previous federated network participation.

### 4.4 Data Analysis Approach

#### 4.4.1 Quantitative Analysis

Response completeness was calculated by country, determining the percentage of applicable questions answered. For questions with multiple respondents per country, responses were aggregated to identify consensus and variation within sites. Challenge ratings were analyzed using a standardized five-level scale (Appendix C), converting descriptive categories to enable cross-site comparison.

#### 4.4.2 Qualitative Analysis

Open text responses were analyzed using thematic content analysis, with attention to technical challenges, governance barriers, resource requirements, infrastructure limitations,



Funded by the  
Nordic Council  
of Ministers

Lessons learned from multi-site federated analysis – VALO Value from Nordic Health Data 14

and improvement recommendations. Direct quotations were preserved and attributed by country and question number to ensure traceability.

#### **4.4.3 Cross-Site Comparison**

Responses were compared across sites to identify common challenges, site-specific issues, differences between active implementers and observers, and variations in technical maturity.

### **4.5 Methodological Considerations**

#### **4.5.1 Response Interpretation**

"Not in my remit" responses were distinguished from missing data, indicating questions outside professional scope rather than implementation gaps. For sites with multiple respondents, conflicting responses were presented as ranges rather than averaged. No statistical testing was performed due to small sample size; findings represent experiential learning rather than hypothesis testing. Generalizability beyond Nordic OMOP implementations should be considered carefully given the specific regional context.

## 5 Results (Enhanced and Comprehensive)

### 5.1 Background Information

Ten respondents representing five Nordic countries participated in the assessment, providing diverse perspectives across technical, clinical, and administrative domains. Active data partners (Norway, Finland, Denmark) contributed seven responses; observer sites (Sweden, Iceland) contributed three responses. Detailed respondent characteristics are provided in Appendix B.

### 5.2 Data Governance

Data governance assessment revealed widespread uncertainty regarding regulatory requirements across participating sites (Table 1). Ethics approval requirements, data permit processes, and approval timelines were frequently reported as "Not known" or unspecified, even among active data partners. OMOP-specific governance frameworks showed similar fragmentation, with only Norway reporting partial integration with existing data warehouse structures. This governance knowledge gap represents a critical finding addressed in Recommendations [Section 7.1](#). Detailed response distributions are provided in Appendix D.

### 5.3 Data Life Cycle

Assessment of data infrastructure across Nordic sites revealed a three-tier data readiness hierarchy. Tier 1 data (diagnoses, demographics, mortality, standard treatments) demonstrated universal availability and high OMOP readiness. Tier 2 data (history, laboratory results, some biomarkers) showed variable availability requiring site-specific mapping efforts. Tier 3 data (patient-reported outcomes, performance status, genetic data, disease progression) exhibited critical gaps across all sites, representing fundamental limitations for precision oncology research. Detailed data source inventories and linkage assessments are provided in Appendix E; the data availability matrix below summarizes readiness status by data category.

#### 5.3.1 Data Availability Matrix Analysis (3, 4)

The systematic assessment of data readiness across types revealed critical gaps.

Table 1: Data Type Readiness Assessment

Data Category	OMOP-Ready Status	Available but Requires Mapping	Not Available	Notes
Diagnoses	<ul style="list-style-type: none"> <li>Norway: Partially ready (1/3 respondents)</li> <li>Denmark: OMOP-ready</li> <li>Finland: Requires mapping</li> </ul>	<ul style="list-style-type: none"> <li>Norway: 2/3 respondents report available</li> <li>Finland: Available in source systems</li> <li>Iceland: Available but unmapped</li> </ul>	<ul style="list-style-type: none"> <li>None</li> </ul>	Universal availability confirmed



Lessons learned from multi-site federated analysis – VALO Value from Nordic Health Data 16

Data Category	OMOP-Ready Status	Available but Requires Mapping	Not Available	Notes
<b>Procedures</b>	<ul style="list-style-type: none"> <li>Denmark: OMOP-ready</li> <li>Others: Not ready</li> </ul>	<ul style="list-style-type: none"> <li>Norway: All respondents confirm availability</li> <li>Finland: In EHR systems</li> <li>Iceland: Available</li> </ul>	<ul style="list-style-type: none"> <li>None</li> </ul>	Mapping effort required
<b>Prescribed drugs</b>	<ul style="list-style-type: none"> <li>Norway: 2/3 OMOP-ready</li> <li>Denmark: Ready</li> <li>Finland: Ready</li> </ul>	<ul style="list-style-type: none"> <li>Norway: 1/3 requires mapping</li> </ul>	<ul style="list-style-type: none"> <li>None</li> </ul>	High readiness level
<b>Administered drugs</b>	<ul style="list-style-type: none"> <li>Denmark, Finland: OMOP-ready</li> <li>Requires mapping: Norway, Iceland, Sweden-KUH</li> <li>Not available: Sweden-KI</li> </ul>	<ul style="list-style-type: none"> <li>Most sites have in clinical systems</li> </ul>	<ul style="list-style-type: none"> <li>Some sites lack integration</li> </ul>	Variable availability
<b>Dispensed drugs</b>	<ul style="list-style-type: none"> <li>Denmark: Registry ready</li> <li>Others: Limited</li> </ul>	<ul style="list-style-type: none"> <li>Available in prescription databases</li> </ul>	<ul style="list-style-type: none"> <li>Hospital dispensing gaps noted</li> </ul>	Registry dependent
<b>Laboratory Results</b>	<ul style="list-style-type: none"> <li>No sites report OMOP-ready status</li> </ul>	<ul style="list-style-type: none"> <li>Universal availability in source systems</li> <li>Norway, Denmark, Finland all confirm presence</li> </ul>	<ul style="list-style-type: none"> <li>None</li> </ul>	Universal mapping need
<b>Imaging Results</b>	<ul style="list-style-type: none"> <li>No OMOP readiness</li> </ul>	<ul style="list-style-type: none"> <li>Denmark: Available</li> <li>Norway: 1/3 has access</li> </ul>	<ul style="list-style-type: none"> <li>Norway: 2/3 report not available</li> <li>Finland: Not accessible</li> </ul>	Limited availability
<b>Pathology Results</b>	<ul style="list-style-type: none"> <li>Limited OMOP mapping</li> </ul>	<ul style="list-style-type: none"> <li>Available in most sites</li> </ul>	<ul style="list-style-type: none"> <li>Some sites lack digital pathology</li> </ul>	Text-heavy format
<b>Biomarker Data</b>	<ul style="list-style-type: none"> <li>No OMOP availability</li> </ul>	<ul style="list-style-type: none"> <li>Denmark: Available in clinical systems</li> </ul>	<ul style="list-style-type: none"> <li>Norway: All 3 respondents report absence</li> <li>Finland: Not available</li> </ul>	Critical gap identified
<b>Patient Reported Outcomes</b>	<ul style="list-style-type: none"> <li>No OMOP availability</li> </ul>	<ul style="list-style-type: none"> <li>No systematic collection reported</li> </ul>	<ul style="list-style-type: none"> <li>Universal absence across all sites</li> </ul>	Not part of routine care
<b>Genetic Data</b>	<ul style="list-style-type: none"> <li>No OMOP availability</li> </ul>	<ul style="list-style-type: none"> <li>Sweden KI: Research settings only</li> </ul>	<ul style="list-style-type: none"> <li>Clinical availability minimal</li> </ul>	Limited to research

### 5.3.2 Coding Systems

All active sites reported ICD-10 usage; additional coding systems included SNOMED/SNOMED CT, national classification systems, and HCPCS at selected sites.

## 5.4 Data Processing

Variable capture assessment revealed differential completeness across sites and data categories. Core oncology variables (cancer diagnoses, demographics, mortality, standard treatments) achieved high capture rates (90-100%), while precision medicine variables (biomarkers, performance status, smoking status) demonstrated critical gaps (<40% complete). Implementation challenges followed a clear difficulty gradient: ETL specification design represented the highest challenge point, while QA/QC processes were facilitated by OHDSI tools. Norway provided the most detailed technical feedback, noting local infrastructure challenges with R environment setup.

### 5.4.1 Variable Capture Completeness

The assessment of which variables were successfully captured within the NSCLC cohort revealed differential data completeness across sites.

Table 2: Clinical Variable Capture Success

Variable Category	Norway	Denmark	Finland	Overall Capture Rate	Quality Notes
<b>Cancer Diagnoses</b>	Captured (all respondents)	Captured	Captured	100% successful	High quality, ICD-10 based
<b>Histology</b>	Partial (2/3 respondents)	Captured	Limited capture	60-70% complete	Variability in documentation
<b>Biomarkers</b>	Not captured systematically	Limited	Not available	<30% complete	Major limitation for precision oncology
<b>Comorbidities</b>	Captured via diagnoses	Captured	Captured	90% successful	Charlson index calculable
<b>Demographics</b>	Complete	Complete	Complete	100% successful	Age, sex universally available
<b>Smoking Status</b>	Not systematically captured	Limited	Limited	<40% complete	Critical gap for lung cancer
<b>Performance Status</b>	Not systematically captured	Not captured	Limited	<20% complete	ECOG/PS largely missing
<b>Treatment Details</b>					
- Chemotherapy	Captured	Captured	Captured	95% successful	Drug codes available

Variable Category	Norway	Denmark	Finland	Overall Capture Rate	Quality Notes
- Immunotherapy	Captured	Captured	Captured	95% successful	ICI identification successful
- Surgery	Procedure codes	Captured	Captured	90% successful	Procedure coding complete
- Radiotherapy	Variable capture	Captured	Partial	60% complete	Site-specific gaps
<b>Death</b>	Registry linkage	Complete	Available	95% successful	High quality mortality data
<b>Disease Progression</b>	Not structured	Limited	Not captured	<20% complete	Requires clinical notes review

### 5.4.2 Implementation Challenge Gradients

Systematic assessment of challenge levels across eight OMOP process stages revealed a clear difficulty gradient.

Table 3: Detailed Challenge Assessment Matrix

Process Stage	Norway (n=3)	Denmark (n=2)	Finland (n=2)	Iceland (n=1)	Interpretation
<b>1. Design OMOP ETL Specification</b>	<ul style="list-style-type: none"> <li>• "Moderately challenging" (2)</li> <li>• "Significant effort but manageable"</li> <li>• Required dedicated resources</li> </ul>	<ul style="list-style-type: none"> <li>• Limited responses</li> <li>• Clinical researchers not involved in technical design</li> </ul>	<ul style="list-style-type: none"> <li>• Responses not provided</li> <li>• Technical team may not have participated</li> </ul>	<ul style="list-style-type: none"> <li>• "Not in my remit"</li> <li>• Observer status</li> <li>• Learning from others' experience</li> </ul>	<p><b>Highest challenge point</b></p> <p>Design phase requires most resources</p>
<b>2. Implement OMOP ETL Specification</b>	<ul style="list-style-type: none"> <li>• "Slightly challenging" (2)</li> <li>• "Minor adjustments needed"</li> <li>• Benefited from design phase</li> </ul>	<ul style="list-style-type: none"> <li>• Technical implementation outside respondent scope</li> </ul>	<ul style="list-style-type: none"> <li>• Limited visibility into technical process</li> </ul>	<ul style="list-style-type: none"> <li>• Not applicable</li> </ul>	<p><b>Moderate challenge</b></p> <p>Easier once design complete</p>
<b>3. QA/QC of OMOP Instance</b>	<ul style="list-style-type: none"> <li>• "Not challenging" (2)</li> <li>• "Routine process"</li> <li>• DataQualityDashboard effective</li> </ul>	<ul style="list-style-type: none"> <li>• Process established but not documented in responses</li> </ul>	<ul style="list-style-type: none"> <li>• QC procedures in place</li> </ul>	<ul style="list-style-type: none"> <li>• Not applicable</li> </ul>	<p><b>Lowest challenge</b></p> <p>OHDSI tools simplify QC</p>



Process Stage	Norway (n=3)	Denmark (n=2)	Finland (n=2)	Iceland (n=1)	Interpretation
<b>4. Study Variable Survey</b>	<ul style="list-style-type: none"> <li>Mixed responses</li> <li>Some variables unavailable</li> <li>Required iterations</li> </ul>	<ul style="list-style-type: none"> <li>Clinical input provided</li> <li>Feasibility challenges noted</li> </ul>	<ul style="list-style-type: none"> <li>Variable availability issues</li> </ul>	<ul style="list-style-type: none"> <li>Observing process</li> </ul>	<p><b>Moderate complexity</b></p> <p>Data availability constrains</p>
<b>5. Diagnostic Package Execution</b>	<ul style="list-style-type: none"> <li>Successfully completed</li> <li>Some debugging required</li> </ul>	<ul style="list-style-type: none"> <li>Limited technical responses</li> </ul>	<ul style="list-style-type: none"> <li>Execution challenges noted</li> </ul>	<ul style="list-style-type: none"> <li>Not applicable</li> </ul>	<p><b>Technical hurdles</b></p> <p>Environment-specific issues</p>
<b>6. Initial Package Execution</b>	<ul style="list-style-type: none"> <li>Completed with delays</li> <li>R environment issues</li> </ul>	<ul style="list-style-type: none"> <li>Not documented</li> </ul>	<ul style="list-style-type: none"> <li>Technical delays reported</li> </ul>	<ul style="list-style-type: none"> <li>Not applicable</li> </ul>	<p><b>Infrastructure dependent</b></p> <p>Local setup affects success</p>
<b>7. Preliminary Results Review</b>	<ul style="list-style-type: none"> <li>Iterative process</li> <li>Multiple rounds needed</li> </ul>	<ul style="list-style-type: none"> <li>Clinical review provided</li> </ul>	<ul style="list-style-type: none"> <li>Data quality issues identified</li> </ul>	<ul style="list-style-type: none"> <li>Learning opportunity</li> </ul>	<p><b>Time intensive</b></p> <p>Quality review crucial</p>
<b>8. Final Package Execution</b>	<ul style="list-style-type: none"> <li>Successful completion</li> <li>Lessons learned applied</li> </ul>	<ul style="list-style-type: none"> <li>Completed successfully</li> </ul>	<ul style="list-style-type: none"> <li>Final execution achieved</li> </ul>	<ul style="list-style-type: none"> <li>Not applicable</li> </ul>	<p><b>Smoother process</b></p> <p>Earlier issues resolved</p>

## 5.5 VALO Pilot Study Process Steps

Survey responses regarding specific OMOP implementation phases (Questions 25-32) revealed limited technical detail from most sites, reflecting the specialized nature of ETL execution and survey respondent profiles. Norway provided the most substantive feedback, noting "We were able to reuse work from previous studies" — highlighting the value of cumulative OMOP experience. The consistent reference to previous answers ("As for 29") in Norwegian responses indicates a unified approach across ETL design, implementation, and QA/QC phases. Iceland confirmed pre-implementation status ("OMOP is not yet implemented in Iceland"), while Denmark and Finland responses were managed by technical teams not represented in the survey. Detailed question-by-question responses are provided in Appendix F.

## 5.6 Future Investments and Improvements

Future investment priorities emerged from respondent feedback, with Norway and Denmark providing the most substantive strategic insights. Key findings include:

**Infrastructure Integration (Norway):** "Having a clinical data warehouse as basis for the OMOP implementation has made the process part of regular operations." Technical



challenges included database compatibility: "We had issues with handling postgres bigints in R, which has a lower limit than what bigint can store."

**Capability Development (Norway):** "OMOP skills have been established in the data warehouse team over the last 3 years" — indicating multi-year investment requirements. Standardization challenges were noted: "Vocabulary versioning is challenging. As a consortium we could define target vocabulary version."

**Methodological Advancement (Denmark):** Two Danish respondents emphasized analytical enhancement needs: "Developing more advanced methods for conducting federated analysis and obtaining aggregated results" and "Improving analytical capabilities into inferential and causal analytics" — highlighting current limitations of primarily descriptive federated analyses.

Detailed question-by-question responses are provided in Appendix G.

### 5.6.1 Summary of Investment Priorities

Based on the responses provided, key investment areas identified by the consortium include:

1. **Infrastructure optimization:** Addressing technical compatibility issues between database systems and analytical environments (Norway's postgres/R challenge)
2. **Capability development:** Building OMOP expertise within technical teams, a multi-year investment as evidenced by Norway's experience
3. **Standardization needs:** Vocabulary versioning coordination across the consortium
4. **Methodological advancement:** Moving beyond descriptive statistics to inferential and causal methods in federated settings
5. **Process integration:** Leveraging existing data warehouse infrastructure to embed OMOP as routine operations

The variation in response detail across sites reflects different stages of OMOP maturity, with active implementation sites providing concrete technical feedback while observer sites focus on learning and planning. The emphasis on both technical infrastructure (database compatibility) and methodological capabilities (advanced analytics) indicates that successful OMOP implementation requires investment across multiple domains.

## 5.7 IQVIA Execution Experience

The IQVIA coordinating center documented operational insights across the complete study execution lifecycle, complementing site-level survey responses. Key quantitative findings include: (1) Timeline calibration: initial technical execution estimates should be multiplied by 1.5-2.0× to account for multi-site coordination complexity, with 30-40% of total timeline allocated to coordination overhead; (2) Patient recruitment reality: final analytical cohorts represented only 10-20% of initial population estimates after applying eligibility criteria—Norway's initial estimate of approximately 1,500 eligible patients reduced dramatically after study period restrictions, lookback windows, and clinical criteria were applied; (3) Success factors: sites with designated accountable coordinators demonstrated 2-3× faster issue



resolution; (4) Communication evolution: transition from email to Microsoft Teams channels (initiated by Oslo University Hospital) proved critical for real-time technical debugging. Detailed risk categories, stakeholder management approaches, and resource coordination experiences are provided in Appendix H.

The VALO study consortium—comprising the five Nordic participating sites (Norway, Finland, Denmark, Sweden, Iceland)—was led by a dedicated Project Manager (PM). The PM served as the primary coordination point both for site interactions and internal team management, with core responsibilities encompassing communication, coordination, and timely project delivery. In addition, the Project Manager was responsible for risk and issue management and mitigation, including identifying potential opportunities for optimization, implementing appropriate support mechanisms, and assisting with troubleshooting operational questions as they arose.

The IQVIA executive team - comprising of data scientists, analytical specialists, epidemiologists and medical writers in addition to the PM - provided complementary operational insights beyond the site-level survey responses. This team, responsible for feasibility assessment, analytical package development, site technical support, and results synthesis, documented experiences across the complete study execution lifecycle through structured interviews and project records.

### **5.7.1 Project Management and Coordination Experience**

The VALO pilot represented a complex multi-site coordination effort spanning five Nordic countries, three active data partners, and two observer sites over a twelve-month timeline (November 2024 - October 2025). The IQVIA PM served as the central coordination point for consortium activities, responsible for communication, timeline management, risk identification and mitigation, operational problem-solving and overall timely and high-quality project delivery. The PM experience revealed several critical lessons about managing federated research networks that extend beyond technical OMOP implementation challenges. A key factor in future success is that the central co-ordinatory project manager has a corresponding contact at each site, a designated team member accountable for coordination, communication and timeline adherence.

#### **Timeline Management and Expectation Calibration**

A key learning emerged in reconciling initial timeline expectations with operational realities. The original project plan anticipated a streamlined execution based on sites' reported OMOP readiness. However, actual timelines extended significantly beyond initial estimates due to factors that were difficult to predict during planning: These experiences underscore that timeline planning for federated studies must incorporate substantial buffer for coordination overhead rather than assuming linear execution. A recommendation emerging from this experience: multiply initial technical execution estimates by 1.5-2.0× to account for multi-site coordination complexity.

#### **Site Coordination**

The marked absence of clearly designated primary coordinators with both authority and dedicated time impacted the efficiency of the pilot conduct. Sites without such roles demonstrate less streamlined execution, as activities and coordination tend to be dispersed among several individuals, resulting in disruption in communication and delays in delivery.



The VALO consortium was invited for an online kick-off meeting, to an online workshop some weeks later, where previous OMOP experiences were shared, and to a second workshop at the end of the study, where pilot results and collected lessons learned were jointly assessed and discussed. In addition, e-mail updates on the study progress were provided for all. More detailed study progress and next steps -communication was provided primarily for the data provider sites.

The project's data science -related communication infrastructure evolved substantially based on operational needs:

**Initial Phase (Email-based):**

- Primary communication via email threads
- **Opportunities identified:** Slow response times, difficulty tracking complex technical discussions, information siloing
- Result: Transition to enhanced communication methods identified as beneficial for real-time technical debugging and collaborative problem-solving

**Evolved Phase (Teams-based):**

- Transition to dedicated Microsoft Teams channels per site (initiated by Oslo University Hospital)
- Advantages: Real-time debugging support, persistent documentation of issues and solutions, improved response times
- Result: Substantial improvement in issue resolution efficiency, identified as critical success factor

Establishing dedicated, persistent communication channels for technical collaboration should be a Day 1 requirement rather than an evolution. The IQVIA operational team recommends provisioning such infrastructure as part of project initiation.

### 5.7.2 Lessons for Future Network Project Management

Based on this experience, future federated network studies should:

- **Allocate even up to 30-40% of total timeline to coordination overhead** rather than assuming pure technical execution time
- **Establish communication infrastructure (Teams/Slack channels) at project initiation** with all key roles onboarded
- **Require sites to designate a primary coordinator** with authority and protected time before project start
- **Conduct early role-mapping exercises** identifying who will respond to different query types (technical, clinical, governance)
- **Build in structured checkpoints** (e.g. weekly in active phases, bi-weekly in stable phases) rather than relying on ad hoc communication



- **Document all technical solutions in shared repositories** to prevent repeated troubleshooting of identical issues
- **Maintain transparent risk registers** visible to all consortium members to enable collaborative mitigation
- **Plan for summer/holiday periods** with adjusted timelines or advance work completion before predictable absence periods

The PM role in federated research networks extends far beyond administrative coordination. It requires active bridging of technical, clinical, and governance domains; continuous calibration of expectations against operational realities; and maintenance of consortium cohesion despite inevitable frustrations. The investment in dedicated, experienced project management proved essential to completing this pilot study and should be considered a **core competency requirement** for future network development.

### 5.7.3 Patient Cohort Assembly Reality

The IQVIA operational team documented substantial discrepancies between initial feasibility estimates and actual patient enrollment. Sites initially projected the availability of substantial NSCLC patient populations receiving immune checkpoint inhibitors. However, application of full study criteria resulted in dramatic attrition:

**Norway (Oslo University Hospital):** Initial estimates suggested approximately 1,500 eligible patients. Final analytical cohort included 10-20% of projected numbers after applying study period restrictions, lookback windows, and clinical criteria.

**Finland (HUS):** Similar magnitude of difference observed between preliminary counts and final cohort size, with final numbers representing 10-20% of initial projections.

**Denmark (Rigshospitalet):** While implementing OMOP for the first time, experienced comparable recruitment challenges to experienced sites.

The attrition was particularly pronounced when stratification by age ( $\geq 75$  years) was applied, with at least one site unable to complete planned stratified analyses due to insufficient patient numbers in subgroups.

### 5.7.4 Technical Execution Challenges

**Database Management System Variability:** The coordinating center identified that bespoke code elements—those not utilizing established OHDSI/DARWIN functions—demonstrated differential performance across sites' database management systems. Specific technical challenges included:

- Arithmetic operations for date calculations functioned on Snowflake but required alternative approaches at other sites
- One site utilized DataBricks, a less-tested platform within OHDSI/DARWIN frameworks, resulting in unexpected function behaviors
- Database-specific adaptations were managed through conditional logic (if/else statements) checking database names, though Rigshospitalet ultimately required a separate package version



**Measurement Characterization Complexity:** The team documented extensive variation in measurement mapping across sites, particularly for biomarkers critical to the study:

- PD-L1 captured through different concept IDs across sites
- Values recorded as numerical, categorical (e.g., "negative"), or relative to thresholds (e.g., ">50%")
- Multiple records on single dates when source data contained ranges
- Measurements spanning both Measurement and Observation tables within single sites

The absence of diagnostic packages capable of profiling measurement variations necessitated bespoke characterization code within the analytical package, increasing development time and debugging requirements.

### 5.7.5 Successful Adaptations and Innovations

Despite challenges, several successful adaptations were noted:

**Protocol Flexibility:** Deliberate maintenance of high-level protocol elements enabled mid-study adaptations, including development of proxy measurements for adjuvant versus neoadjuvant chemotherapy when direct capture proved unavailable.

**Code Reusability:** Site-specific adaptations were successfully managed within a single codebase using conditional logic, avoiding the need for multiple package versions across most sites.

**Accelerated Timeline Achievement:** The team noted that treatment pattern analyses comparable to those from the DYNASTY study (2-year timeline)(5) were completed within the 3-month VALO pilot timeframe, validating the efficiency of the federated approach once operational.

### 5.7.6 Genomic Vocabulary Implementation

One site's implementation of OMOP Genomics vocabulary presented both advancement and challenge. While this approach retained maximum information through allele-level mutation recording, it complicated network-level analysis by requiring:

- Comprehensive clinical review to identify eligible allele mutations
- Site-specific concept lists rather than hierarchical queries
- Additional development time for biomarker characterization

The IQVIA team noted this as an example where technical best practice at the site level created network-level analytical complexity.



## 6 Discussion

This assessment documents OMOP CDM implementation experiences across Nordic healthcare systems, providing empirical evidence for federated research network development. Survey responses from ten participants across five countries, complemented by IQVIA operational team execution documentation, establish implementation realities and actionable guidance for future network expansion.

### 6.1 Principal Findings and Interpretation

This comprehensive assessment of OMOP CDM implementation across Nordic countries documents technical, organizational, and methodological implementation experiences that extend beyond the common data model itself. The survey responses from ten participants across five countries provide insights into establishing into the realities of establishing federated research networks in heterogeneous healthcare systems. These site-level perspectives are further validated and expanded by the Consortium lead's execution experience, which documented the operational realities of translating OMOP implementation into scientific output.

A key finding is the implementation phase gradient in OMOP adoption, with ETL design representing the highest barrier (moderately challenging), followed by implementation (slightly challenging), and QA/QC emerging as routine. This progression demonstrates that early investment in design enables efficiency in subsequent phases—a pattern explicitly confirmed by Norway's ability to "reuse work from previous studies." The IQVIA operational team's experience reinforces this finding, noting that protocol flexibility deliberately maintained during design phase enabled critical mid-study adaptations, including development of proxy measurements for adjuvant versus neoadjuvant chemotherapy when direct capture proved unavailable (document in the main study report). This finding has profound implications for network development strategies, indicating that shared design resources and templates could accelerate adoption across new sites.

The disconnect between technical infrastructure challenges and OMOP methodology success represents a critical insight. Norway's specific issues—postgres bigint handling in R, analytical environment setup delays—were infrastructure-related rather than conceptual. The coordinating center documented additional technical complexity: bespoke code elements demonstrated differential performance across database management systems, with arithmetic operations functioning on Snowflake but requiring alternative approaches at other sites. One site's use of DataBricks, a less-tested platform within OHDSI/DARWIN frameworks, necessitated ultimately running a separate package version. This distinction is crucial: sites struggle not with understanding OMOP principles but with technical implementation in their specific environments. This finding reframes the support needed for new implementations from conceptual training to technical troubleshooting and environment configuration.

A notable finding emerged from patient recruitment, with a variation between initial site estimates of eligible NSCLC patients and the significantly smaller number of patients who fulfilled all study criteria. To advance learning and improve future study outcomes, it is essential to implement more robust feasibility assessments to mitigate the risk of similar attrition. Moreover, conducting thorough investigations into the underlying factors that contributed to the unexpected findings—such as the notably high attrition rate observed



during the pilot study—will provide valuable insights and inform methodological refinements for subsequent research.

## 6.2 Data Readiness and Systematic Gaps

The data availability matrix reveals an important pattern: while basic clinical data (diagnoses, procedures, medications) show high availability and varying degrees of OMOP readiness, advanced data types critical for modern oncology research (biomarkers, performance status, PROs) are systematically absent. This creates a fundamental tension between the promise of comprehensive real-world evidence and the reality of available data.

### Three-Tier Data Readiness Hierarchy Emerged:

**Tier 1 - Ready or Near-Ready:** Prescribed drugs (>60% OMOP-ready), diagnoses (30% ready, universally available), and core demographics represent the foundation layer where OMOP implementation succeeds.

**Tier 2 - Available but Requiring Substantial Effort:** Laboratory results, procedures, and administered drugs exist in source systems but require significant mapping effort. The universal availability of laboratory data with 0% OMOP readiness exemplifies this challenge—data exists but lacks standardization.

**Tier 3 - Systematically Absent:** Biomarkers (Norway: all three respondents report absence), PROs (universal absence), and performance status (<20% capture) represent critical gaps that cannot be solved through technical mapping alone. These require changes in clinical documentation practices.

The implications are profound: studies requiring Tier 3 data elements will fail regardless of OMOP implementation quality. This suggests a need for stratified feasibility assessment based on data tier requirements before study initiation.

### Understanding the Nature of Tier 3 Data Gaps

The systematic absence of Tier 3 data elements reflects fundamentally different challenges than Tier 2 mapping requirements. Based on site responses and execution experience, three distinct gap mechanisms emerged:

#### 1. Clinical Documentation Practices (Biomarkers, Performance Status):

These data elements often exist in clinical practice but are not captured in structured, query-able formats. The operational team's experience with biomarker heterogeneity illustrates this challenge: when PD-L1 data was available, it appeared through different concept IDs across sites, with values recorded as numerical results, categorical interpretations (e.g., "negative"), or threshold-based descriptions (e.g., ">50%"). Some sites captured biomarkers in both Measurement and Observation tables, creating multiple records on single dates when source data contained ranges.

This heterogeneity indicates that biomarkers are being **tested and documented** but not in standardized formats compatible with systematic research extraction. Performance status



faces similar challenges—ECOG scores are assessed during clinical encounters but frequently documented in narrative clinical notes rather than discrete fields, making systematic capture infeasible without natural language processing capabilities not currently deployed.

## **2. Structural Absence from Routine Care (Patient-Reported Outcomes):**

PROs represent a different challenge: systematic absence from routine clinical workflows. Unlike biomarkers that are captured inconsistently, PROs are generally **not collected** as part of standard care delivery in Nordic healthcare systems. The universal unavailability across all sites indicates this is not a technical mapping problem but reflects that these data elements are not generated during routine care encounters. Digital infrastructure for PRO collection, while technically feasible, has not been implemented in the clinical settings represented in this consortium.

## **3. Research Versus Clinical Context (Genetic Data):**

Certain data elements, particularly comprehensive genetic/genomic data, remain predominantly within research contexts rather than integrated into routine clinical databases. Sweden (KI) noted availability in research settings, but clinical availability remains minimal. This segregation between research and clinical data systems represents an architectural challenge distinct from documentation or collection issues.

### **Implications for Feasibility Assessment:**

These mechanistic differences require tailored solutions:

- **Documentation heterogeneity** (biomarkers, performance status) may be addressable through enhanced EHR templates, clinical workflow modifications, and standardized capture protocols, though such changes require clinical engagement and multi-year implementation timelines.
- **Structural absence** (PROs) requires new infrastructure and workflows, representing more fundamental intervention than technical mapping.
- **Research-clinical segregation** (genetic data) requires data architecture decisions about integration versus federated access across systems.

Studies dependent on Tier 3 data must evaluate not just whether elements are available but **which gap mechanism** applies, as this determines feasibility of enhancement efforts within study timelines.

## **6.3 Governance and Regulatory Uncertainty**

The widespread "Not known" responses regarding ethics approval requirements and timelines reveal a concerning disconnect between technical teams implementing OMOP and institutional governance structures. This knowledge gap—evident even in experienced sites like Norway where 2/3 respondents were uncertain about ethics requirements—suggests that OMOP implementation often proceeds in technical isolation from regulatory frameworks.

The variation in responses within single countries (Norway: one "Yes," two "Not known" for ethics requirements) indicates that governance knowledge is role-dependent rather than



institutionally embedded. This fragmentation poses risks for study execution and suggests the need for cross-functional team structures that bridge technical and regulatory domains.

## 6.4 The Learning Curve and Capability Development

Norway's report that "OMOP skills have been established in the data warehouse team over the last 3 years" provides the only quantified learning timeline in the survey. This three-year capability development period represents crucial intelligence for institutions planning OMOP adoption. It establishes that sustainable OMOP implementation is a multi-year capability building journey rather than a short-term project requiring sustained organizational commitment. The IQVIA operational team experience adds a critical dimension to this learning curve: despite the extended capability development timeline, the team successfully completed treatment pattern analyses within three months that were comparable to the DYNASTY study's two-year effort(5). This demonstrates that front-loaded capability investment yields substantial returns in study execution efficiency once infrastructure becomes operational.

The contrast between active sites and observers illuminates different stages of the adoption curve. Iceland's consistent "OMOP is not yet implemented" responses, paired with high survey completion rate (89%), demonstrates engaged learning despite lack of hands-on experience. Sweden's dual perspective—research infrastructure at Karolinska Institutet versus clinical systems at Karolinska University Hospital—highlights the additional complexity of bridging research and care delivery systems. The IQVIA's operational team documented that even experienced sites developed deeper understanding of their actual data capabilities through implementation experience. Sites expressed confidence in measurement linkage to primary tumors during initial discussions, but execution revealed that linkages were partial, inconsistent, or required additional development. This evolution from perceived to validated capabilities establishes that the learning curve extends beyond technical skills to encompass deep understanding of data realities.

The IQVIA operational team also documented the evolution of collaboration infrastructure as part of the learning process. The transition from email-based communication to dedicated Microsoft Teams channels, initiated by Oslo, transformed debugging efficiency and became a model for other sites. This organizational learning—discovering optimal communication modalities through experience—represents an often-overlooked dimension of capability development that extends beyond technical OMOP skills to encompass collaborative workflows essential for network success.

## 6.5 Network Coordination Challenges

The vocabulary versioning challenge identified by Norway—"As a consortium we could define target vocabulary version"—reveals a critical network coordination need. Without synchronized vocabulary versions across sites, the promise of federated analysis is compromised by semantic inconsistencies. This technical detail has strategic implications: successful networks require not just compatible data models but synchronized maintenance schedules.

The limited response rates for technical execution questions (29-32) suggest another coordination challenge: the separation between those who implement OMOP and those



who participate in network governance. This disconnect may explain why critical technical lessons remain undocumented and unshared across the network.

## 6.6 From Descriptive to Inferential Analytics

Denmark's articulation of the need for "improving analytical capabilities into inferential and causal analytics" and "developing more advanced methods for conducting federated analysis" identifies the next frontier beyond basic OMOP implementation. Current federated analyses remain largely descriptive, limiting the scientific value of the network investment.

This methodological gap represents a fundamental limitation: even perfect OMOP implementation and complete data cannot answer causal questions without appropriate analytical methods. The challenge is compounded in federated settings where patient-level data pooling is impossible, requiring novel approaches to adjustment and inference.

## 6.7 Infrastructure Optimization as Success Factor

Infrastructure alignment with OMOP requirements emerged as a critical success determinant. Norway's experience demonstrates both dimensions: leveraging existing clinical data warehouse infrastructure enabled OMOP to become "part of regular operations," while addressing specific technical compatibility requirements (postgres/R data type handling) refined implementation capabilities. This pattern establishes that sites with mature data infrastructure accelerate OMOP adoption when paired with proactive technical environment assessment.

The IQVIA's operational team documented infrastructure considerations that inform realistic timeline planning. Several sites experienced extended timelines for study-specific database instance provisioning, with IT governance and outsourcing arrangements requiring advance planning in project schedules. These experiences established that infrastructure readiness assessment should precede project commitment, enabling appropriate resource allocation and timeline development. Sites successfully addressed these requirements through early engagement with IT teams, documented technical specifications, and contingency planning for environment configuration.

### **Balancing Technical Advancement with Network Interoperability**

HUS's implementation of OMOP Genomics vocabulary exemplifies an important network design consideration: the relationship between site-level technical sophistication and network-level analytical approaches. HUS retained allele-level mutation detail—representing technical best practice for genomics data capture—which maximizes information preservation for site-specific analyses. However, this granular approach required comprehensive clinical review to identify eligible allele mutations for each biomarker, necessitating specialized concept lists rather than standard hierarchical queries for network studies. This experience establishes that sites can optimize implementation depth for local analytical requirements while the network develops harmonization approaches that leverage site-specific capabilities within federated frameworks. The IQVIA team successfully managed this through site-specific characterization within analytical packages, demonstrating that technical heterogeneity is addressable through adaptive analytical strategies.



## 6.8 Implications for Future Network Studies

### 6.8.1 Feasibility Assessment Must Be Data-Tier Specific

Studies requiring only Tier 1 data (diagnoses, medications, demographics) can proceed with confidence across all active sites. Studies requiring Tier 2 data (laboratories, procedures) need extended timelines for mapping. Studies requiring Tier 3 data (biomarkers, PROs, performance status) may not be feasible without fundamental changes to clinical documentation.

### 6.8.2 Front-Load Design Investment

The challenge gradient from design to implementation to QA/QC argues for concentrating resources in the design phase. Creating reusable design artifacts—as Norway did successfully—transforms subsequent implementations from creative exercises to execution tasks.

### 6.8.3 Bridge Technical and Governance Domains

The disconnect between technical implementation and regulatory knowledge requires integrated project teams. No respondent had complete visibility across technical, clinical, and regulatory dimensions, suggesting that successful implementation requires explicit cross-functional coordination.

### 6.8.4 Plan for Three-Year Capability Development

Norway's three-year timeline for establishing OMOP skills should inform institutional planning. This is not a six-month project but a multi-year capability development journey requiring sustained investment and patience.

## 6.9 Strengths and Limitations of This Assessment

**Strengths:** This assessment captures diverse perspectives across roles, countries, and implementation stages. The inclusion of observer sites provides valuable pre-implementation insights. The mixed-methods approach combining quantitative challenge ratings with qualitative experiences enables nuanced understanding.

**Limitations:** Response rates varied by question, with technical execution questions showing limited responses. The single time-point assessment may not capture evolution of challenges over time. The separation between survey respondents and technical implementers may have resulted in undocumented lessons. The small sample size (ten respondents) limits statistical analysis, though it represents 100% of consortium countries.

## 6.10 Future Directions

These findings identify eight critical development areas: standardized feasibility assessment tools, shared design repositories, technical configuration guides, governance



Funded by the  
Nordic Council  
of Ministers

## Lessons learned from multi-site federated analysis – VALO Value from Nordic Health Data 31

integration frameworks, advanced analytical methods for federated causal inference, coordinated vocabulary governance, capability development curricula, and clinical documentation enhancement for Tier 3 data elements. The path forward requires recognizing that OMOP implementation is necessary but not sufficient for successful network research—equal attention must be paid to infrastructure configuration, governance integration, capability development, and methodological advancement. Detailed, actionable recommendations addressing each area are presented in [Section 7](#).



## 7 Recommendations

Based on the systematic analysis of implementation experiences across the Nordic OMOP network, we present stratified recommendations for different stakeholders and implementation phases.

### 7.1 Technical Infrastructure Recommendations

#### 7.1.1 Pre-Implementation Infrastructure Assessment

Sites planning OMOP implementation should conduct a detailed technical environment assessment addressing the specific compatibility issues identified in this study:

- **Database-Analytics Compatibility:** Verify data type handling between database systems and R/Python environments, specifically addressing the postgres bigint limitation identified by Norway
- **Memory and Computing Resources:** Allocate sufficient resources for analytical environments, anticipating the infrastructure upgrades Norway required mid-project
- **Environment Configuration:** Establish dedicated analytical environments with documented package dependencies and version control before project initiation

#### 7.1.2 Infrastructure Standardization

Based on Norway's success with clinical data warehouse integration, we recommend:

- **Data Warehouse First Strategy:** Sites with existing data warehouses should leverage these as the foundation for OMOP implementation rather than creating parallel systems
- **Standardized Technical Stack:** Consortium should define recommended technology stacks with known compatibility, reducing site-specific troubleshooting
- **Configuration Templates:** Develop and share environment configuration templates that address known issues

### 7.2 Governance Framework Recommendations

#### 7.2.1 Integrated Governance Structure

The disconnect between technical and regulatory knowledge requires structural solutions:

- **Cross-Functional Implementation Teams:** Mandate participation from technical, clinical, and regulatory representatives from project inception



- **Governance Documentation Requirements:** Create standardized documentation templates that capture ethics requirements, approval timelines, and data permit needs
- **Role-Based Knowledge Maps:** Develop clear RACI (Responsible, Accountable, Consulted, and Informed) matrices defining who needs what knowledge about governance processes, and the project conduct

### 7.2.2 Regulatory Pathway Clarification

Given the widespread uncertainty about ethics and data permit requirements:

- **Country-Specific Regulatory Guides:** Each country should develop clear guidance documents outlining OMOP-specific regulatory requirements
- **Approval Timeline Documentation:** Establish realistic timeline expectations based on actual experiences, replacing current uncertainty with data-driven planning
- **Regulatory Liaison Roles:** Designate specific team members as regulatory liaisons to bridge technical and compliance domains

## 7.3 Data Readiness and Quality Recommendations

### 7.3.1 Tiered Feasibility Assessment Protocol

The systematic data gaps identified across sites necessitate a structured approach to study feasibility assessment. Based on the three-tier data hierarchy that emerged from our analysis, we recommend implementing a mandatory feasibility assessment protocol that aligns study requirements with data reality. The coordinating center's experience documenting 10-20% actual enrollment versus predicted patient counts mandates a fundamental shift in feasibility timing: sites must execute actual cohort definitions during protocol development, not after protocol finalization.

Table 4: Tiered Feasibility Assessment Framework

Data Tier	Data Elements	Readiness Level	Timeline Adjustment	Confidence Level	Feasibility Decision Criteria
<b>Tier 1</b>	<ul style="list-style-type: none"> <li>• Diagnoses (ICD-10)</li> <li>• Prescribed medications</li> <li>• Demographics</li> <li>• Mortality</li> </ul>	<ul style="list-style-type: none"> <li>• &gt;60% OMOP-ready</li> <li>• Universal availability</li> <li>• Standardized coding</li> <li>• Minimal site variation</li> </ul>	<ul style="list-style-type: none"> <li>• Standard timeline</li> <li>• No additional mapping</li> <li>• 2-3 month setup</li> </ul>	<p><b>High confidence</b></p> <ul style="list-style-type: none"> <li>• Cross-site comparable</li> <li>• Minimal harmonization</li> <li>• Proven success</li> </ul>	<ul style="list-style-type: none"> <li>• Proceed with standard protocol</li> <li>• Minimal site adaptation</li> <li>• Regular monitoring sufficient</li> </ul>



Data Tier	Data Elements	Readiness Level	Timeline Adjustment	Confidence Level	Feasibility Decision Criteria
<b>Tier 2</b>	<ul style="list-style-type: none"> <li>Laboratory results</li> <li>Procedures</li> <li>Administered drugs</li> <li>Radiotherapy</li> </ul>	<ul style="list-style-type: none"> <li>0-30% OMOP-ready</li> <li>Available in source</li> <li>Requires mapping</li> <li>Site-specific formats</li> </ul>	<ul style="list-style-type: none"> <li>Add 3-6 months</li> <li>Iterative refinement</li> <li>Site-specific logic</li> <li>Multiple validation cycles</li> </ul>	<p><b>Moderate confidence</b></p> <ul style="list-style-type: none"> <li>Harmonization needed</li> <li>Quality varies</li> <li>Completeness uncertain</li> </ul>	<ul style="list-style-type: none"> <li>Detailed source assessment</li> <li>Pilot extraction required</li> <li>Site-specific protocols</li> <li>Extended QC phase</li> </ul>
<b>Tier 3</b>	<ul style="list-style-type: none"> <li>Biomarkers</li> <li>Performance status</li> <li>PROs</li> <li>Disease progression</li> <li>Genetic data</li> </ul>	<ul style="list-style-type: none"> <li>&lt;20% available</li> <li>Not standardized</li> <li>Often in free text</li> <li>May not exist</li> </ul>	<ul style="list-style-type: none"> <li>May not be feasible</li> <li>6-12 month development</li> <li>Alternative strategies needed</li> </ul>	<p><b>Low confidence</b></p> <ul style="list-style-type: none"> <li>Major gaps identified</li> <li>Requires enhancement</li> <li>Study modification likely</li> </ul>	<ul style="list-style-type: none"> <li>Consider study redesign</li> <li>Evaluate alternatives</li> <li>Prospective collection</li> <li>Narrow to available sites</li> </ul>

### Implementation Process for Feasibility Assessment:

Each proposed study must undergo systematic evaluation against this framework before resource commitment. The assessment should begin with cataloging all required data elements and assigning each to the appropriate tier. A critical addition from IQVIA team experience is that feasibility assessments must include execution of full cohort definitions with all inclusion/exclusion criteria during the protocol development phase. The IQVIA operational team documented that simple queries (e.g., "non-small cell lung cancer patients on ICIs") yielding 1,500 patients reduced to 150-300 when study period restrictions, lookback windows, and clinical criteria were applied. This order-of-magnitude reduction was consistent across sites, suggesting systematic overestimation when relying on high-level queries.

Studies requiring predominantly Tier 1 data can proceed with confidence and standard timelines. However, studies dependent on Tier 2 elements require detailed technical assessment at each site, including test extractions to verify data quality and completeness. The IQVIA team experience with measurement heterogeneity—PD-L1 appearing through different concept IDs, value formats, and operators across sites—confirms that Tier 2 feasibility must include detailed profiling of specific measurements critical to primary outcomes. Given the absence of diagnostic packages for measurement profiling, the coordinating center recommends prioritizing key study measurements (e.g., PD-L1, Body Mass Index (BMI), stage) rather than attempting comprehensive characterization of all potential biomarkers.



### 7.3.2 Two-Phase Study Design Framework

Based on IQVIA operational team experience, a two-phase approach is recommended for network studies:

Phase 1 - Comprehensive Characterization (3-4 months):

- Execute full cohort identification with all criteria
- Profile available variables at measurement level detail
- Identify and develop proxy definitions where needed
- Establish site-specific adaptations required
- Document actual versus expected data availability
- Create reusable cohort definitions for Phase 2

Phase 2 - Comparative Effectiveness or Advanced Analytics (3-6 months):

- Leverage characterized cohorts from Phase 1
- Apply sophisticated analytical methods
- Focus on specific research questions
- Utilize established variable definitions

The IQVIA operational team noted that this approach would have prevented the late discovery of missing variables (smoking status, ECOG, linkage) that complicated analytical package development. Furthermore, this phased approach aligns with funding realities, allowing demonstration of feasibility before committing to complex analyses.

### 7.3.3 Strategic Data Enhancement Initiatives

Based on documented data gaps, we recommend targeted enhancement strategies for critical oncology variables.

#### **Biomarker Data Capture Enhancement:**

Biomarker capture (absent across all sites) requires consortium-wide initiatives:

- Development of standardized biomarker documentation templates that can be integrated into existing EHR systems, ensuring structured capture at the point of care rather than relying on retrospective extraction from clinical notes
- Creation of OMOP-compatible biomarker vocabularies that accommodate both standard markers (PD-L1, EGFR, ALK) and emerging markers, with clear versioning to track evolution of testing methods and thresholds
- Implementation of automated extraction pipelines for laboratories already performing biomarker testing, addressing the current disconnect between laboratory information systems and clinical databases



- Establishment of minimum biomarker documentation standards for oncology studies, ensuring that critical markers are consistently captured across all sites

**Performance Status Systematic Documentation:**

Performance status (<20% capture) requires:

- Integration of ECOG performance status as a required field in oncology clinic note templates, transforming it from optional narrative documentation to structured data capture
- Development of automated prompts in EHR systems that trigger performance status documentation at key clinical events (diagnosis, treatment initiation, progression)
- Training programs for clinical staff emphasizing the importance of consistent performance status documentation for both clinical care and research purposes
- Validation studies comparing documented performance status with other functional indicators to ensure data quality

**Patient-Reported Outcomes Infrastructure Development:**

PRO infrastructure requires prospective development:

- Selection of validated PRO instruments compatible with OMOP CDM structure, prioritizing instruments with established mappings to standard vocabularies
- Implementation of digital PRO collection platforms that integrate with existing clinical systems, enabling real-time capture without disrupting clinical workflows
- Development of PRO collection protocols that balance research needs with patient burden, potentially using adaptive sampling strategies
- Creation of OMOP CDM extensions specifically designed for PRO data, addressing the unique characteristics of patient-generated data

**7.4 Capacity Building Recommendations**

**7.4.1 Comprehensive Three-Year Capability Development Framework**

Based on Norway's empirical report that "OMOP skills have been established in the data warehouse team over the last 3 years," we propose a structured capability development program that acknowledges the extended timeline required for sustainable implementation.

Table 5: Structured Capability Development Roadmap

Phase	Timeline	Technical Competencies	Organizational Capabilities	Deliverables	Success Metrics
<b>Foundation Phase</b>	Months 1-12	<ul style="list-style-type: none"> <li>• OMOP CDM fundamentals</li> <li>• Basic SQL for OMOP</li> </ul>	<ul style="list-style-type: none"> <li>• Team formation</li> <li>• Governance structure</li> </ul>	<ul style="list-style-type: none"> <li>• Pilot disease area mapped</li> <li>• Initial data quality report</li> </ul>	<ul style="list-style-type: none"> <li>• One disease area in OMOP</li> <li>• 80% team trained</li> </ul>



Phase	Timeline	Technical Competencies	Organizational Capabilities	Deliverables	Success Metrics
		<ul style="list-style-type: none"> <li>• Vocabulary navigation</li> <li>• Simple ETL patterns</li> <li>• Data quality basics</li> </ul>	<ul style="list-style-type: none"> <li>• Resource allocation</li> <li>• Stakeholder engagement</li> </ul>	<ul style="list-style-type: none"> <li>• Basic cohort definitions</li> <li>• Team training completed</li> </ul>	<ul style="list-style-type: none"> <li>• First quality assessment</li> <li>• Governance established</li> </ul>
<b>Expansion Phase</b>	Months 13-24	<ul style="list-style-type: none"> <li>• Complex ETL development</li> <li>• Multi-source integration</li> <li>• Advanced cohort building</li> <li>• Achilles implementation (6, 7)</li> <li>• Network study basics</li> </ul>	<ul style="list-style-type: none"> <li>• Cross-functional coordination</li> <li>• Change management</li> <li>• Documentation standards</li> <li>• Knowledge management</li> </ul>	<ul style="list-style-type: none"> <li>• 3-5 disease areas mapped</li> <li>• Automated quality checks</li> <li>• Participation in network study</li> <li>• Internal documentation</li> </ul>	<ul style="list-style-type: none"> <li>• Multiple therapeutic areas</li> <li>• &lt;5% data quality issues</li> <li>• One network study</li> <li>• Reduced mapping time</li> </ul>
<b>Maturation Phase</b>	Months 25-36	<ul style="list-style-type: none"> <li>• Performance optimization</li> <li>• Custom analytics</li> <li>• Methods development</li> <li>• Troubleshooting expertise</li> <li>• Knowledge transfer</li> </ul>	<ul style="list-style-type: none"> <li>• Sustainable operations</li> <li>• Continuous improvement</li> <li>• External collaboration</li> <li>• Innovation capacity</li> </ul>	<ul style="list-style-type: none"> <li>• Full OMOP implementation</li> <li>• Reusable components library</li> <li>• Training materials</li> <li>• Network contributions</li> </ul>	<ul style="list-style-type: none"> <li>• Self-sufficient operations</li> <li>• Leading network studies</li> <li>• Training other sites</li> <li>• Methods contributions</li> </ul>

### Detailed Implementation Guidance:

The Foundation Phase (Year 1) should focus on a single disease area with protected team time (minimum 50% allocation) and regular connection with experienced sites.

The Expansion Phase (Year 2) shifts focus from learning to execution, emphasizing automation and scalability.

The Maturation Phase (Year 3) transforms sites from consumers to contributors, exemplified by Norway's ability to identify technical challenges and propose network solutions.



## 7.4.2 Comprehensive Knowledge Management Systems

The opportunity for reuse identified by Norway—"We were able to reuse work from previous studies"—requires systematic knowledge capture and sharing mechanisms.

### **Design Pattern Library Development:**

A centralized, searchable repository of OMOP implementation artifacts should be established containing:

- **ETL Specifications:** Detailed documentation of extract, transform, and load logic for common scenarios, tagged by source system type, therapeutic area, and data types. Each specification should include source-to-concept mappings, transformation rules, handling of edge cases, and known limitations
- **Cohort Definitions:** Validated cohort definitions for common conditions, treatments, and outcomes, with clear documentation of inclusion/exclusion logic, concept sets used, and validation results from multiple sites
- **Concept Mappings:** Curated mappings from local vocabularies to OMOP standard concepts, including mapping rationale, assumptions made, and situations where one-to-many or many-to-one mappings are required
- **Quality Check Protocols:** Standardized approaches for validating ETL output, including expected distributions, plausibility checks, and cross-site comparison methods

### **Lessons Learned Database Structure:**

Beyond static documentation, sites need access to experiential knowledge. We recommend establishing a structured lessons learned database containing:

- **Technical Solutions:** Specific fixes for identified problems (like Norway's postgres/R compatibility issue), including problem description, root cause analysis, solution implemented, and generalizability assessment
- **Process Improvements:** Workflow optimizations discovered during implementation, with before/after comparisons and effort savings quantified
- **Pitfalls and Warnings:** Common mistakes and how to avoid them, particularly those that may not be caught by standard quality checks
- **Resource Estimates:** Actual versus planned timelines and effort for different implementation components, enabling more accurate project planning

### **Variable Survey Optimization:**

**The coordinating center identified that variable surveys arrived "quite later on after we had coded out a lot of the package," resulting in redundant development and required revisions. Based on this experience, we recommend:**

Variable surveys must be completed during Phase 1 characterization, before any analytical package development

Surveys should distinguish between "available in source" versus "available in OMOP" versus "could be mapped with effort"

For each critical measurement, sites must specify: concept IDs used, value formats (numeric/categorical), operators, units, and known limitations

Survey responses should be verified through test queries before package development begins

## 7.5 Network Coordination Recommendations

### 7.5.1 Comprehensive Vocabulary Governance Framework

The vocabulary versioning challenge identified by Norway—requiring consortium-wide coordination—demands a formal governance structure that balances standardization needs with site autonomy.

Table 6: Vocabulary Governance Structure

Component	Current State	Recommended Approach	Implementation Requirements	Expected Outcomes
<b>Version Management</b>	<ul style="list-style-type: none"> <li>• Sites on different versions</li> <li>• Uncoordinated updates</li> <li>• Version conflicts discovered late</li> </ul>	<ul style="list-style-type: none"> <li>• Quarterly synchronized updates</li> <li>• 6-week testing window</li> <li>• Formal version release notes</li> <li>• Rollback procedures</li> </ul>	<ul style="list-style-type: none"> <li>• Central coordination role</li> <li>• Testing environments</li> <li>• Version tracking system</li> <li>• Communication protocols</li> </ul>	<ul style="list-style-type: none"> <li>• Network-wide consistency</li> <li>• Predictable update schedule</li> <li>• Reduced conflicts</li> <li>• Clear audit trail</li> </ul>
<b>Custom Concepts</b>	<ul style="list-style-type: none"> <li>• Site-specific mappings</li> <li>• No sharing mechanism</li> <li>• Duplicate effort across sites</li> </ul>	<ul style="list-style-type: none"> <li>• Centralized custom concept repository</li> <li>• Review and approval process</li> <li>• Promotion to standard pathway</li> <li>• Attribution and tracking</li> </ul>	<ul style="list-style-type: none"> <li>• Governance committee</li> <li>• Submission templates</li> <li>• Quality criteria</li> <li>• Repository infrastructure</li> </ul>	<ul style="list-style-type: none"> <li>• Reduced duplication</li> <li>• Quality improvement</li> <li>• Knowledge sharing</li> <li>• Eventual standardization</li> </ul>
<b>Quality Assurance</b>	<ul style="list-style-type: none"> <li>• Site-specific validation</li> </ul>	<ul style="list-style-type: none"> <li>• Standardized validation suite</li> </ul>	<ul style="list-style-type: none"> <li>• Validation scripts</li> <li>• Reporting templates</li> </ul>	<ul style="list-style-type: none"> <li>• Consistent quality</li> <li>• Early issue detection</li> </ul>



Component	Current State	Recommended Approach	Implementation Requirements	Expected Outcomes
	<ul style="list-style-type: none"> <li>No cross-site comparison</li> <li>Inconsistent approaches</li> </ul>	<ul style="list-style-type: none"> <li>Cross-site consistency checks</li> <li>Automated testing pipelines</li> <li>Regular audits</li> </ul>	<ul style="list-style-type: none"> <li>Dashboard development</li> <li>Audit protocols</li> </ul>	<ul style="list-style-type: none"> <li>Continuous improvement</li> <li>Trust in data</li> </ul>
<b>Change Management</b>	<ul style="list-style-type: none"> <li>Ad hoc communication</li> <li>Variable implementation</li> <li>Impact assessment gaps</li> </ul>	<ul style="list-style-type: none"> <li>Formal change process</li> <li>Impact analysis requirements</li> <li>Staged rollout protocols</li> <li>Success metrics</li> </ul>	<ul style="list-style-type: none"> <li>Change control board</li> <li>Impact templates</li> <li>Communication plan</li> <li>Success criteria</li> </ul>	<ul style="list-style-type: none"> <li>Managed transitions</li> <li>Minimal disruption</li> <li>Learning capture</li> <li>Continuous refinement</li> </ul>

### Implementation Process:

Implementation requires version control systems, automated testing pipelines, designated vocabulary stewards at each site, and a network-level governance committee.

Quarterly synchronization follows a structured three-month cycle: release and testing (Month 1), resolution (Month 2), and final implementation (Month 3).

### 7.5.2 Enhanced Communication and Collaboration Structures

Formal communication structures are needed to capture and disseminate technical knowledge. The coordinating center's experience provides specific validated solutions for effective collaboration.

#### Technical Working Group Framework:

Establish role-specific working groups that meet regularly to address implementation challenges:

- ETL Developers Group:** Monthly sessions focusing on technical challenges, code reviews, and pattern sharing. Norway's postgres/bigint issue would be discussed here, with solutions documented and shared
- Data Quality Group:** Quarterly reviews of quality metrics, validation approaches, and cross-site consistency. Focus on translating DataQualityDashboard outputs into actionable improvements
- Analytics Group:** Bi-monthly sessions on analytical methods, particularly addressing Denmark's need for "inferential and causal analytics" in federated settings
- Governance Group:** Quarterly meetings bridging technical and regulatory domains, addressing the knowledge gaps evident in ethics and permit requirements



Each group should maintain action logs and solution repositories with rotating site leadership.

#### **Enhanced Site-Level Collaboration:**

The IQVIA operational team experience revealed that dedicated collaboration infrastructure significantly improved execution efficiency. Based on Norway's successful innovation, we recommend establishing formal communication channels for each site that move beyond email to enable real-time technical support and persistent documentation of solutions. The transition from asynchronous to synchronous communication reduced issue resolution time from days to hours.

#### **Essential Team Composition:**

The IQVIA team experience is that successful network participation requires defined roles at each site:

- **Project coordination:** A designated point of contact with authority to engage team members and manage timelines
- **Data expertise:** Deep knowledge of source systems, mapping decisions, and data limitations
- **Technical execution:** Capability to run packages, interpret errors, and provide debugging feedback
- **Clinical oversight:** Domain expertise for protocol decisions and result interpretation

The absence of any role created bottlenecks, as noted by the coordinating center: "Once we found someone who would respond, we just went to them," leading to overwhelming single individuals rather than distributed expertise. Clear role definition at study initiation prevents this dysfunction while respecting sites' autonomy in specific staffing decisions.

#### **Regular Progress Verification:**

The IQVIA operational team reported periods where "communication would drop off at points and we thought progress was happening when it wasn't." This experience mandates establishing regular touchpoints between coordinating centers and sites to verify progress versus assumptions, identify emerging issues before they become critical, and maintain project momentum across distributed teams.

## **7.6 Methodological Advancement Recommendations**

### **7.6.1 Comprehensive Framework for Advanced Federated Analytics**

Denmark's articulated need for "developing more advanced methods for conducting federated analysis" and "improving analytical capabilities into inferential and causal analytics" identifies a critical gap between current descriptive capabilities and scientific ambitions.



Table 7: Methodological Development Roadmap

Analytical Capability	Current State	Development Requirements	Technical Approach	Timeline	Success Indicators
<b>Descriptive Statistics</b>	<ul style="list-style-type: none"> <li>Fully operational</li> <li>Basic aggregations work</li> <li>Simple stratification possible</li> </ul>	<ul style="list-style-type: none"> <li>Documentation of current capabilities</li> <li>Standardization across sites</li> <li>Efficiency improvements</li> </ul>	<ul style="list-style-type: none"> <li>Optimize existing scripts</li> <li>Create reusable templates</li> <li>Automate report generation</li> </ul>	Immediate	<ul style="list-style-type: none"> <li>Consistent results across sites</li> <li>Reduced analysis time</li> <li>Automated reporting</li> </ul>
<b>Adjusted Analyses</b>	<ul style="list-style-type: none"> <li>Limited capability</li> <li>Site-level confounding only</li> <li>No individual adjustment</li> </ul>	<ul style="list-style-type: none"> <li>Statistical method development</li> <li>Validation frameworks</li> <li>Software implementation</li> </ul>	<ul style="list-style-type: none"> <li>Distributed regression methods</li> <li>Meta-analysis approaches</li> <li>Privacy-preserving techniques</li> </ul>	6-12 months	<ul style="list-style-type: none"> <li>Validated adjustment methods</li> <li>Published methodology</li> <li>Network adoption</li> </ul>
<b>Propensity Scores</b>	<ul style="list-style-type: none"> <li>Not currently possible</li> <li>Requires patient-level data</li> <li>Privacy constraints</li> </ul>	<ul style="list-style-type: none"> <li>Novel federated PS methods</li> <li>Simulation studies</li> <li>Validation against pooled data</li> </ul>	<ul style="list-style-type: none"> <li>Distributed PS calculation</li> <li>Federated matching algorithms</li> <li>Balance diagnostics</li> </ul>	12-18 months	<ul style="list-style-type: none"> <li>Functioning PS pipeline</li> <li>Validation publications</li> <li>Case studies completed</li> </ul>
<b>Causal Inference</b>	<ul style="list-style-type: none"> <li>Not implemented</li> <li>Methods underdeveloped</li> <li>Expertise limited</li> </ul>	<ul style="list-style-type: none"> <li>Theoretical framework</li> <li>Methods development</li> <li>Training programs</li> <li>Validation studies</li> </ul>	<ul style="list-style-type: none"> <li>Instrumental variables</li> <li>Difference-in-differences</li> <li>Regression discontinuity</li> <li>Target trial emulation(8)</li> </ul>	18-24 months	<ul style="list-style-type: none"> <li>Published methods</li> <li>Successful applications</li> <li>Network expertise</li> <li>External recognition</li> </ul>
<b>Machine Learning</b>	<ul style="list-style-type: none"> <li>Not feasible currently</li> <li>Requires pooled data</li> <li>Privacy barriers</li> </ul>	<ul style="list-style-type: none"> <li>Federated learning methods (9)</li> <li>Infrastructure development</li> <li>Validation frameworks</li> </ul>	<ul style="list-style-type: none"> <li>Distributed gradient boosting</li> <li>Federated neural networks</li> <li>Privacy-preserving ML</li> </ul>	24-36 months	<ul style="list-style-type: none"> <li>ML models deployed</li> <li>Performance validated</li> <li>Publications</li> <li>Tool availability</li> </ul>

**Detailed Development Strategy:**



Progression from descriptive to causal methods requires systematic capability building, starting with distributed regression techniques.

Propensity score methods require innovation including distributed calculation with secure aggregation, federated matching algorithms, and validation studies comparing federated versus pooled approaches.

Causal inference methods must be adapted for the constraints of federated analysis. Target trial emulation frameworks appear particularly promising, as they can be implemented through careful cohort definition without requiring complex statistical machinery (8). Instrumental variable approaches may be feasible when instruments can be defined at the site level. These methods require not just technical development but also training in causal thinking and study design.

### 7.6.2 Scientific Infrastructure Development

Beyond methods development, successful advancement requires supporting infrastructure:

#### **Statistical Support Core Establishment:**

A centralized statistical support function should provide:

- **Consultation Services:** Expert guidance on study design, analytical approaches, and interpretation for network studies, addressing the current gap where sites lack advanced analytical expertise
- **Methods Development:** Dedicated resources for developing and validating new federated analytical methods, with focus on practical implementation rather than purely theoretical advancement
- **Training Programs:** Regular workshops on advanced methods, causal inference, and federated analytics, building network-wide capacity rather than relying on site-specific expertise
- **Quality Review:** Independent review of analytical plans and results for network studies, ensuring methodological rigor and appropriate interpretation

#### **Methods Validation Framework:**

Establishing credibility for federated methods requires systematic validation:

- **Simulation Studies:** Comprehensive evaluation of method performance under various scenarios, including different types of confounding, missing data patterns, and heterogeneity
- **Comparative Studies:** Direct comparison of federated versus pooled analyses using datasets where both are possible, quantifying any systematic differences
- **Sensitivity Analyses:** Standard approaches for assessing robustness of federated results to assumptions and analytical choices

- **Publication Standards:** Development of reporting guidelines specific to federated analyses, ensuring transparency about limitations and assumptions

## 7.7 Implementation Prioritization Strategy

### 7.7.1 Immediate Priorities (0-3 months)

The most urgent actions focus on capturing and sharing existing knowledge before it is lost:

**Knowledge Preservation Initiative:** Norway's reusable ETL specifications represent valuable intellectual property that should be immediately documented and shared. This includes creating detailed documentation of design decisions, mapping rationales, and edge case handling. The specifications should be packaged with clear implementation guides, dependency requirements, and known limitations. Version control should be established to track evolution and improvements.

**Technical Environment Assessment Protocol:** Based on Norway's infrastructure challenges, develop a comprehensive checklist for technical environment assessment including:

- Database system specifications and known compatibility issues, particularly data type limitations and character encoding
- Analytical environment requirements including R/Python versions, package dependencies, and memory requirements
- Network and security configurations that may affect federated analysis execution
- Performance benchmarks and optimization opportunities

**Cross-functional Team Formation:** The governance knowledge gaps require immediate establishment of integrated teams including:

- Technical leads responsible for implementation architecture and troubleshooting
- Clinical experts ensuring clinical validity of mappings and cohort definitions
- Regulatory specialists managing ethics and data protection requirements
- Project managers coordinating across domains and managing dependencies

### 7.7.2 Short-term Priorities (3-12 months)

These priorities focus on establishing sustainable processes and addressing identified gaps:

Table 8: Short-term Implementation Plan

Priority Area	Specific Actions	Responsible Parties	Success Metrics	Risk Mitigation
<b>Vocabulary Synchronization</b>	<ul style="list-style-type: none"> <li>• Establish governance committee</li> </ul>	<ul style="list-style-type: none"> <li>• Network coordinator</li> </ul>	<ul style="list-style-type: none"> <li>• Committee formed</li> </ul>	<ul style="list-style-type: none"> <li>• Phased rollout</li> </ul>



Priority Area	Specific Actions	Responsible Parties	Success Metrics	Risk Mitigation
	<ul style="list-style-type: none"> <li>• Define synchronization schedule</li> <li>• Develop testing protocols</li> <li>• Create rollback procedures</li> </ul>	<ul style="list-style-type: none"> <li>• Site vocabulary stewards</li> <li>• Technical working group</li> </ul>	<ul style="list-style-type: none"> <li>• Schedule published</li> <li>• First synchronized update</li> <li>• &lt;5% version conflicts</li> </ul>	<ul style="list-style-type: none"> <li>• Extensive testing</li> <li>• Clear communication</li> <li>• Rollback capability</li> </ul>
<b>Regulatory Guidance</b>	<ul style="list-style-type: none"> <li>• Document country requirements</li> <li>• Create template applications</li> <li>• Establish timelines</li> <li>• Share approval strategies</li> </ul>	<ul style="list-style-type: none"> <li>• Country regulatory leads</li> <li>• Legal advisors</li> <li>• Ethics specialists</li> </ul>	<ul style="list-style-type: none"> <li>• Guides published</li> <li>• Templates available</li> <li>• Timelines documented</li> <li>• Approval success &gt;90%</li> </ul>	<ul style="list-style-type: none"> <li>• Legal review</li> <li>• Regular updates</li> <li>• Country variations</li> <li>• Change tracking</li> </ul>
<b>Capability Development</b>	<ul style="list-style-type: none"> <li>• Design curricula</li> <li>• Identify trainers</li> <li>• Create materials</li> <li>• Schedule programs</li> </ul>	<ul style="list-style-type: none"> <li>• Education committee</li> <li>• Experienced sites</li> <li>• External experts</li> </ul>	<ul style="list-style-type: none"> <li>• Curricula approved</li> <li>• Materials developed</li> <li>• First cohort trained</li> <li>• Competency assessed</li> </ul>	<ul style="list-style-type: none"> <li>• Multiple modalities</li> <li>• Practical focus</li> <li>• Ongoing support</li> <li>• Regular updates</li> </ul>
<b>Technical Working Groups</b>	<ul style="list-style-type: none"> <li>• Charter groups</li> <li>• Recruit members</li> <li>• Establish schedules</li> <li>• Create repositories</li> </ul>	<ul style="list-style-type: none"> <li>• Network coordinator</li> <li>• Domain experts</li> <li>• Site representatives</li> </ul>	<ul style="list-style-type: none"> <li>• Groups chartered</li> <li>• Regular meetings</li> <li>• Knowledge captured</li> <li>• Solutions shared</li> </ul>	<ul style="list-style-type: none"> <li>• Clear scope</li> <li>• Active facilitation</li> <li>• Action tracking</li> <li>• Value demonstration</li> </ul>

### 7.7.3 Long-term Investments (12+ months)

Strategic investments for sustainable network development:

**Advanced Analytics Platform Development:** Building on Denmark's identified needs, establish a comprehensive advanced analytics platform including:

- Development environment for federated methods with secure testing capabilities
- Validation datasets enabling method comparison and benchmarking



Funded by the  
Nordic Council  
of Ministers

Lessons learned from multi-site federated analysis – VALO Value from Nordic Health Data 46

- Production environment with appropriate security and privacy controls
- Documentation and training resources for method implementation

**Clinical Documentation Enhancement Program:** Addressing Tier 3 data gaps requires long-term investment in clinical documentation:

- Stakeholder engagement to build clinical buy-in for documentation improvements
- EHR template modifications to capture structured data at point of care
- Training programs for clinical staff on documentation importance
- Quality monitoring and feedback loops to ensure sustained improvement
- Incentive structures that reward complete documentation

**Sustainable Network Funding Model:** Long-term sustainability requires diversified funding including:

- Core infrastructure support from government or foundation sources
- Fee-for-service model for commercial studies leveraging network capabilities
- Grant funding for methods development and innovation
- In-kind contributions from participating sites
- Intellectual property frameworks that incentivize contribution while enabling sharing



## 8 Conclusion

### 8.1 Synthesis of Key Findings

This comprehensive assessment of OMOP CDM implementation across the Nordic region provides crucial empirical evidence about the realities of establishing federated research networks in heterogeneous healthcare environments. Through systematic analysis of experiences from ten respondents across five countries—including three active data partners and two observer sites—we have identified fundamental patterns that transcend individual site experiences and reveal systemic challenges and opportunities in federated network development. The integration of IQVIA operational team perspectives, captured through structured interviews with the team responsible for project management, analytical package development and execution, provides essential validation and operational context for site-reported experiences. This dual perspective—combining site-level strategic views with coordinating center execution realities—offers unprecedented insight into both the aspirational and operational dimensions of federated network implementation.

The journey from OMOP concept to operational capability emerges as a multi-year organizational transformation, with Norway's three-year capability development timeline providing essential benchmarks for future implementations.

The three-tier data availability hierarchy identified across sites fundamentally constrains study feasibility, with critical elements for precision oncology (biomarkers, performance status, PROs) systematically absent from routine documentation.

### 8.2 Transformative Insights for Network Development

The disconnect between technical excellence and infrastructure readiness represents perhaps the most actionable insight from this assessment. Norway's specific challenges with postgres bigint handling in R, local environment configuration issues, and memory limitations reveal that sites struggle not with OMOP concepts but with technical implementation details. This distinction fundamentally reframes support needs from conceptual training to technical troubleshooting and infrastructure optimization. The success of sites with established clinical data warehouses in making OMOP "part of regular operations" further emphasizes that infrastructure alignment, not just infrastructure availability, determines implementation success.

Governance knowledge gaps—with ethics requirements uncertain even among experienced teams—underscore the need for cross-functional teams bridging technical, clinical, and regulatory domains.

Network coordination challenges, including vocabulary versioning, require consortium-level solutions that elevate coordination from administrative function to scientific necessity.

#### 8.2.1 Strategic Implications for Healthcare Research

The evolution from descriptive to inferential analytical capabilities represents the next frontier for federated networks, requiring methodological innovation in privacy-preserving approaches to causal inference.



Observer site participation (Sweden, Iceland) demonstrates the value of inclusive network development that enables learning before commitment.

### 8.2.2 Contributions to the Field

This assessment makes several distinct contributions to the growing literature on federated research networks and common data models:

**Empirical Quantification of Implementation Timelines:** The three-year capability development timeline reported by Norway provides the first empirical benchmark for institutional planning, replacing optimistic projections with evidence-based expectations. This finding has immediate implications for funding models, project planning, and stakeholder expectation management.

**Systematic Categorization of Data Readiness:** The three-tier hierarchy of data availability provides a framework for feasibility assessment that can be applied across therapeutic areas and research questions. This categorization enables more realistic study planning and highlights systematic gaps requiring strategic intervention rather than technical solutions.

**Identification of Infrastructure Versus Conceptual Challenges:** The distinction between technical infrastructure challenges and OMOP conceptual understanding reframes support strategies and resource allocation. This finding suggests that successful networks require technical support teams with deep infrastructure expertise rather than solely focusing on OMOP education.

**Documentation of Network Coordination Requirements:** The vocabulary versioning challenge and its proposed solution exemplify network-level requirements that emerge only through multi-site implementation. These findings inform governance structures and coordination mechanisms for emerging networks.

### 8.2.3 Limitations and Future Directions

While this assessment provides valuable insights, several limitations must be acknowledged. The single time-point survey captures a snapshot of experiences that continue to evolve. The separation between survey respondents and hands-on implementers may have resulted in undocumented technical lessons. The small sample size, while representing 100% of consortium countries, limits statistical analysis and generalizability beyond Nordic contexts.

Future research should pursue longitudinal assessment of network maturation, tracking how challenges and capabilities evolve over time. Comparative studies across different federated networks would identify which findings are universal versus context-specific. Development and validation of objective maturity models would enable systematic assessment of network readiness for different study types. Economic evaluation of implementation approaches would inform optimal resource allocation strategies. Most critically, correlation of implementation approaches with scientific outputs would establish evidence-based best practices.



### 8.2.4 The Path Forward

The transformation of healthcare data into scientific evidence through federated networks represents one of the most promising yet challenging frontiers in medical research. This assessment reveals that success requires simultaneous advancement across multiple dimensions: technical infrastructure properly configured for OMOP requirements; organizational capability developed over multiple years; governance structures that bridge technical and regulatory domains; network coordination mechanisms that ensure semantic consistency; methodological innovation that enables causal inference; and clinical engagement that improves documentation completeness.

The lessons learned provide a roadmap for accelerating adoption: systematic knowledge capture enables reuse, documented technical solutions enable preemptive problem-solving, and quantified timelines enable realistic planning.

### 8.3 Final Reflections

The establishment of federated research networks using OMOP CDM represents a fundamental shift in how real-world evidence is generated, moving from isolated institutional studies to coordinated multi-site investigations. This assessment demonstrates that while technical challenges are substantial, they are surmountable through systematic approach, knowledge sharing, and sustained investment. The greater challenges lie in organizational transformation, governance integration, and methodological advancement—challenges that require strategic leadership, institutional commitment, and network-level coordination.

The Nordic experience, with its combination of successful implementation, identified challenges, and proposed solutions, provides valuable intelligence for the global community pursuing federated research capabilities. The active participation of observer sites alongside implementing sites creates a learning ecosystem that accelerates capability development across the network. Completing analyses in three months compared to two-year requirement in traditional studies demonstrates that despite documented challenges, the federated approach delivers substantial efficiency gains once operational. Their practical innovations, from Teams-based collaboration to single-codebase site adaptations, provide tested solutions that complement sites' strategic insights. The candid acknowledgment of limitations, from missing biomarker data to governance uncertainties, demonstrates the maturity to confront rather than minimize implementation challenges.

As healthcare systems globally grapple with the imperative to transform clinical data into scientific evidence, these lessons gain urgency. Though implementation timelines will differ substantially based on starting infrastructure and institutional readiness, the consistent finding across sites that capability development extends beyond initial projections argues for realistic planning and early engagement. The systematic data gaps identified require strategic enhancement initiatives that extend beyond technical solutions. The network coordination challenges demand governance structures that balance standardization with site autonomy. The methodological limitations necessitate investment in analytical innovation that preserves privacy while enabling sophisticated analyses.

The vision of comprehensive, real-time evidence generation from routine clinical care remains compelling despite the challenges documented in this assessment. The path from



Funded by the  
Nordic Council  
of Ministers

## Lessons learned from multi-site federated analysis – VALO Value from Nordic Health Data 50

vision to reality, while longer and more complex than initially anticipated, is now better mapped through the empirical experiences of early adopters. The Nordic OMOP network, through its willingness to systematically document and share both successes and struggles, provides an invaluable service to the global research community. Their experiences transform OMOP implementation from an uncertain journey into a mapped expedition, with known challenges, proven solutions, and realistic timelines.

This assessment concludes that successful OMOP implementation and federated network participation requires recognition of the true scope of the undertaking: not merely technical implementation but organizational transformation; not just data standardization but documentation enhancement; not simply descriptive statistics but methodological innovation; not individual excellence but network coordination. With these lessons learned, future implementations can proceed with clarity about requirements, confidence in eventual success, and commitment to the sustained investment necessary for transformation. The journey from fragmented healthcare data to integrated scientific evidence remains challenging, but it is now illuminated by the experience of those who have walked this path before.



## 9 References

1. Trinh NT, Houghtaling J, Bernal FL, Hayati S, Maglanoc LA, Lupattelli A, et al. Harmonizing Norwegian registries onto OMOP common data model: Mapping challenges and opportunities for pregnancy and COVID-19 research. *International Journal of Medical Informatics*. 2024;191:105602.
2. Voss EA, Blacketer C, van Sandijk S, Moinat M, Kallfelz M, Van Speybroeck M, et al. European Health Data & Evidence Network—learnings from building out a standardized international health data network. *Journal of the American Medical Informatics Association*. 2024;31(1):209-19.
3. Belenkaya R, Gurley MJ, Golozar A, Dymshyts D, Miller RT, Williams AE, et al. Extending the OMOP common data model and standardized vocabularies to support observational cancer research. *JCO Clinical Cancer Informatics*. 2021;5:12-20.
4. Reich C, Ostropolets A, Ryan P, Rijnbeek P, Schuemie M, Davydov A, et al. OHDSI Standardized Vocabularies—a large-scale centralized reference ontology for international data harmonization. *Journal of the American Medical Informatics Association*. 2024;31(3):583-90.
5. Öjlert Å, Bhatnagar P, Hendriks L, Ogliari F, Acker F, Bolton E, et al. DINASTY in NSCLC—a multicenter retrospective study on real-world data using federated analysis. *medRxiv*. 2025:2025.02. 22.25321981.
6. Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. *Egems*. 2016;4(1):1244.
7. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Studies in health technology and informatics*. 2015;216:574.
8. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *American journal of epidemiology*. 2016;183(8):758-64.
9. Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. *NPJ digital medicine*. 2020;3(1):119.

## Appendices

### 9.1 Appendix A: Survey Structure and Content

Section	Questions	Focus Areas	Question Types
Background	1-5	Respondent identification, Country/role, OMOP/RWE experience	Open text, Multiple choice
Data Governance	6-13	Contracting, Ethics, Data permits, OMOP governance	Yes/No/Not known, Open text
Data Life Cycle	14-20	Data sources, Linkage, Coding systems, Availability	Checkbox, Matrix
Data Processing	21-24	Variables, Script execution, Challenges	Checkbox, Ratings
VALO Study Process	25-32	ETL feedback, QA/QC, Package execution	Open text
Future Investments	33-36	Infrastructure, Capabilities, Recommendations	Open text

### 9.2 Appendix B: Comprehensive Respondent Characteristics (Questions 1-5)

Country	N	Professional Roles and Responsibilities	OMOP/RWE Experience	Previous Federated Network Participation
Norway	3	<ul style="list-style-type: none"> <li>Data analyst with direct OMOP implementation experience</li> <li>Data analyst/statistician focusing on analytical frameworks</li> <li>Data manager overseeing data governance and infrastructure</li> </ul>	<ul style="list-style-type: none"> <li>2.5 years documented experience with OMOP CDM implementation (1 respondent)</li> <li>Experience level not specified (2 respondents)</li> </ul>	<ul style="list-style-type: none"> <li>Confirmed participation in previous federated networks (2 respondents)</li> <li>One respondent did not specify prior experience</li> </ul>
Denmark	2	<ul style="list-style-type: none"> <li>Clinical researcher with focus on RWE generation</li> <li>Clinical researcher involved in study design and execution</li> </ul>	<ul style="list-style-type: none"> <li>Both respondents did not specify exact years</li> </ul>	<ul style="list-style-type: none"> <li>Previous participation not documented</li> </ul>
Finland	2	<ul style="list-style-type: none"> <li>Clinical researcher bridging clinical and technical aspects</li> <li>Chief medical officer for clinical auditing at HUS</li> </ul>	<ul style="list-style-type: none"> <li>Specific years not provided</li> <li>Senior positions evident from roles</li> </ul>	<ul style="list-style-type: none"> <li>Not specified by either respondent</li> </ul>
Sweden - Karolinska	1	<ul style="list-style-type: none"> <li>Precision Cancer Medicine Researcher and Translation</li> </ul>	<ul style="list-style-type: none"> <li>Years not quantified</li> </ul>	<ul style="list-style-type: none"> <li>Not documented</li> </ul>



Country	N	Professional Roles and Responsibilities	OMOP/RWE Experience	Previous Federated Network Participation
<b>Institutet (KI)</b>		and Health Data Lead at SciLifeLab		
<b>Sweden - Karolinska University Hospital (KUH)</b>	1	<ul style="list-style-type: none"> <li>Senior health data expert</li> </ul>	<ul style="list-style-type: none"> <li>Years not quantified</li> </ul>	<ul style="list-style-type: none"> <li>Not documented</li> </ul>
<b>Iceland</b>	1	<ul style="list-style-type: none"> <li>Data manager responsible for data infrastructure planning</li> </ul>	<ul style="list-style-type: none"> <li>Experience years not specified</li> </ul>	<ul style="list-style-type: none"> <li>Explicitly indicated no previous federated network participation</li> </ul>

### 9.3 Appendix C: Challenge Rating Standardization

Descriptive Rating	Level	Interpretation
Not challenging / Routine process	1	No significant obstacles
Slightly challenging / Minor adjustments	2	Manageable with internal resources
Moderately challenging / Significant effort	3	Required dedicated effort
Challenging / Required external support	4	Substantial obstacles
Extremely challenging / Major obstacles	5	Critical barriers
Not in my remit	N/A	Outside respondent's scope

### 9.4 Appendix D: Ethics and Regulatory Approval Landscape (Questions 7-9)

Country	Ethics Committee Approval Required	Approval Timeline	Separate Data Permit Required	Permit Timeline
<b>Norway</b>	<ul style="list-style-type: none"> <li>Not known (2 respondents)</li> <li>Yes (1 respondent)</li> </ul>	<ul style="list-style-type: none"> <li>Not known (2 respondents)</li> <li>Timeline not specified (1 respondent)</li> </ul>	<ul style="list-style-type: none"> <li>Not known (all 3 respondents)</li> </ul>	<ul style="list-style-type: none"> <li>Timeline not specified</li> </ul>
<b>Denmark</b>	<ul style="list-style-type: none"> <li>Requirements not specified</li> </ul>	<ul style="list-style-type: none"> <li>No timeline information provided</li> </ul>	<ul style="list-style-type: none"> <li>Permit requirements not documented</li> </ul>	<ul style="list-style-type: none"> <li>Timeline not specified</li> </ul>
<b>Finland</b>	<ul style="list-style-type: none"> <li>Not specified by either respondent</li> </ul>	<ul style="list-style-type: none"> <li>Approval timeline not documented</li> </ul>	<ul style="list-style-type: none"> <li>Data permit status not specified</li> </ul>	<ul style="list-style-type: none"> <li>Timeline not specified</li> </ul>
<b>Sweden</b>	<ul style="list-style-type: none"> <li>Both institutions did not specify</li> </ul>	<ul style="list-style-type: none"> <li>Not provided</li> </ul>	<ul style="list-style-type: none"> <li>Requirements not specified</li> </ul>	<ul style="list-style-type: none"> <li>Not applicable</li> </ul>



Country	Ethics Committee Approval Required	Approval Timeline	Separate Data Permit Required	Permit Timeline
Iceland	• Not known (1 respondent)	• Timeline not specified	• Not known	• Not applicable

## 9.5 Appendix E: Data Source Architecture by Country (Question 14)

Country	Primary Data Sources	Secondary Sources	Data Integration Status
Norway	<ul style="list-style-type: none"> <li>• <b>Electronic Health Records (Secondary Care):</b> - Reported by 2/3 respondents</li> <li>• <b>National Registries:</b> - Reported by 1 respondent</li> <li>• <b>Disease/Quality Registries:</b> - Reported by 1 respondent</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Primary Care EHR:</b> - Reported by 1 respondent</li> </ul>	<ul style="list-style-type: none"> <li>• Responses varied on integration status</li> </ul>
Denmark	<ul style="list-style-type: none"> <li>• <b>National Registries</b></li> <li>• <b>Electronic Health Records (Secondary Care)</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>Clinical Support Systems:</b> - Laboratory - Microbiology - Pathology - Radiology</li> </ul>	<ul style="list-style-type: none"> <li>• Integration status not specified</li> </ul>
Finland	<ul style="list-style-type: none"> <li>• <b>Electronic Health Records:</b> - Both primary and secondary care</li> <li>• <b>Data Warehouse/SPE</b></li> </ul>	<ul style="list-style-type: none"> <li>• Registry connections mentioned</li> </ul>	<ul style="list-style-type: none"> <li>• Integration approach varied</li> </ul>
Sweden	<ul style="list-style-type: none"> <li>• <b>KI - National Registries:</b> - Disease/quality registries</li> <li>• <b>KUH - Response:</b> - "None. See Q21"</li> </ul>	<ul style="list-style-type: none"> <li>• Research databases mentioned</li> </ul>	<ul style="list-style-type: none"> <li>• Variable by institution</li> </ul>
Iceland	<ul style="list-style-type: none"> <li>• <b>Response provided:</b> - "Most data sources available but not yet mapped to OMOP"</li> </ul>	<ul style="list-style-type: none"> <li>• Not specified</li> </ul>	<ul style="list-style-type: none"> <li>• Pre-OMOP status</li> </ul>

### 9.5.1 Data Linkage Infrastructure Assessment (Questions 15-16)

Country	Linkage Capability	Identified Limitations (Q16)
Norway	<ul style="list-style-type: none"> <li>• Yes (2/3 respondents)</li> <li>• Not known (1/3)</li> </ul>	<ul style="list-style-type: none"> <li>• Limited specific limitations documented in responses</li> </ul>
Denmark	<ul style="list-style-type: none"> <li>• Yes (2/2 respondents)</li> </ul>	<ul style="list-style-type: none"> <li>• No specific limitations reported</li> </ul>
Finland	<ul style="list-style-type: none"> <li>• Yes (1/2)</li> <li>• Not known (1/2)</li> </ul>	<ul style="list-style-type: none"> <li>• Some restrictions noted but not detailed</li> </ul>
Sweden	<ul style="list-style-type: none"> <li>• KI: Yes</li> </ul>	<ul style="list-style-type: none"> <li>• Limited detail provided</li> </ul>



Country	Linkage Capability	Identified Limitations (Q16)
	• KUH: Not known	
Iceland	• Not known	• Not specified

### 9.5.2 Data Availability Matrix Analysis (Question 17)

Data Category	OMOP-Ready Status	Available but Requires Mapping	Not Available	Notes
<b>Diagnoses</b>	<ul style="list-style-type: none"> <li>Norway: Partially ready (1/3 respondents)</li> <li>Denmark: OMOP-ready</li> <li>Finland: Requires mapping</li> </ul>	<ul style="list-style-type: none"> <li>Norway: 2/3 respondents report available</li> <li>Finland: Available in source systems</li> <li>Iceland: Available but unmapped</li> </ul>	• None	Universal availability confirmed
<b>Procedures</b>	<ul style="list-style-type: none"> <li>Denmark: OMOP-ready</li> <li>Others: Not ready</li> </ul>	<ul style="list-style-type: none"> <li>Norway: All respondents confirm availability</li> <li>Finland: In EHR systems</li> <li>Iceland: Available</li> </ul>	• None	Mapping effort required
<b>Prescribed drugs</b>	<ul style="list-style-type: none"> <li>Norway: 2/3 OMOP-ready</li> <li>Denmark: Ready</li> <li>Finland: Ready</li> </ul>	<ul style="list-style-type: none"> <li>Norway: 1/3 requires mapping</li> </ul>	• None	High readiness level
<b>Administered drugs</b>	<ul style="list-style-type: none"> <li>Denmark, Finland: OMOP-ready</li> <li>Requires mapping: Norway, Iceland, Sweden-KUH</li> <li>Not available: Sweden-KI</li> </ul>	<ul style="list-style-type: none"> <li>Most sites have in clinical systems</li> </ul>	• Some sites lack integration	Variable availability
<b>Dispensed drugs</b>	<ul style="list-style-type: none"> <li>Denmark: Registry ready</li> <li>Others: Limited</li> </ul>	<ul style="list-style-type: none"> <li>Available in prescription databases</li> </ul>	• Hospital dispensing gaps noted	Registry dependent
<b>Laboratory Results</b>	<ul style="list-style-type: none"> <li>No sites report OMOP-ready status</li> </ul>	<ul style="list-style-type: none"> <li>Universal availability in source systems</li> <li>Norway, Denmark, Finland all confirm presence</li> </ul>	• None	Universal mapping need



Data Category	OMOP-Ready Status	Available but Requires Mapping	Not Available	Notes
<b>Imaging Results</b>	• No OMOP readiness	• Denmark: Available • Norway: 1/3 has access	• Norway: 2/3 report not available • Finland: Not accessible	Limited availability
<b>Pathology Results</b>	• Limited OMOP mapping	• Available in most sites	• Some sites lack digital pathology	Text-heavy format
<b>Biomarker Data</b>	• No OMOP availability	• Denmark: Available in clinical systems	• Norway: All 3 respondents report absence • Finland: Not available	Critical gap identified
<b>Patient Reported Outcomes</b>	• No OMOP availability	• No systematic collection reported	• Universal absence across all sites	Not part of routine care
<b>Genetic Data</b>	• No OMOP availability	• Sweden KI: Research settings only	• Clinical availability minimal	Limited to research

### 9.5.3 Clinical Variable Capture Success (Question 21)

Variable Category	Norway	Denmark	Finland	Overall Capture Rate	Quality Notes
<b>Cancer Diagnoses</b>	Captured (all respondents)	Captured	Captured	100% successful	High quality, ICD-10 based
<b>Histology</b>	Partial (2/3 respondents)	Captured	Limited capture	60-70% complete	Variability in documentation
<b>Biomarkers</b>	Not captured systematically	Limited	Not available	<30% complete	Major limitation for precision oncology
<b>Comorbidities</b>	Captured via diagnoses	Captured	Captured	90% successful	Charlson index calculable
<b>Demographics</b>	Complete	Complete	Complete	100% successful	Age, sex universally available
<b>Smoking Status</b>	Not systematically captured	Limited	Limited	<40% complete	Critical gap for lung cancer
<b>Performance Status</b>	Not systematically captured	Not captured	Limited	<20% complete	ECOG/PS largely missing
<b>Treatment Details</b>					
- Chemotherapy	Captured	Captured	Captured	95% successful	Drug codes available



Variable Category	Norway	Denmark	Finland	Overall Capture Rate	Quality Notes
- Immunotherapy	Captured	Captured	Captured	95% successful	ICI identification successful
- Surgery	Procedure codes	Captured	Captured	90% successful	Procedure coding complete
- Radiotherapy	Variable capture	Captured	Partial	60% complete	Site-specific gaps
Death	Registry linkage	Complete	Available	95% successful	High quality mortality data
Disease Progression	Not structured	Limited	Not captured	<20% complete	Requires clinical notes review

## 9.6 Appendix F1: Implementation Challenge Gradients (Question 23)

Process Stage	Norway (n=3)	Denmark (n=2)	Finland (n=2)	Iceland (n=1)	Interpretation
<b>1. Design OMOP ETL Specification</b>	<ul style="list-style-type: none"> <li>• "Moderately challenging" (2)</li> <li>• "Significant effort but manageable"</li> <li>• Required dedicated resources</li> </ul>	<ul style="list-style-type: none"> <li>• Limited responses</li> <li>• Clinical researchers not involved in technical design</li> </ul>	<ul style="list-style-type: none"> <li>• Responses not provided</li> <li>• Technical team may not have participated</li> </ul>	<ul style="list-style-type: none"> <li>• "Not in my remit"</li> <li>• Observer status</li> <li>• Learning from others' experience</li> </ul>	<p><b>Highest challenge point</b></p> <p>Design phase requires most resources</p>
<b>2. Implement OMOP ETL Specification</b>	<ul style="list-style-type: none"> <li>• "Slightly challenging" (2)</li> <li>• "Minor adjustments needed"</li> <li>• Benefited from design phase</li> </ul>	<ul style="list-style-type: none"> <li>• Technical implementation outside respondent scope</li> </ul>	<ul style="list-style-type: none"> <li>• Limited visibility into technical process</li> </ul>	<ul style="list-style-type: none"> <li>• Not applicable</li> </ul>	<p><b>Moderate challenge</b></p> <p>Easier once design complete</p>
<b>3. QA/QC of OMOP Instance</b>	<ul style="list-style-type: none"> <li>• "Not challenging" (2)</li> <li>• "Routine process"</li> <li>• DataQualityDashboard effective</li> </ul>	<ul style="list-style-type: none"> <li>• Process established but not documented in responses</li> </ul>	<ul style="list-style-type: none"> <li>• QC procedures in place</li> </ul>	<ul style="list-style-type: none"> <li>• Not applicable</li> </ul>	<p><b>Lowest challenge</b></p> <p>OHDSI tools simplify QC</p>
<b>4. Study Variable Survey</b>	<ul style="list-style-type: none"> <li>• Mixed responses</li> <li>• Some variables unavailable</li> <li>• Required iterations</li> </ul>	<ul style="list-style-type: none"> <li>• Clinical input provided</li> <li>• Feasibility challenges noted</li> </ul>	<ul style="list-style-type: none"> <li>• Variable availability issues</li> </ul>	<ul style="list-style-type: none"> <li>• Observing process</li> </ul>	<p><b>Moderate complexity</b></p> <p>Data availability constrains</p>
<b>5. Diagnostic Package Execution</b>	<ul style="list-style-type: none"> <li>• Successfully completed</li> <li>• Some debugging required</li> </ul>	<ul style="list-style-type: none"> <li>• Limited technical responses</li> </ul>	<ul style="list-style-type: none"> <li>• Execution challenges noted</li> </ul>	<ul style="list-style-type: none"> <li>• Not applicable</li> </ul>	<p><b>Technical hurdles</b></p> <p>Environment-specific issues</p>



Process Stage	Norway (n=3)	Denmark (n=2)	Finland (n=2)	Iceland (n=1)	Interpretation
<b>6. Initial Package Execution</b>	<ul style="list-style-type: none"> <li>Completed with delays</li> <li>R environment issues</li> </ul>	<ul style="list-style-type: none"> <li>Not documented</li> </ul>	<ul style="list-style-type: none"> <li>Technical delays reported</li> </ul>	<ul style="list-style-type: none"> <li>Not applicable</li> </ul>	<p><b>Infrastructure dependent</b></p> <p>Local setup affects success</p>
<b>7. Preliminary Results Review</b>	<ul style="list-style-type: none"> <li>Iterative process</li> <li>Multiple rounds needed</li> </ul>	<ul style="list-style-type: none"> <li>Clinical review provided</li> </ul>	<ul style="list-style-type: none"> <li>Data quality issues identified</li> </ul>	<ul style="list-style-type: none"> <li>Learning opportunity</li> </ul>	<p><b>Time intensive</b></p> <p>Quality review crucial</p>
<b>8. Final Package Execution</b>	<ul style="list-style-type: none"> <li>Successful completion</li> <li>Lessons learned applied</li> </ul>	<ul style="list-style-type: none"> <li>Completed successfully</li> </ul>	<ul style="list-style-type: none"> <li>Final execution achieved</li> </ul>	<ul style="list-style-type: none"> <li>Not applicable</li> </ul>	<p><b>Smoother process</b></p> <p>Earlier issues resolved</p>

## 9.7 Appendix F2: VALO Study Process Steps (Questions 25-32)

### 9.7.1 F2.1 Design of OMOP ETL Specification (Question 25)

Country	Response
Norway	"We were able to reuse work from previous studies"
Denmark	Limited technical responses from clinical researchers. Design phase managed by technical team not represented in survey
Finland	Responses focused on data availability challenges rather than design process
Iceland	"OMOP is not yet implemented in Iceland"
Sweden	Observer perspective on design requirements. No specific design feedback provided

### 9.7.2 F2.2 Implementation of OMOP ETL Specification (Question 26)

Country	Response
Norway	"As for 29"
Denmark	Implementation feedback not provided by respondents
Finland	Limited implementation details in responses
Iceland	"OMOP is not yet implemented in Iceland"
Sweden	No implementation feedback as observer sites



### 9.7.3 F2.3 QA/QC of OMOP Instance (Question 27)

Country	Response
Norway	"As for 29"
Denmark	QC process completed but detailed feedback not provided
Finland	Quality process feedback not elaborated
Iceland	"OMOP is not yet implemented in Iceland"
Sweden	Observer status, no QC implementation

### 9.7.4 F2.4 Study Variable Survey and Package Execution (Questions 28-32)

Question	Summary of Responses
Study Variable Survey (Q28)	Limited responses were provided for this question across all sites, with most respondents either not answering or indicating it was outside their remit
Diagnostic Package Execution (Q29)	Limited or no responses from all sites
Initial Package Execution (Q30)	Limited or no responses from all sites
Preliminary Results Review (Q31)	Limited or no responses from all sites
Final Package Execution (Q32)	Limited or no responses from all sites

*Note: The consistent reference to "As for 29" in Norwegian responses indicates a unified approach across ETL design, implementation, and QA/QC phases, with the ability to reuse previous work being a critical success factor.*

## 9.8 Appendix G: Future Investments and Improvements (Questions 33-36)

### 9.8.1 G.1 Data Infrastructure Requirements (Question 33)

Country	Response
Norway (R1)	"Having a clinical data warehouse as basis for the OMOP implementation has made the process part of regular operations"
Norway (R2)	"Database: We had issues with handling postgres bigints in R, which has a lower limit than what bigint can store"
Denmark	Infrastructure feedback not provided by clinical researcher respondents
Finland	Infrastructure assessment not detailed in responses



<b>Sweden</b>	Observer sites did not provide infrastructure requirements
<b>Iceland</b>	Infrastructure planning informed by consortium experiences but specific requirements not detailed

### 9.8.2 G.2 Data Conversion Needs (Question 34)

Limited responses were received for this question across all sites, with most respondents not providing specific data conversion requirements or challenges.

### 9.8.3 G.3 Current OMOP Capabilities Assessment (Question 35)

Country	Response
<b>Norway (Skills)</b>	"OMOP skills have been established in the data warehouse team over the last 3 years"
<b>Norway (Standardization)</b>	"Vocabulary versioning is challenging. As a consortium we could define target vocabulary version"
<b>Denmark</b>	Capability assessment not provided by respondents
<b>Finland</b>	OMOP maturity status not detailed
<b>Sweden</b>	Observer perspective on capabilities, specific assessment not provided
<b>Iceland</b>	Pre-implementation phase, learning from consortium experiences

### 9.8.4 G.4 Recommendations for Future Improvements (Question 36)

Country	Response
<b>Denmark (R1)</b>	"Developing more advanced methods for conducting federated analysis and obtaining aggregated results"
<b>Denmark (R2)</b>	"Improving analytical capabilities into inferential and causal analytics"
<b>Norway</b>	Future recommendations not specifically provided in Question 36 responses
<b>Finland</b>	Recommendations not detailed in responses
<b>Sweden</b>	Observer sites provided limited forward-looking recommendations
<b>Iceland</b>	Future planning based on lessons learned but specific recommendations not documented



## 9.9 Appendix H: IQVIA Execution Experience Details

### 9.9.1 H.1 Communication Strategy Evolution

Phase	Description
<b>Initial Phase (Email-based)</b>	Primary communication via email threads. Opportunities identified: Slow response times, difficulty tracking complex technical discussions, information siloing. Transition to enhanced communication methods identified as beneficial.
<b>Evolved Phase (Teams-based)</b>	Transition to dedicated Microsoft Teams channels per site (initiated by Oslo University Hospital). Advantages: Real-time debugging support, persistent documentation of issues and solutions, improved response times. Day 1 requirement: IQVIA operational team recommends provisioning.

### 9.9.2 H.2 Stakeholder Management Requirements

Stakeholder Group	Requirements
Technical teams	Needed clarity on exactly what was required and by when
Clinical experts	Needed context on why certain data elements were essential
Institutional leadership	Needed realistic timelines and resource requirements
Governance / compliance teams	Needed early visibility into regulatory requirements
Funder/Sponsor	Required projections, timeline adjustments due to site capacity constraints and inherent complexity revealed through execution

### 9.9.3 H.3 Resource Coordination Challenges

Challenge Category	Description
Multi-Role Dependencies	Coordination required across data scientists, ETL experts, clinical experts, and governance specialists
Competing Priorities	Sites with dedicated research infrastructure and protected time demonstrated more consistent engagement
Expert Availability	Specialized expertise was at risk of becoming bottlenecks



#### 9.9.4 H.4 Lessons for Future Network Project Management

Recommendation
Allocate 30-40% of total timeline to coordination overhead
Establish communication infrastructure (Teams/Slack channels) at project initiation
Require sites to designate a primary coordinator
Conduct early role-mapping exercises
Build in structured checkpoints (weekly in active phases, bi-weekly in stable phases) rather than relying on ad hoc communication
Document all technical solutions in shared repositories
Maintain transparent risk registers

*Note: A recommendation emerging from this experience: multiply initial technical execution estimates by 1.5-2.0x to account for multi-site coordination complexity.*