

Advancing Transparency-by-design in Clinical Trial Disclosure: Real-World Applications for Regulatory submissions

Obaraboye Olude, Privacy Analytics (an IQVIA company), Ottawa, Canada

Chantal Din-Bell, Sanofi, Paris France

23rd March 2026 at PHUSE US CONNECT in Austin, USA.



PRIVACY
ANALYTICS

an IQVIA company

sanofi

Disclaimer - IQVIA

- The analyses, their interpretation, and related information contained herein are made and provided subject to the assumptions, methodologies, caveats, and variables described in this report and are based on third party sources and data reasonably believed to be reliable. No warranty is made as to the completeness or accuracy of such third party sources or data.
- As with any attempt to estimate future events, the forecasts, projections, conclusions, and other information included herein are subject to certain risks and uncertainties, and are not to be considered guarantees of any particular outcome.
- All reproduction rights, quotations, broadcasting, publications reserved. The contents of this presentation are confidential, and no part of this presentation may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without express written consent of IQVIA.
- Copyright © 2026 IQVIA. All rights reserved. IQVIA® is a registered trademark of IQVIA Inc. in the United States and various other countries.

Introducing this sessions first speaker

Obaraboye Olude MBBS, MPH
Sr Advisor and Technical Manager of Privacy Analytics'
Clinical Trial Transparency Team

Providing anonymization and redaction services to clinical
trial sponsors for **7+ years**

Supported **over 210** regulatory document submissions

Background in Medicine and Public Health

Based in Ottawa Canada



Session Agenda



Identify

Key checkpoints for integrating transparency across a products lifecycle



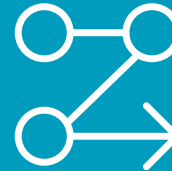
Review

Practical strategies for incorporating transparency into the development of redaction/anonymization-friendly documents



Highlight

Approaches to enhance cross-functional stakeholder collaboration



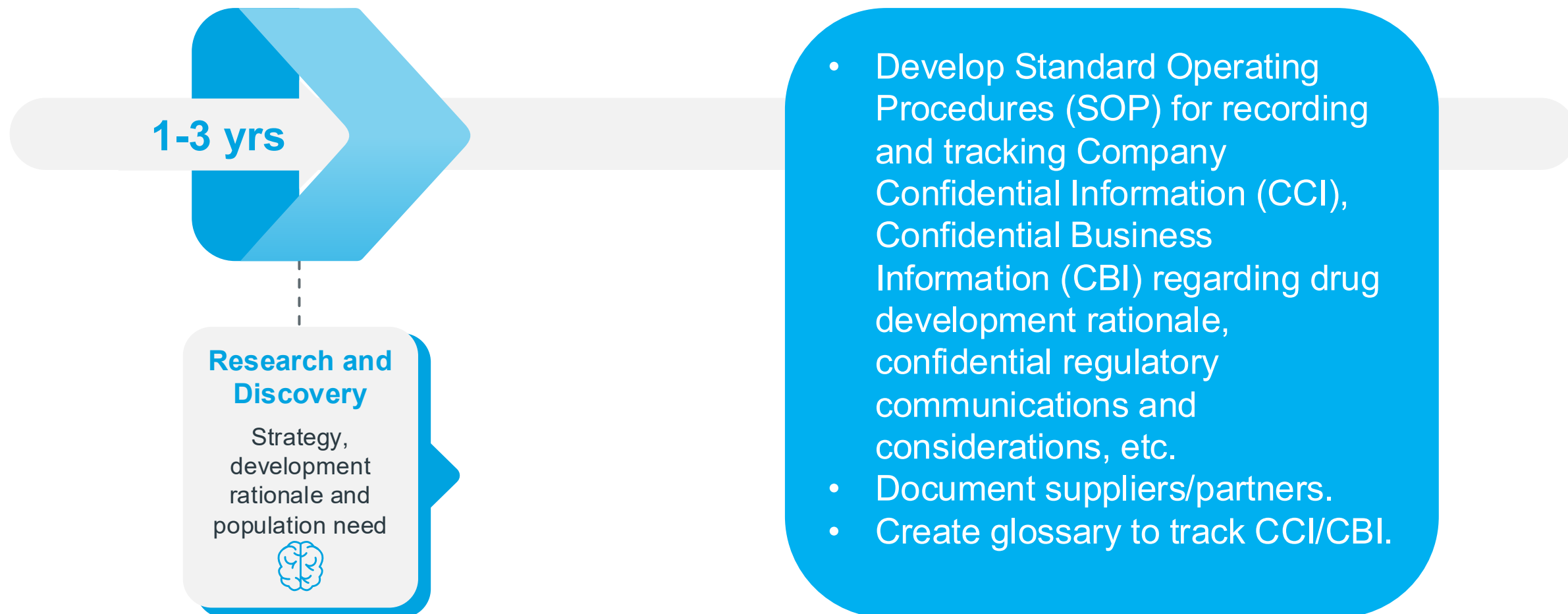
Outline

Strategies for developing and monitoring organizational transparency guidelines

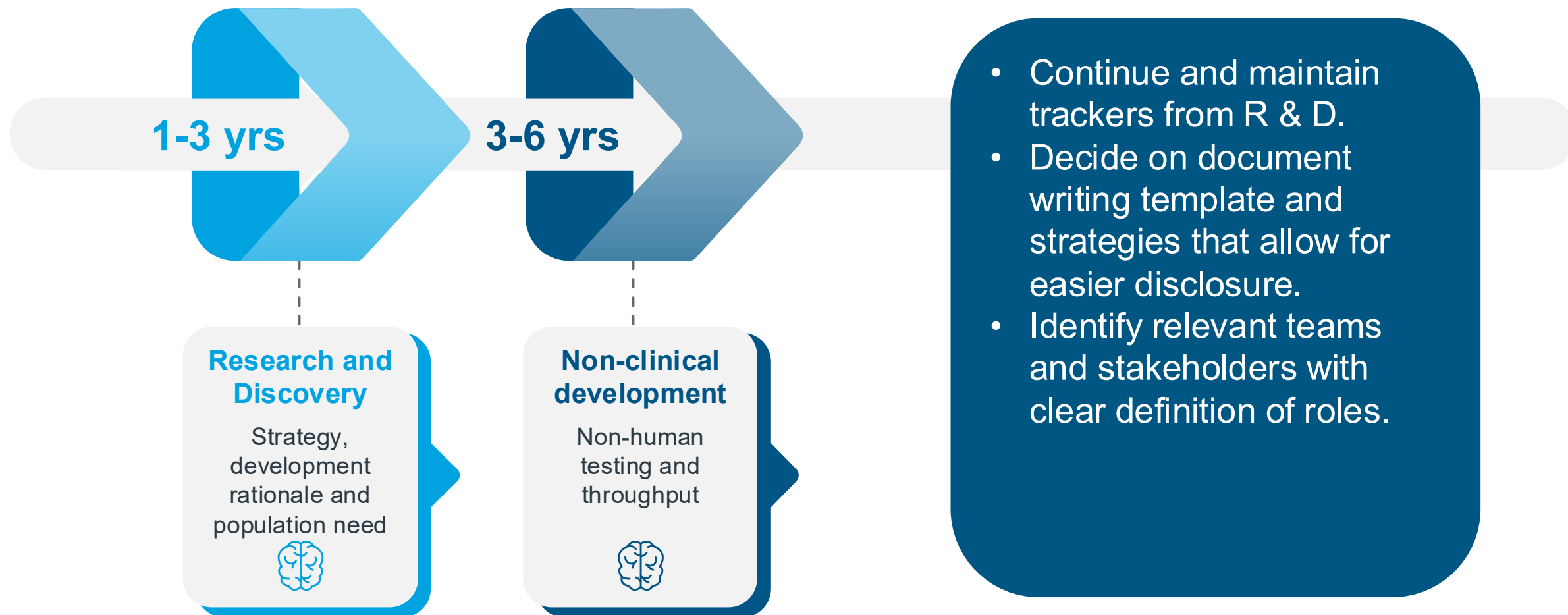
1. Identify: Key points for integrating transparency across the product lifecycle

Research and Discovery (R & D)

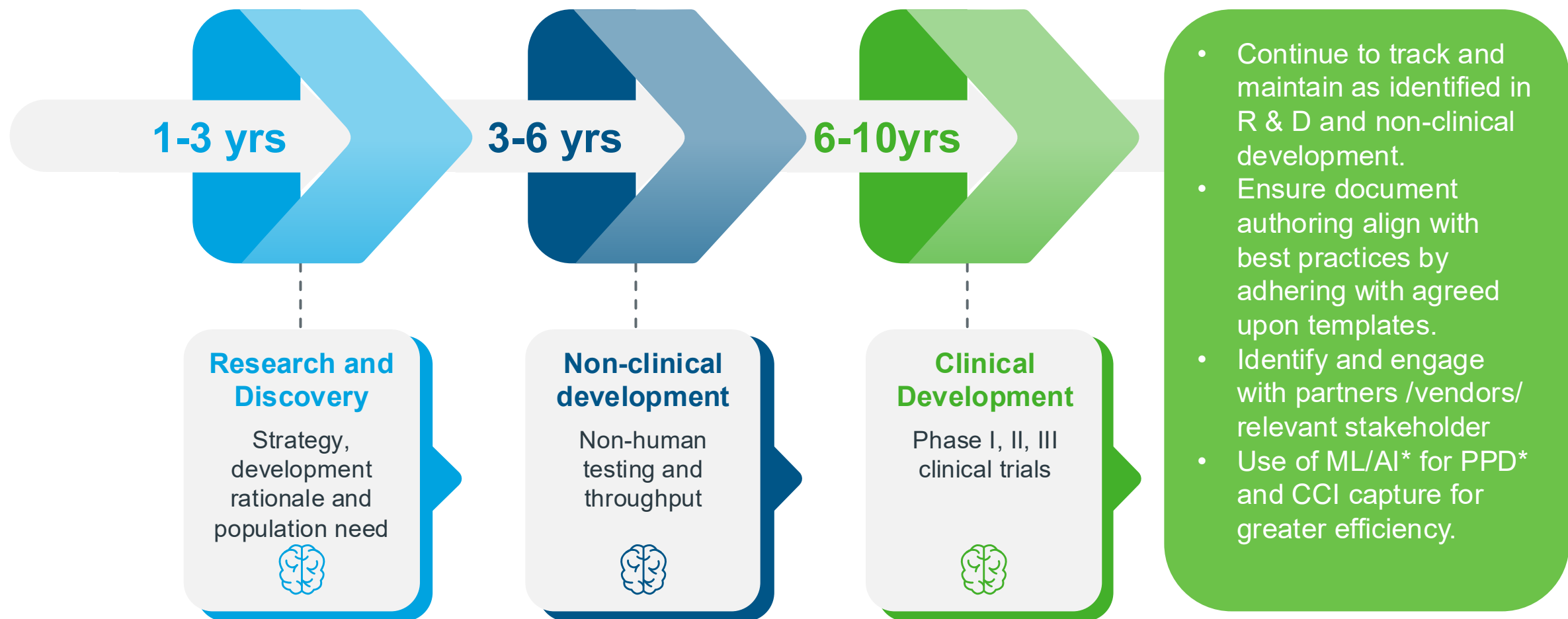
Transparency checkpoints and steps during Research and Discovery phase of drug development



From Research and Discovery to Non-clinical Development

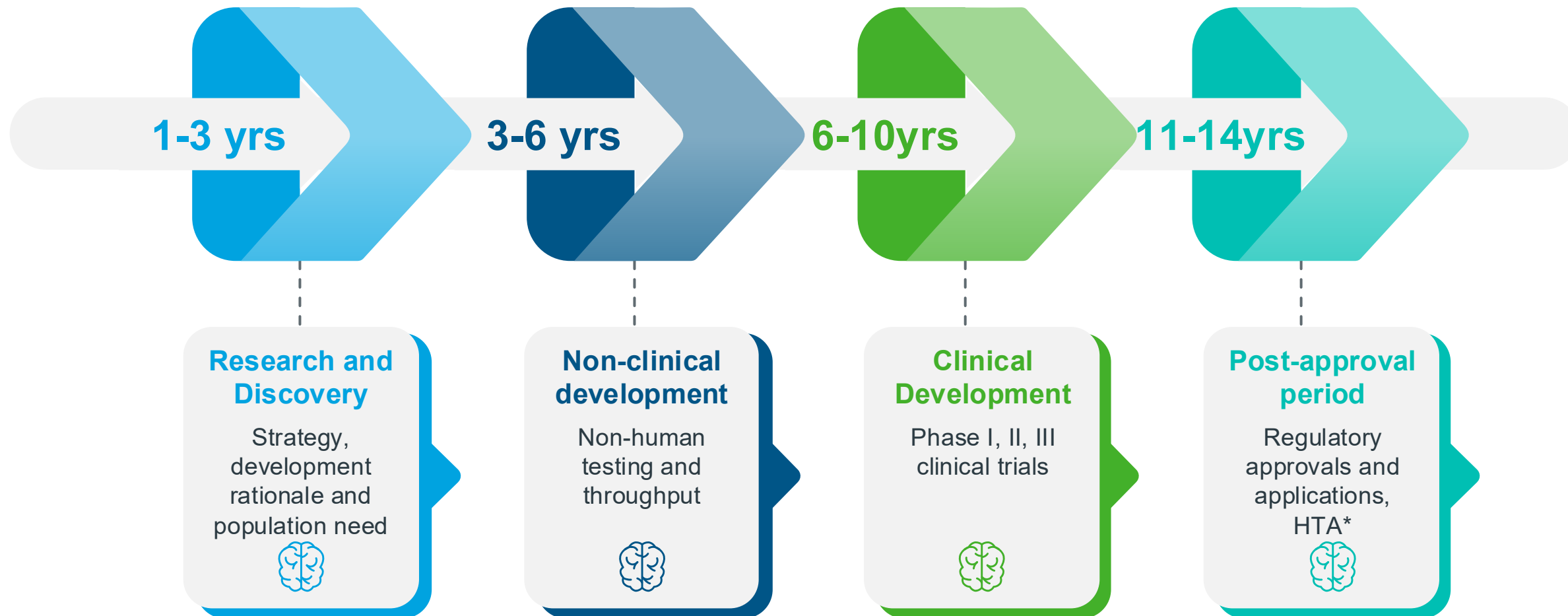


From Research and Discovery to Clinical Development



Overview of Key Drug Development Milestones

Key points for transparency considerations in document writing occur during the clinical development and post-approval stages.



2. Review: Practical strategies for the development of redaction- and anonymization-friendly documents, facilitating transparency and disclosure

Considerations for clinical document disclosure



Privacy Protection

Protection of participant data requires **high-accuracy** data capture.

Considerations for clinical document disclosure



Privacy Protection

Protection of participant data requires **high-accuracy** data capture.



Utility Preservation

Under EMA CDP & HC PRCI, utility-preserving **statistical anonymization** is preferred.

Privacy and utility preservation requires high-accuracy data capture and classification



PI is detected in the documents **using a fine-tuned LLM human-in-the-loop optimization.**

Subject 899-456 was a 34-year-old female with a medical history of coeliac disease.

Privacy and utility preservation requires high-accuracy data capture and classification



PI is detected in the documents **using a fine-tuned LLM human-in-the-loop optimization.**

Subject 899-456 was a 34-year-old female with a medical history of coeliac disease.

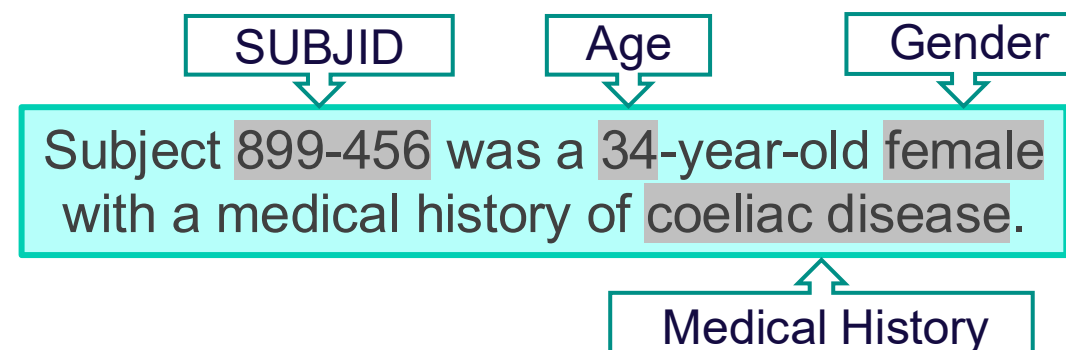
Privacy and utility preservation requires high-accuracy data capture and classification



PI is detected in the documents **using a fine-tuned LLM human-in-the-loop optimization.**



PI is **classified** to determine contribution to re-identification risk.



Privacy and utility preservation requires high-accuracy data capture and classification



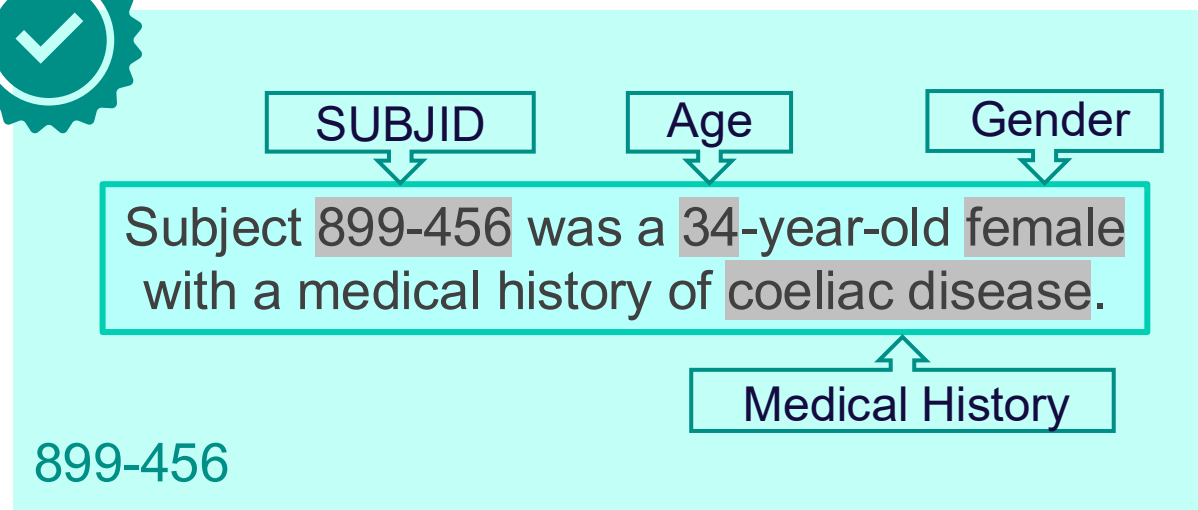
PI is detected in the documents **using a fine-tuned LLM human-in-the-loop optimization.**



PI is **classified** to determine contribution to re-identification risk.



Detected PI is then associated with the **correct participant ID.**



Privacy and utility preservation requires high-accuracy data capture and classification



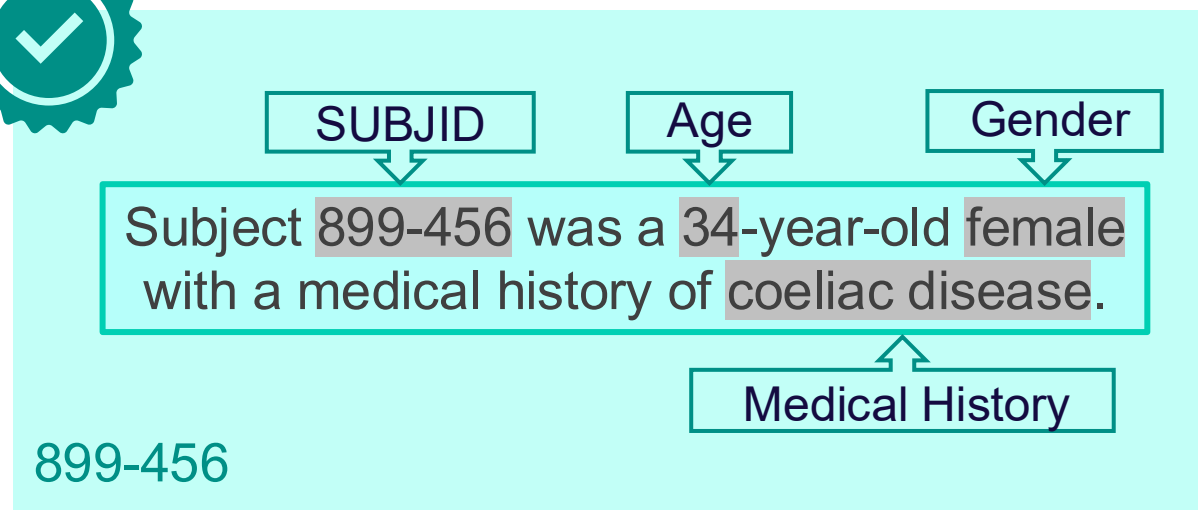
PI is detected in the documents **using a fine-tuned LLM human-in-the-loop optimization.**



PI is **classified** to determine contribution to re-identification risk.



Detected PI is then associated with the **correct participant ID.**

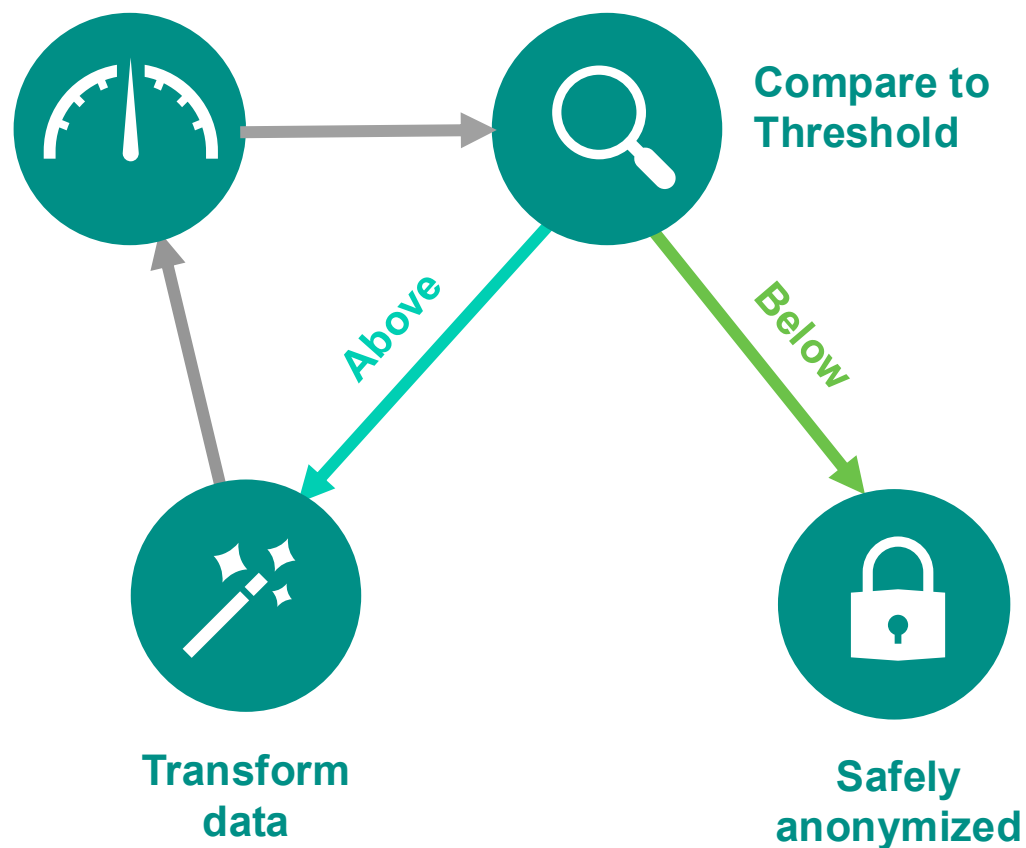


A database containing all individuals represented in the document is created, upon which the re-identification risk assessment can be performed.

Re-identification risk assessment and mitigation

Measure re-identification risk:

- ✓ Uniqueness
- ✓ Reference population
- ✓ Recall



Statistical Anonymization

Subject 899-456 was a 34-year-old female with a medical history of coeliac disease.



Subject 454-994 was a 30-35-year-old female with a medical history of autoimmune disorders.

Considerations for clinical document disclosure



Privacy Protection

Protection of participant data requires **high-accuracy** data capture.



Utility Preservation

Under EMA CDP & HC PRCI, utility-preserving **statistical anonymization** is preferred.



CCI/CBI Protection

Protection of commercial data is complex, requiring **multi-stakeholder consultation and alignment.**

Considerations for clinical document disclosure



Privacy Protection

Protection of participant data requires **high-accuracy** data capture.



Utility Preservation

Under EMA CDP & HC PRCI, utility-preserving **statistical anonymization** is preferred.



CCI/CBI Protection

Protection of commercial data is complex, requiring **multi-stakeholder consultation and alignment**.



Timelines

Adhering to timelines whilst managing all these priorities is challenging.

The cornerstones of transparency-by-design

Writing clinical documents with disclosure in mind

Minimization

Only include necessary personal and commercial data in documents.

Only include necessary documents in submission package.



The cornerstones of transparency-by-design

Writing clinical documents with disclosure in mind

Minimization

Only include necessary personal and commercial data in documents.

Only include necessary documents in submission package.



Normalization

Present variables in a limited number of predictable formats for ease of detection, classification and protection.

The cornerstones of transparency-by-design

Writing clinical documents with disclosure in mind

Minimization

Only include necessary personal and commercial data in documents.

Only include necessary documents in submission package.



Normalization

Present variables in a limited number of predictable formats for ease of detection, classification and protection.

Standardization

Use the ICHE3 guidance for CSR structuring and include personal data only in pre-specified sections.

Minimalization



**PRIVACY
ANALYTICS**

an IQVIA company

Data minimization in clinical documents

Avoid the inclusion of data (personal and commercial) which are **not strictly required** for the interpretation of the clinical study results or a regulatory decision.

Avoid the inclusion of personal details which **not medically relevant**, such as

Data minimization in clinical documents



Family or living
situation

The participants' step-daughter
called an ambulance from
outside his apartment complex.



The participant was taken by
ambulance to hospital.

Data minimization in clinical documents



Family or living situation

The participants' step-daughter called an ambulance from outside his apartment complex.



The participant was taken by ambulance to hospital.



Career or job details

The participant was diagnosed with depression after losing their teaching job.



The participant was diagnosed with depression due to personal circumstances.

Data minimization in clinical documents



Family or living situation

The participants' step-daughter called an ambulance from outside his apartment complex.



The participant was taken by ambulance to hospital.



Career or job details

The participant was diagnosed with depression after losing their teaching job.



The participant was diagnosed with depression due to personal circumstances.



Sensitive Information

- ⚠ Visible deformities
- ⚠ Culturally-specific attributes
- ⚠ Sexual health or fertility
- ⚠ Mental health
- ⚠ Religious/political affiliation
- ⚠ Criminal records

Additional recommendations: Data minimization



Avoid repetition of information

- *Especially* if presented with slight differences, in multiple locations
 - Use cross-references where possible
-

Additional recommendations: Data minimization



Avoid repetition of information

- *Especially* if presented with slight differences, in multiple locations
- Use cross-references where possible



Avoid inclusion of non-machine-readable personal data

- E.g., embedded in figures/images
- Precludes the use of automation and necessitates manual redaction

Additional recommendations: Data minimization



Avoid repetition of information

- *Especially* if presented with slight differences, in multiple locations
- Use cross-references where possible



Avoid inclusion of non-machine-readable personal data

- E.g., embedded in figures/images
- Precludes the use of automation and necessitates manual redaction



Do not include identifiers in filenames, bookmarks, headers or footers

- Narratives are often bookmarked by subject ID – avoid!
- Avoid study personnel names or Subject IDs in footer/header

Normalization



**PRIVACY
ANALYTICS**

an IQVIA company

Heterogeneity of data representation

Patient ID	ABC-001-652
Sex	F
Age	51
Race	White
Weight	68.0 kg
Height	163 cm
BMI	25.7
Treatment Start Date	18FEB19 (Day 0)
AE Start Date	20FEB19 (Day 2)
Reason for Narrative	Cellulitis

Patient 001-652

Patient 652, a 51-year-old Caucasian female, with a medical history of hypertension and arthritis complained of influenza-like symptoms on 20-FEB-2019, Day 2 of treatment. These symptoms included a high fever, chills, scratchy throat and rhinitis. On Day 3 (21-02-2019), the patient reported that her symptoms had worsened significantly, and she was admitted to hospital. On examination, it was noted that the patient had a severe rash on her right foot. Cellulitis was diagnosed and the patient received IV fluids and antibiotics. By Day 6 (24-FEB), the patient had lost a significant amount of weight and currently stood at 140lbs. By Day 10 the patient was discharged.

Heterogeneity of data representation

Patient ID	ABC-001-652
Sex	F
Age	51
Race	White
Weight	68.0 kg
Height	163 cm
BMI	25.7
Treatment Start Date	18FEB19 (Day 0)
AE Start Date	20FEB19 (Day 2)
Reason for Narrative	Cellulitis

Patient 001-652

Patient 652, a 51-year-old Caucasian female, with a medical history of hypertension and arthritis complained of influenza-like symptoms on 20-FEB-2019, Day 2 of treatment. These symptoms included a high fever, chills, scratchy throat and rhinitis. On Day 3 (21-02-2019), the patient reported that her symptoms had worsened significantly, and she was admitted to hospital. On examination, it was noted that the patient had a severe rash on her right foot. Cellulitis was diagnosed and the patient received IV fluids and antibiotics. By Day 6 (24-FEB), the patient had lost a significant amount of weight and currently stood at 140lbs. By Day 10 the patient was discharged.



Subject ID – 3 formats



Sex – 2 formats



Age – 1 format



Race – 2 formats



Weight – 2 formats



Height -1 format



BMI - 1 format

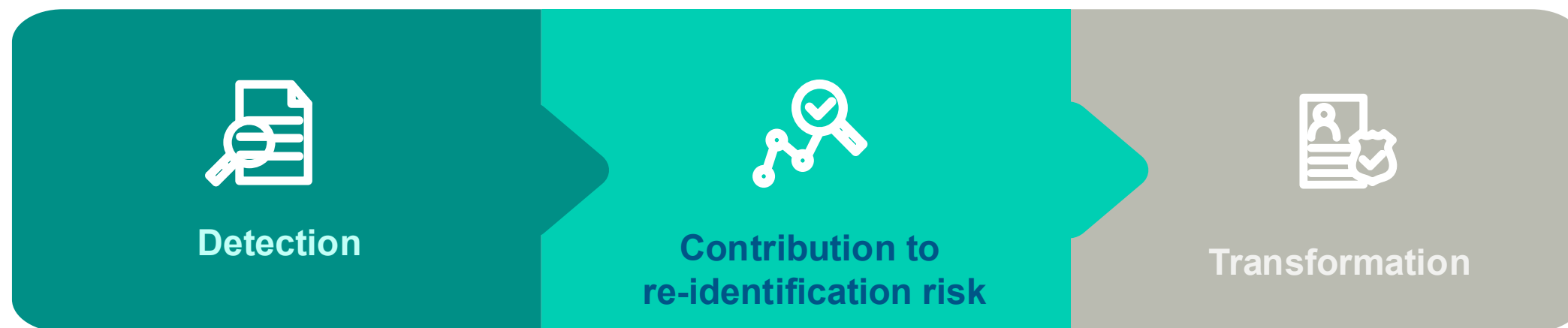


Date – 5 formats



Medical Details – AE and MH

Importance of data normalization



Even automated approaches are more **effective** when data are represented in a **consistent and predictable** fashion.

To build a data model correctly, information must be **classified properly**, and **de-duplicated**.

Risk-based transformation of the information requires **accurate detection and risk assessment**.

Recommendations for data normalization by category – Subject IDs



Subject IDs

- Use a **single ID format** throughout the document
- Standardize the format with the **USUBJID** format in the datasets
- Avoid the use of short-form IDs (e.g., subject 001), as there may be **subjects from multiple sites or multiple studies with the same ID**.

Recommendations for data normalization by category – Subject IDs



Subject IDs

- Use a **single ID format** throughout the document
- Standardize the format with the **USUBJID** format in the datasets
- Avoid the use of short-form IDs (e.g., subject 001), as there may be **subjects from multiple sites or multiple studies with the same ID**.

Take for example, Study A101.

Subject **001-001** is subject 001 from site 001.

Subject **002-001** is subject 001 from site 002.

Without the site ID, we might confuse them.

In a clinical summary document, **without the study number**, we may confuse them further with individuals from a different study with the same ID assignment convention.

Best practice:

- ✓ Study-Site-Subject
- ✓ Maintain the same number of leading zeroes

Recommendations for data normalization by category – Subject IDs, Dates



Subject IDs

- Use a **single ID format** throughout the document
- Standardize the format with the **USUBJID** format in the datasets
- Avoid the use of short-form IDs (e.g., subject 001), as there may be **subjects from multiple sites or multiple studies with the same ID**.



Dates

- For ease of date-offsetting, a **single format** should be used.
- **Avoid partial dates** (e.g., DD-MM)
- **Avoid combining multiple dates in a single entry** (e.g., 1st to 15th April 2015, instead use 01-Apr-2015 to 15-Apr-2015)
- Alternatively, **relative dates** (e.g., Day 5) alone could be used.

Recommendations for data normalization by category – Subject IDs, Dates, Demographic information



Subject IDs

- Use a **single ID format** throughout the document
- Standardize the format with the **USUBJID** format in the datasets
- Avoid the use of short-form IDs (e.g., subject 001), as there may be **subjects from multiple sites or multiple studies with the same ID**.



Dates

- For ease of date-offsetting, a **single format** should be used.
- **Avoid partial dates** (e.g., DD-MM)
- **Avoid combining multiple dates in a single entry** (e.g., 1st to 15th April 2015, instead use 01-Apr-2015 to 15-Apr-2015)
- Alternatively, **relative dates** (e.g., Day 5) alone could be used.



Demographics

- **Avoid verbal presentation** of ages
- Use a **single race/ethnicity format**
- For height/weight, use either **metric or imperial** (metric is preferred).
- Avoid grouping **multiple subject IDs** by demography
- **Avoid “yes/no” or checkbox presentation** of characteristics

Recommendations for data normalization by category – Subject IDs, Dates, Demographic information



Subject IDs

- Use a **single ID format** throughout the document
- Standardize the format with the **USUBJID** format in the datasets
- Avoid the use of short-form IDs (e.g., subject 001), as there may be **subjects from multiple sites or multiple studies with the same ID**.



Dates

- For ease of date-offsetting, a **single format** should be used.
- **Avoid partial dates** (e.g., DD-MM)
- **Avoid combining multiple dates in a single entry** (e.g., 1st to 15th April 2015, instead use 01-Apr-2015 to 15-Apr-2015)
- Alternatively, **relative dates** (e.g., Day 5) alone could be used.



Demographics

- **Avoid verbal presentation** of ages
- Use a **single race/ethnicity format**
- For height/weight, use either **metric or imperial** (metric is preferred).
- Avoid grouping **multiple subject IDs** by demography
- **Avoid “yes/no” or checkbox presentation** of characteristics

Race:	
White	
Asian	X
Black or African American	
American Indian or Alaska Native	
Native Hawaiian or Other Pacific Islander	
Multiple	

Recommendations for data normalization by category – Subject IDs, Dates, Demographic information, medical details



Subject IDs

- Use a **single ID format** throughout the document
- Standardize the format with the **USUBJID** format in the datasets
- Avoid the use of short-form IDs (e.g., subject 001), as there may be **subjects from multiple sites or multiple studies with the same ID**.



Dates

- For ease of date-offsetting, a **single format** should be used.
- **Avoid partial dates** (e.g., DD-MM)
- **Avoid combining multiple dates in a single entry** (e.g., 1st to 15th April 2015, instead use 01-Apr-2015 to 15-Apr-2015)
- Alternatively, **relative dates** (e.g., Day 5) alone could be used.



Demographics

- **Avoid verbal presentation** of ages
- Use a **single race/ethnicity format**
- For height/weight, use either **metric or imperial** (metric is preferred).
- Avoid grouping **multiple subject IDs** by demography
- **Avoid “yes/no” or checkbox presentation** of characteristics



Medical details

- Consider presenting medical histories in **tables** to minimize replacement text offset in paragraphs.
- Where possible, ensure information is coded to **MedDRA, and limit inclusion of verbatim terms**.
- Clearly differentiate between **adverse events and medical histories**.

Standardization



PRIVACY
ANALYTICS

an IQVIA company

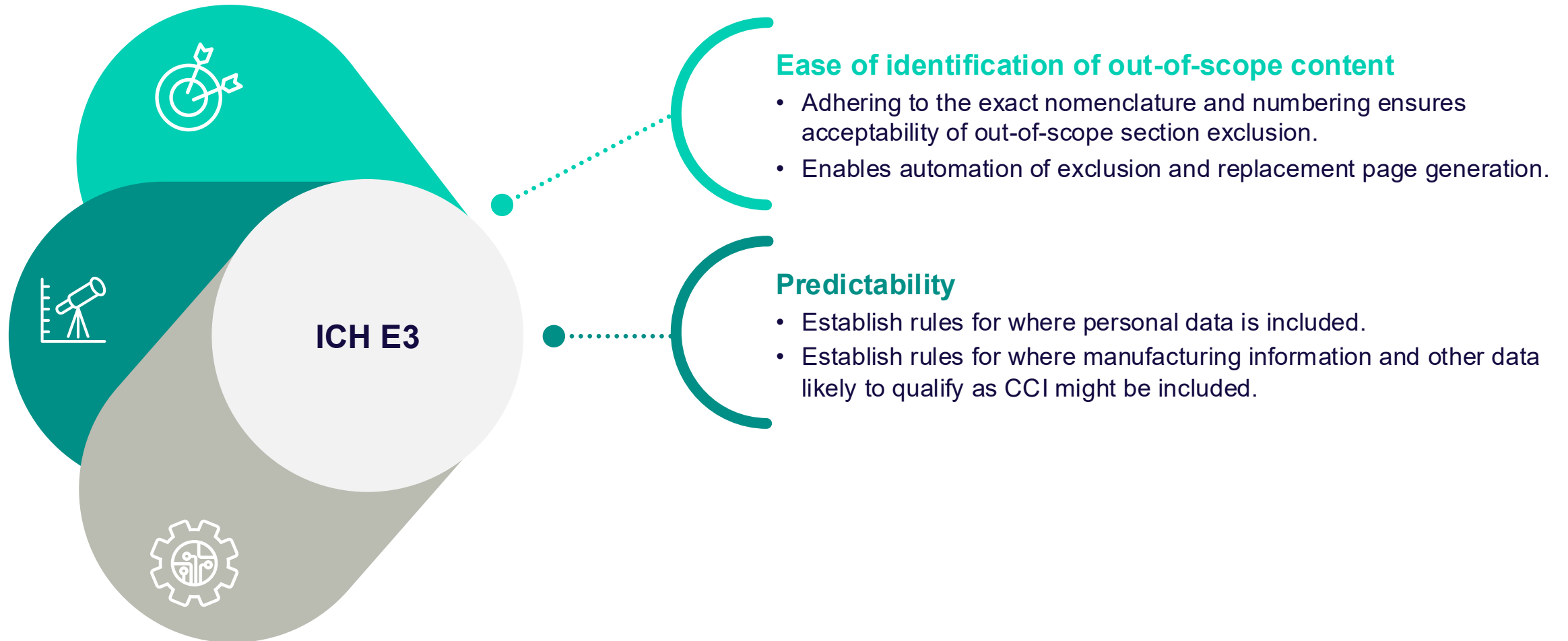
Using ICH E3 guidelines for CSR (and appendix) structuring



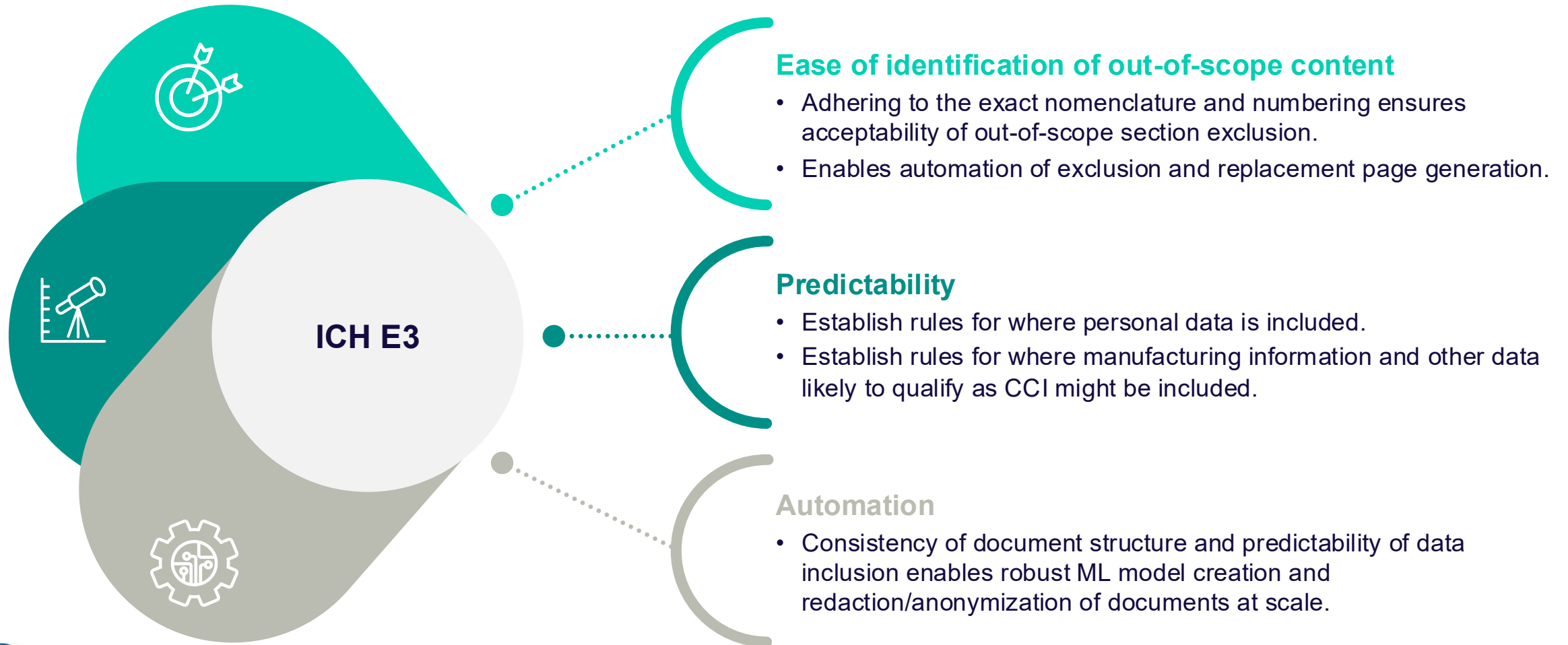
Ease of identification of out-of-scope content

- Adhering to the exact nomenclature and numbering ensures acceptability of out-of-scope section exclusion.
- Enables automation of exclusion and replacement page generation.

Using ICH E3 guidelines for CSR (and appendix) structuring



Using ICH E3 guidelines for CSR (and appendix) structuring



Summary



**PRIVACY
ANALYTICS**

an IQVIA company

The three cornerstones of transparency-by-design

Writing clinical documents with disclosure in mind

Minimization

Only include necessary personal and commercial data in documents.

Only include necessary documents in submission package.



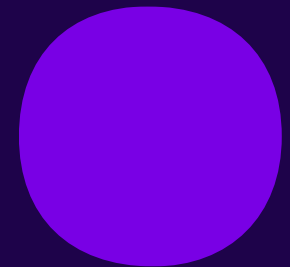
Normalization

Present variables in a limited number of predictable formats for ease of detection, classification and protection.

Standardization

Use the ICHE3 guidance for CSR structuring and include personal data only in pre-specified sections.

3. Highlight: Approaches to enhance cross-functional stakeholder collaboration



Disclaimer - Sanofi

- The views and opinions expressed in this presentation are those of the authors and do not necessarily reflect the official policy or position of Sanofi.
- The information provided is for educational and informational purposes only and is based on the authors' professional experience.
- No confidential or proprietary information is disclosed.
- This presentation does not constitute legal, regulatory, or professional advice.
- Sanofi assumes no liability for any actions taken based on the information presented.



Chantal Din-Bell

Senior Medical Writer, Sanofi

- Clinical Documentation Activity Lead for IB/IMPD and Document De-identification
- Subject Matter Expert for EU-CTR Document Redaction
- Business Lead for regulatory submission of anonymized clinical documents
- 20+ years in clinical research (Biostatistics → Medical Writing)
- MSc in Public Health (Biostatistics), University of Paris V

Clinical Documentation

6 Locations: US, France, Hungary, India, China, Japan



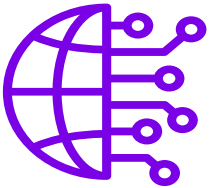
sanofi

A solid and successful 9-year partnership

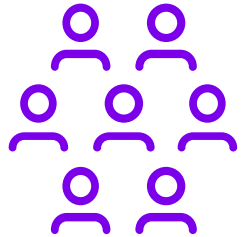


Initiated as proactive response to evolving regulatory landscape

- EMA Policy 0070,
- HC PRCI

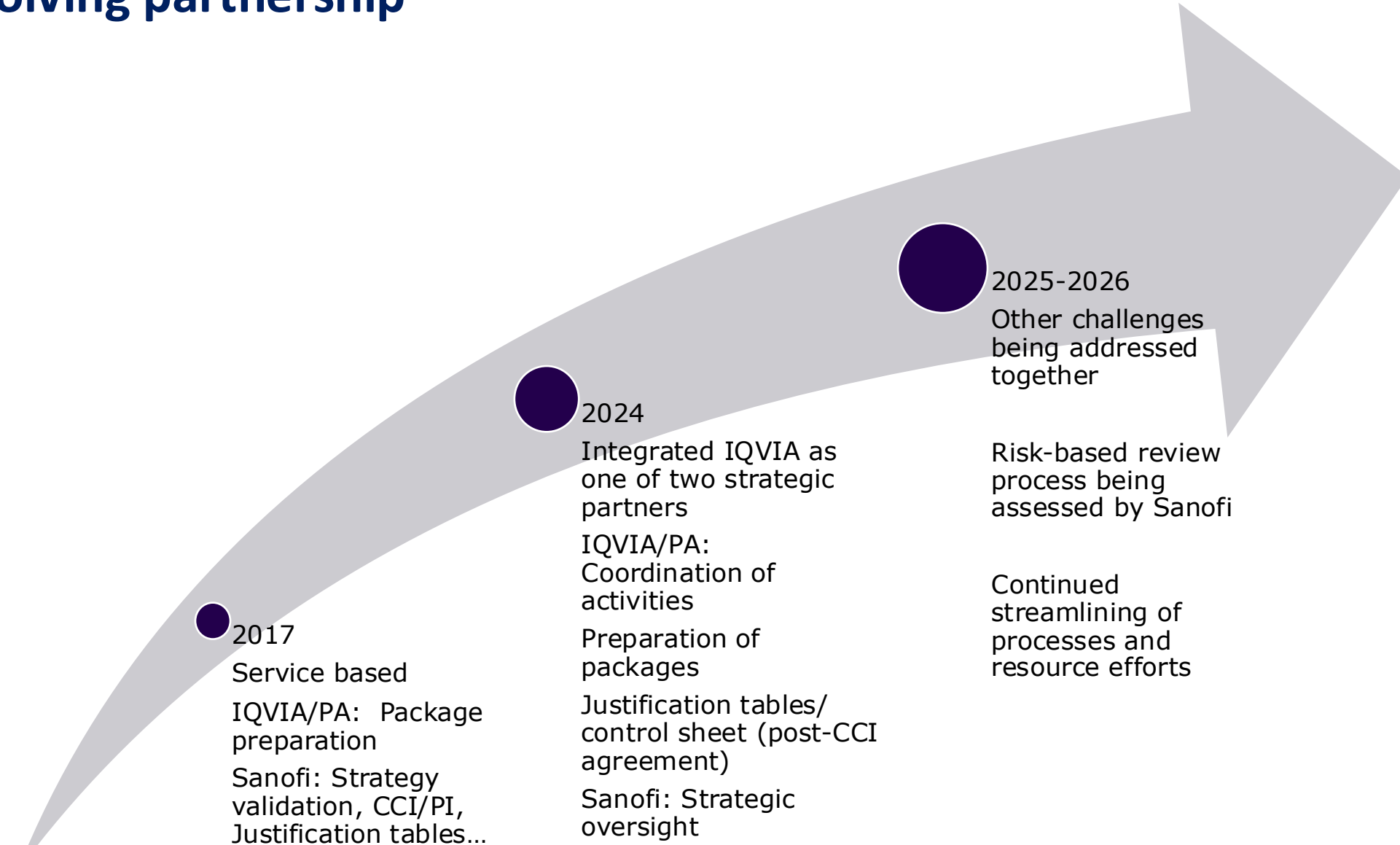


- Leverage external expertise

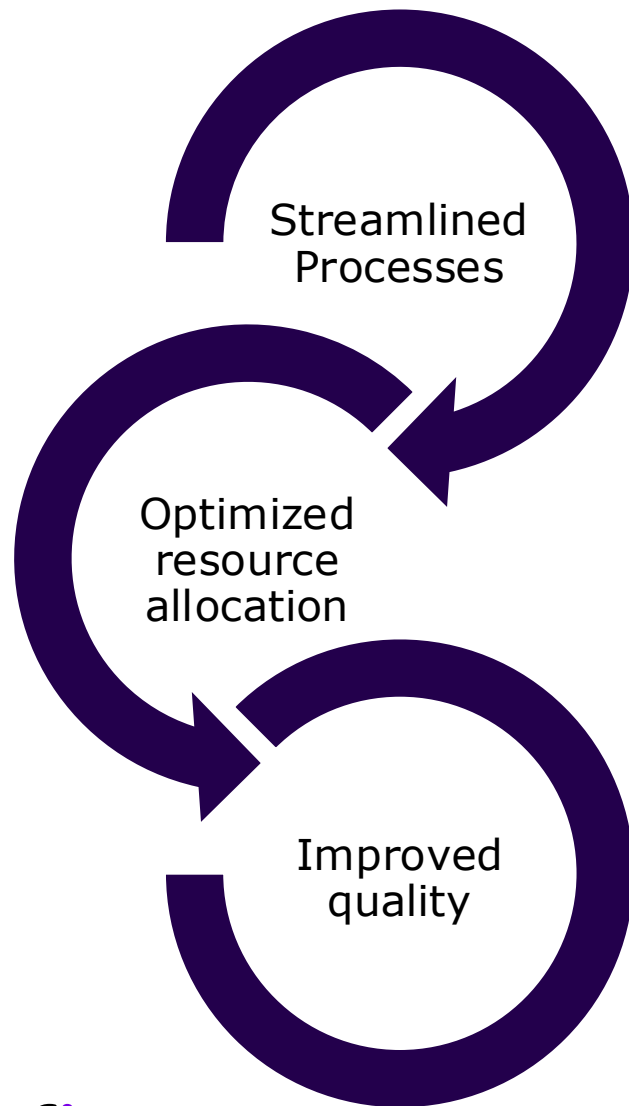


- Strong cross-functional Sanofi team
 - › Regulatory
 - › Clinical
 - › 'Data owners': Biostatistics, PK etc.
 - › Patent
 - › Publications
 - › Clinical Documentation (Medical writing, electronic document specialists, trial disclosure managers)

An evolving partnership



Solutions-based collaboration, and many packages submitted



- Secure document workflow:
- Proactive identification of scope
- End-to-end coordination

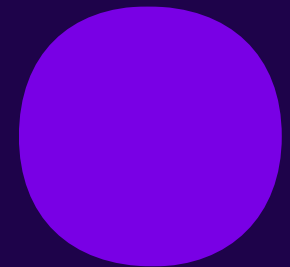
2 EMA
submitted

- Project forecasting
- Clear roles /responsibilities
- IQVIA/PA: Meetings, package prep, control sheets, document updates
- Sanofi: Strategic decisions, CCI identification, HA communication, approvals, submissions

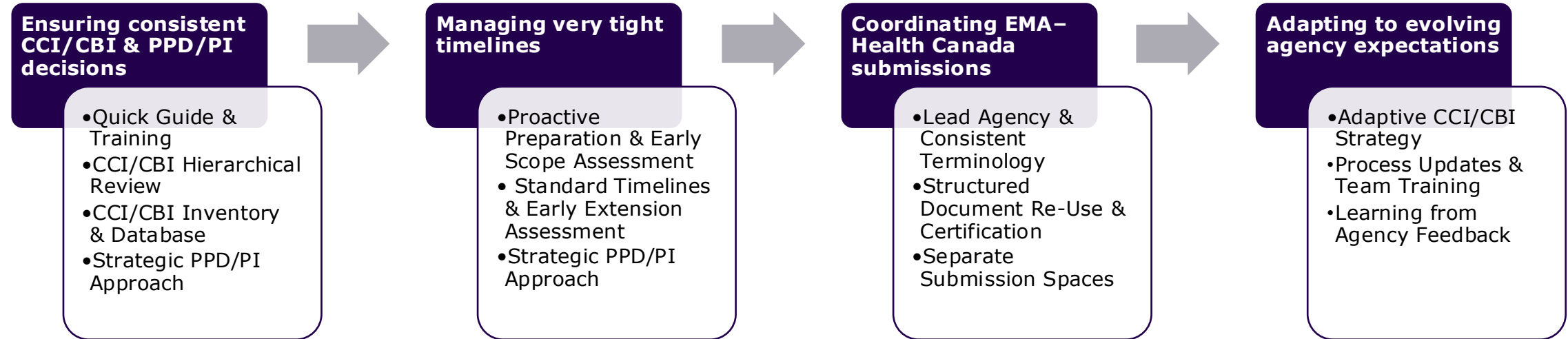
- Well defined QC stages
- Leverage combined expertise
- Trust and recognition of IQVIA/PA regulatory expertise

20 HC
submitted

4. Outline: Strategies for developing and monitoring organizational transparency guidelines



Key Challenges — And How We Addressed Them



From Complexity to Clarity: Our CCI/CBI & PPD/PI Review Toolkit

Key Sanofi-developed resources for P0070/PRCI reviews

Resource	Description
Quick Guide on P0070/PRCI Review	Comprehensive resource covering PPD/PI strategic review approach and CCI/CBI identification principles
Redaction Playbook	Step-by-step guidance for the entire redaction process with real-world examples
EMA P70 / HC PRCI Operational Guidance	Operations-focused guide for document preparation & compliance
How to Identify CCI/CBI	Self-training resource with accepted/rejected examples, lessons learned and best practices
Dedicated CCI/CBI Training	Focused session organized 1 week before document review starts
CCI Inventory	Tracks CCI/CBI at product level and HA decisions, ensuring consistency across documents
Redaction SharePoint	Central hub including KOM slides, Job Aids, CCI training slides and microplanning

A learning partnership

...and the journey hasn't finished!



👉 Strategic partnership → Long collaboration built on trust, expertise & mutual learning

🎯 Proven track record → High volume of complex dossiers for multiple agencies

📋 Robust toolkit → Sanofi-developed resources ensuring consistency and compliance

🚀 Continuous improvement → Evolving processes to meet regulatory challenges

Thank
you

sanofi



To connect with either of our speakers:

Obaraboye Olude –

oolude@privacyanalytics.com

Chantal Din-Bell –

Chantal.Din-Bell@sanofi.com