IMPERIAL
PROJECTS

The NHS Federated Data Platform Evaluation Strategy

An independent evaluation of the NHS Federated Data Platform commissioned by NHS England and led by a team of independent experts from Imperial College London

Table of Contents

INTRODUCTION **4**

EVALUATION OVERVIEW **5**

WP1: Logic model refinement and scoping 7

WP2: FDP's Achievement of Objectives, Outcomes, and Impacts 8

WP3: Exploring FDP's Implementation, Unintended Outcomes, and Wider Societal Impacts 11

WP4: Integrated National Survey of FDP Implementation, Barriers, Enablers and Wider Impacts 14

WP5: Evaluating FDP's Value for Money 15

ETHICS **16**

TIMELINE **16**

RISKS **17**

.

This document presents an initial proposal for the evaluation methodology of the NHS Federated Data Platform. It is a working document, subject to review and refinement. Any reference to this document should acknowledge its draft status and should not be presented as final findings or recommendations.

INTRODUCTION

This document sets out the refined evaluation strategy for the **independent evaluation** of the **NHS Federated Data Platform (FDP)**, commissioned by NHS England (NHSE) and led by a team of independent experts from Imperial College London. It provides an overview of the evaluation approach, its purpose, and the rationale for the revision following the tender stage.

The strategy is a **revised version** of the evaluation approach **submitted at tender**. The overall scope, timelines, and sequencing **remain unchanged**, with the evaluation continuing to be structured across five work packages (WPs) to be carried out throughout three years. The revised strategy aligns with Government Major Projects Portfolio (GMPP) evaluation standards and is designed to deliver robust, actionable evidence to support learning, accountability, and future decision-making.

As this is an independent evaluation, its methodological approach remains under the ownership of the independent experts from Imperial College London who will continue to refine the evaluation as it develops. The version presented here reflects the **current stage of the process** and will be **updated accordingly**; it should **not** be treated as final or fixed for the remainder of the evaluation. Changes may occur in response to emerging findings as the work progresses (e.g. data availability and key stakeholder input).

EVALUATION OVERVIEW

The evaluation methodology is structured across **five WPs** (*see Figure 1*).

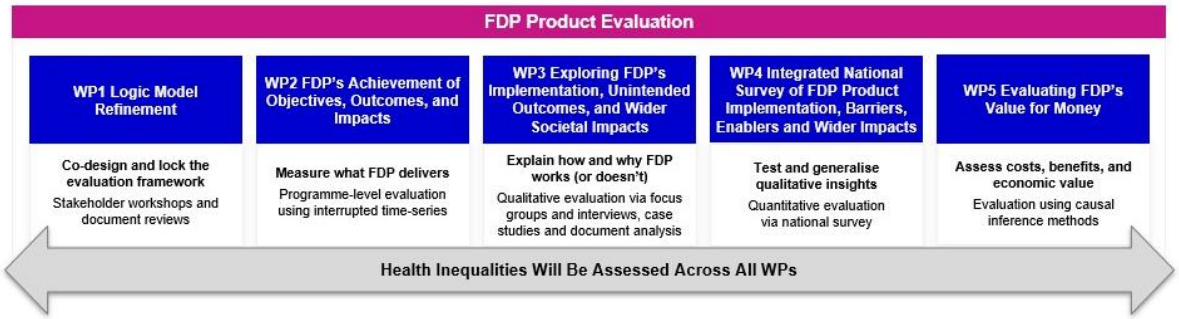

Figure 1: Evaluation Strategy Overview

WP1 refines the programme theory and product-level logic models; WP2 evaluates the impact and benefits of FDP products using quantitative methods; WP3 explores implementation, unintended consequences, and wider societal impacts through qualitative research; WP4 scales these insights nationally via surveys; and WP5 assesses value for money (VfM) through integrated economic appraisal. WP3-5 will focus on the products prioritised in WP1-2, with ongoing evaluation.

While **WP2 initially proceeds with analyses for CCS IP Theatre Management and OPTICA,** discussions on three **additional products** are underway. The delivery plan assumes that data access for these products will be granted by the **end of Q3 of Year 1 (January 2027),** enabling quantitative evaluation of additional products within the WP timeframe. Data access requests will be submitted, with prioritisation agreed upon with NHSE well before this milestone. The timely provision of access depends on NHSE. The evaluation of **health inequalities** remains a fundamental component of this programme; accordingly, **it has been embedded across all WPs rather than delivered as a standalone WP**. Each WP includes a targeted health-inequalities component aligned with NHSE priorities, assessing whether FDP deployment reduces disparities in access and outcomes and identifying opportunities to enhance benefits for underserved populations. Access to health inequalities data is currently being explored with NHSE. The proposed analyses are therefore subject to the availability, completeness, and quality of the relevant data sources and may be refined as these are confirmed during the data access phase.

Table 1 summarises **the evaluation framework**, mapping each evaluation question to the relevant WP and the methods used to generate the required evidence.

Table 1: Evaluation questions, WPs and methods			
#	Evaluation question	Lead WP	Potential Methods
1	How well has FDP achieved its objectives, intended outcomes and long-term impacts and are these sustainable?	WP2 (informed by WP1)	Product-level logic models, programme-level logic model, trust-level interrupted time series (controlled where feasible), with meta-analytic pooling, difference-in-differences (DiD) analyses and statistical process control (SPC) charts
2	What factors have contributed positively and negatively to the implementation of FDP, and what are the lessons for future digital programmes?	WP3, WP4	Cross-Trust focus groups and semi-structured interviews; documentation review; national survey; triangulated synthesis
3	What wider societal impacts is FDP having, including social value, local economies, and cultural change?	WP3, WP4	Cross-Trust focus groups and semi-structured interviews; case studies; national survey; triangulated synthesis
4	What have been the unintended and unexpected outcomes from FDP, positive and negative?	WP3, WP4	Cross-Trust focus groups and semi-structured interviews; national survey; case studies; triangulated synthesis and comparison against product-level logic model expectations
5	What impact does FDP have in reducing health inequalities and what more can it do?	Cross-cutting (all WPs)	Stratified quantitative analyses and interaction terms (Indices of Multiple Deprivation (IMD), ethnicity), subject to data completeness; inequalities lens in logic models; survey.
6	Comparing costs with outcomes, what is FDP's short- and long-term value for money and where should resources be targeted?	WP5 (drawing on WP2)	Product-specific economic appraisal; incremental costs vs monetisable outcomes; cash-releasing vs non-cash-releasing savings; heterogeneity by site and uptake.

WP1: Logic model refinement and scoping

WP1 establishes the analytical foundation for the evaluation. For each product, we will develop a **draft "version 0" programme logic model** and **outcome framework in advance**, grounded in programme documentation, product specifications and published evidence. These refined logic models will guide outcome selection, data requirements, analyses, and interpretation for each product. They will be treated as **living frameworks** and iteratively refined throughout the evaluation lifecycle.

Alongside the product-level models, an **overarching FDP logic model**, synthesising the shared platform-level mechanisms and strategic outcomes common across products, will be maintained and refined **once the product-level models have been agreed**, providing the programme-level theory of change required for the final evaluation report.

Facilitated workshops with NHSE experts, product owners, delivery leads from selected NHS regions and Trusts, and frontline operational users will then be used to agree and refine these models, rather than build them from scratch. The aim is to **maximise the value** of participants' time, develop a **shared understanding** of FDP products, operational context, workflows, evaluation metrics, and priority inequality measures, ensuring alignment with the evaluation questions and both formative and summative aims.

Workshop discussions will be recorded with participants' agreement and supplemented with facilitator notes. This data will be **analysed thematically** using a rapid deductive approach structured around workshop questions, whilst remaining open to themes that fall outside this framework. We will use **programme theory** to outline the **causal pathways** between inputs (data and people layers) and mechanisms, illustrating how these FDP products are expected to lead to **short-** and **long-term outcomes**. Preconditions for benefit realisation will be made explicit, and each outcome will be marked for suitability for quantitative or qualitative exploration or measurement, as well as for data feasibility. Health inequalities are incorporated as **a cross-cutting lens** in each logic model, and approaches to assessing differential impacts are also agreed upon, where data are available.

WP1 will also determine which Trusts to include in the quantitative analyses and to **purposively select case study Trusts for qualitative work** in subsequent WPs. **Selection criteria have been defined by the evaluation team** and reflect variation in FDP adoption

(earlier vs. later), NHS digital maturity (via published metrics), case-mix, waiting lists, hospital type, geography, system relationships, populations served (including levels of socioeconomic deprivation and other characteristics relevant to health inequities, where available and provided to us in the data extracts), and existence of baseline data to compare pre- and post-implementation. **Candidate outcome metrics for each product have been drafted by the evaluation team** and are grounded in the product-level logic models, product specifications and prior evaluation work. The **prioritisation of these metrics** will take place at the workshops with stakeholder input and expertise. For the quantitative analysis, all deployed Trusts with **sufficient available pre- and post-implementation data will be included**, with Trust-level profiling of this data to characterise the cohort and support the interpretation of heterogeneity. To mitigate positive selection bias, sites where implementation has been slower than expected will be deliberately included. This is to ensure **maximal representation** of hospital-, patient-, and system-level factors across the country.

WP2: FDP's Achievement of Objectives, Outcomes, and Impacts

WP2 will deliver the **programme-level evaluation of impact and outcomes**, focusing on whether FDP products are achieving their stated objectives for patients, staff, the wider system, and their sustainability over time. We will conduct a **quantitative evaluation of FDP data,** using national and Trust-level data stored within the FDP, alongside product-usage metrics, and national datasets.

The evaluation will be conducted on a product-by-product basis, and the same methodology will be applied to subsequent FDP products as the evaluation progresses.

The methodological approaches proposed below are subject to data access, completeness and quality, and depend on availability constraints where relevant. Since no pre-intervention baseline exists for FDP-generated usage metrics, these metrics will be used as adoption-intensity moderators. We will evaluate the suitability of various methods based on the characteristics of the data, proceeding from descriptive to inferential approaches, including:

- **Descriptive** and **exploratory analysis** will form the foundation of the quantitative work. For each product and Trust, we will describe the available data, visualise outcome series over time, summarise activity and performance pre- and post-

deployment, and analyse the distribution of product usage and adoption depth across Trusts and specialties. This will provide an accessible first view of what the data show, inform the specification of the subsequent statistical models and can flag data quality issues before they can bias later estimates.

- Wherever possible, Trust-level **statistical process control (SPC) charts** will provide an additional **descriptive layer** for each Trust, distinguishing routine (common-cause) variation from significant change (special-cause) and revealing the timing and shape of any shift.
- **Quasi-experimental methods** such as i**nterrupted time series (ITS)** analysis will be used at the Trust level to estimate both level and trend changes in outcomes following each Trust's deployment date and assesslong-term impacts and the sustainability of change. If suitable, Trust-level estimates will then be pooled in a **meta-analysis** and presented on a forest plot, making **heterogeneity across Trusts a finding in its own right rather than just a limitation**. Pooling will include all Trusts for which a valid Trust-level estimate can be produced, with small Trusts accommodated within the meta-analysis through precision weighting. Trust characteristics (e.g. Trust size, digital maturity score quartile, baseline pressure) will **support exploration** of that heterogeneity, subject to the number of contributing Trusts. This structure also accommodates the **staggered rollout across Trusts**. Where data allow (e.g., sufficient baseline/historical data), we will aim to use the **same primary analytical design across all products**, with the main difference being the outcomes assessed, which will be selected for each product in WP1 to reflect its nature.
- Where possible, a **controlled ITS** analysis will be conducted, comparing with Trusts that have not deployed the relevant product, depending on the granularity and coverage of publicly available data sources. Where a suitable public dataset covers non-deployed Trusts, comparator Trusts will be identified through propensity score matching on Trust characteristics. Where no public comparator series exists, a within-Trust design comparing specialities with high and low product usage or adoption depth will be utilised. High and low product usage will be defined *a priori*.
- Where usage-intensity data permit, we will additionally examine dose-response relationships through **difference-in-differences (DiD) comparisons** as a secondary objective.

Metrics will be guided by the logic models developed in WP1. Data completeness is treated as an explicit analytical risk: several relevant fields in the product specifications are recommended or optional rather than required, and completeness will be assessed per Trust before analysis proceeds. **Health-inequalities analyses** will assess outcomes by Indices of Multiple Deprivation (IMD) quintile, ethnicity, and other proxies, using stratified designs and interaction terms where feasible to test whether FDP deployment reduces or widens disparities. These analyses are conditional on data access and completeness.

Data quality and completeness

Robust evaluation depends on the quality and completeness of the underlying data. FDP products draw on multiple data sources, each with varying degrees of completeness, definitional consistency, and sensitivity to changes in recording practices across products, trusts, and specialties over time. As with any evaluation reliant on routinely collected administrative data, some degree of missingness, miscoding or definitional change is expected and cannot be fully eliminated.

The evaluation will aim to identify and address these issues as thoroughly as possible ahead of outcome analysis. Data quality checks, including completeness, internal consistency and cross-source validation, will be conducted for each dataset prior to analysis. Material discrepancies or Trust-level reporting anomalies will be documented and, where necessary, addressed through sensitivity analysis. Fields with substantial or systematic missingness will be flagged transparently rather than imputed without justification.

While these steps improve confidence in the reliability of the underlying data, they cannot guarantee complete or error-free data across all trusts and time periods. Findings will therefore be reported with due regard to any known data quality limitations affecting specific metrics, Trusts or time periods.

Strength of evidence and causal inference

ITS analysis is **among the strongest quasi-experimental designs for evaluating system-level interventions** and represents the most feasible and proportionate approach in the context of the operational deployment of FDP products across NHS Trusts, where randomised allocation

is neither feasible nor appropriate. As with any observational design, it **cannot establish definitive causality,** and observed changes may be influenced by concurrent initiatives, secular trends, seasonal effects, and changes in case mix that coincide with implementation.

This evaluation design seeks to **tackle these limitations as thoroughly as possible.** For example, study design, outcome definitions, and analytic decisions (including comparator selection and the classification of high vs low adoption) will be pre-specified in a statistical analysis plan agreed before outcome analysis commences. Where feasible, **controlled ITS** incorporating comparator Trusts or **DiD comparisons** will account for underlying temporal trends unrelated to the intervention. Whenever possible, **Trust-level SPC charts** will further support interpretation (helping to distinguish genuine special-cause change from routine variation or changes in recording practice) and, where sufficient methodological and clinical comparability exists, meta-analytic synthesis across Trusts will improve the precision of effect estimates and test the consistency of impacts. Effects that replicate across many Trusts, deployment dates and contexts are less plausibly explained by local confounding than any single-Trust result. Furthermore, upcoming qualitative and survey work packages will ground these impact analyses on the real experiences of those using the products.

While these approaches **strengthen the robustness of the evaluation** and **increase confidence** that observed changes are attributable to the FDP, **they cannot fully eliminate residual confounding or establish direct causality**. Findings, therefore, will be reported with appropriate language regarding causal attribution, consistent with Magenta Book guidance on evidence from quasi-experimental designs.

WP3: Exploring FDP's Implementation, Unintended Outcomes, and Wider Societal Impacts

This WP will use a ***single, integrated programme*** **of qualitative enquiry to examine three evaluation questions:** a) factors influencing implementation; b) unintended outcomes from FDP (not captured in classic quantitative metrics); c) wider societal impacts. The use of an integrated programme to address these aspects will allow comprehensive coverage, while **minimising participant burden and maximising participation.** Site and participant selection

will be informed by Trust profiling (WP1) and emerging quantitative findings from WP2, to ensure representation across hospital-, patient-, and system-level factors nationally (see WP1). Additionally, **documentation reviews** and **case studies** will be conducted to capture additional nuance, as required (*see Figure 2*).

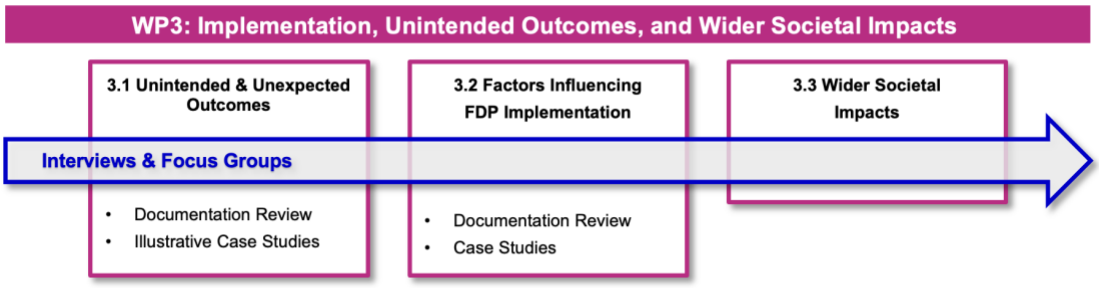

Figure 2: WP3 Activities Mapped to the Research Questions

The qualitative design will combine focus groups and interviews to balance collective reflection with in-depth individual insight. It will include the products identified in WP1, enabling comparison of implementation factors across products and organisational contexts within a single WP. **Four cross-Trust focus groups** (n=6 per group; total n=24) will be held to support structured brainstorming and cross-site comparison, with recruitment spanning all participating Trusts and at least one group anchored to each product. In addition, semi-structured interviews will be conducted using a tiered approach: approximately five interviews at each case-study Trust (one per product), complemented by one or two key informant interviews at the remaining selected Trusts (n=5-6). This will provide an anticipated 35-40 interviews in total. This **concentrates in-depth enquiry where variation is greatest** while retaining coverage of all selected sites and remains proportionate for the number of products examined. Interviews will capture contextual details, dissenting views, and role-specific nuances that are less likely to emerge in group settings. Participants will include NHSE programme and delivery teams, implementation leads, clinical and operational users, analysts, digital workforce leads, and, where relevant, equity champions or patient representatives to explore **effects on underserved populations**. A single modular topic guide will cover all products evaluated in WP2 to date to address the three evaluation questions outlined below in the focus groups/interviews.

3.1 Unintended and unexpected outcomes. Targeted questions will capture unintended risks or harms (e.g., increased workload) and emergent benefits (e.g., novel use cases). Thematic analysis will map outcomes against FDP objectives, illustrating both positive contributions and tensions. Findings will be triangulated with a structured review of relevant documentation, including risk logs, incident reports, user feedback and governance papers to systematically identify reported issues and adaptations associated with FDP use. Additionally, two or three illustrative case studies of unexpected use cases or impacts in selected Trusts will be developed, examining the context, mechanisms, and outcomes in depth, highlighting transferable lessons and conditions under which unintended effects are more likely. Outputs will include a **structured typology of unintended outcomes** (positive and negative) and **recommendations** to amplify benefits, mitigate risks, and improve horizon-scanning for future unintended effects.

3.2. Factors influencing FDP implementation. Topic guide questions will explore operational barriers, enablers, tensions, and trade-offs that influence the successful implementation of FDP products. Areas of focus and improvement will include product design (e.g., usability), existing onboarding processes, configuration, workflow integration, skills and training across different organisational contexts. The aim is to generate practical learning on what has worked well, what has not, and why. An implementation framework, such as the Consolidated Framework for Implementation Research, will inform this work. Findings will be complemented by document reviews (e.g., onboarding materials, deployment guides, and issue logs). Triangulation of interview and documentary findings will support an additional two or three comparative case studies across different contexts, producing actionable recommendations, such as refined use cases, deployment and equity adjustments.

3.3. Wider societal impacts. Topic guide questions will explore: a) social value for patients and communities (e.g., improved care coordination, reduced harm from long waits or delayed discharge, more equitable access and experience); b) local economic benefits (avoided duplication, capacity and efficiencies reinvested into frontline care and local community services), with attention to how benefits are equitably distributed; c) cultural and system change (e.g., strengthened data literacy, equitable use of data to identify and address disparities, and inclusive digital participation).

WP4: Integrated National Survey of FDP Implementation, Barriers, Enablers and Wider Impacts

A structured online survey, **explicitly informed by findings from WP1-WP3,** will be developed and distributed nationally to test and generalise key themes identified in WP3, providing **national-level evidence**. The survey will be organised into modular sections aligned with the three evaluation questions explored in WP3 across the FDP products evaluated to date, enabling respondents to contribute across multiple domains. It will be disseminated via Qualtrics to FDP users, local implementation teams, and programme or system leaders (target n = 60-80), with minimum representation across stakeholder groups and products.

The survey will be designed to test the **transferability of WP3 themes to a broader national audience**; the target, therefore, reflects coverage across stakeholder groups, products and organisational contexts. Response monitoring will track respondent composition and include targeted follow-up to address gaps. Non-response bias will be assessed by comparing early and late responders on key items as a proxy for non-respondent views, and the achieved sample will be benchmarked against the known characteristics of the deployed Trusts. Where the sample is skewed on known dimensions, key results will be reported with **stratified sensitivity analyses** to test the stability of conclusions across subgroups. Given the absence of a defined user-level sampling frame, no non-response weighting will be applied, and the characteristics of the achieved sample will be reported transparently and taken into account in interpretation.

Quantitative survey data will be summarised using **descriptive statistics**, while free-text responses will be **analysed thematically using a coding framework** aligned with the selected implementation model. Convergent data triangulation of quantitative patterns (WP4) and qualitative themes (WP3) will enhance robustness, mitigate source biases, reconcile dissonant findings, and generate **a multi-method synthesis of evidence** on barriers, enablers, unintended outcomes, and wider impacts influencing FDP implementation.

WP5: Evaluating FDP's Value for Money

VfM will be assessed through a stepwise approach, with depth of analysis for each product **determined by data availability**. The **primary focus** is an assessment of **operational impact** (e.g., list utilisation, length of stay), including downstream effects (e.g., discharge destination, readmissions) where data allows, primarily as balancing measures to sense-check operational gains rather than as additional monetised benefits. Where feasible, these operational improvements will be translated into indicative monetary values using published unit costs and national reference data. A broader quantification of costs relative to patient or health benefits will be addressed qualitatively (WPs 2-3), with any additional quantitative assessment contingent on data availability. This approach will be applied consistently as data availability is confirmed and as the evidence base develops.

The evaluation will be conducted on a **product-by-product** basis, with CCS IP Theatre Management Module and OPTICA identified as initial priority products, followed by 1-3 subsequent FDP products as scoping progresses. Where meaningful, **analyses will examine heterogeneity** by Trust and uptake to **characterise the range of VfM observed and the factors associated with stronger or weaker performance**. Methodological requirements and data feasibility will be scoped upfront with input from NHSE stakeholders.

We will assess the feasibility of using a range of datasets to support this. For the operational tier, this includes: (I) FDP rollout timing and usage intensity metrics and (II) Trust impact estimates from WP2 and HES. For the cost tier, this includes activity and cost/financial data, supplemented by standard unit costs where required and where access permits.

Monetisable operational outcomes will focus on those that can be **robustly estimated from the WP2 analyses** and will be tailored to each product, valued using published unit costs from publicly available sources. Staff time savings will be reported separately as non-cash-releasing productivity gains to avoid overstating financial benefit. **Economic values will be transparently derived**, with **sensitivity analyses** conducted around key assumptions.

Where **FDP-enabled changes translate into measurable patient health impacts** identified in WP2, these will be incorporated as **health gains in natural units** and monetised where appropriate, using the unit sources above. Outcomes that are too many causal steps removed from what a product directly achieves (e.g. avoided emergency admissions, complications,

mortality) will not be monetised. To **calculate net VfM,** economic benefits (operational and where estimable, patient-level) will be set against incremental deployment and running costs, drawing on Trust-level cost data where available and plausible estimates where it is not.

ETHICS

As per **Health Research Authority guidance**, the evaluation is classified as **a service evaluation** and is supported as such by NHSE. The classification will be documented alongside Imperial College London's institutional processes, and any elements subsequently identified as requiring further approvals will be addressed before the relevant data collection begins. For the qualitative work, NHSE will manage participant invitations and consent, with the research team obtaining additional informed consent before qualitative data collection begins. No identifiable participant information will be stored as part of the qualitative work packages.

TIMELINE

In accordance with the contract, the final report will be shared at the **end of the project**, in **April 2029**.

RISKS

We are using a **live risk register** to track likelihood, impact and mitigation. Risks will be reviewed with NHSE monthly.

IMPERIAL
PROJECTS

Contact us:

Imperial Projects
E: iproj@imperial.ac.uk