DRAFT IMPLEMENTATION PLAN FOR A LINKAGE TOOL

euvabeco

Document Properties

Grant	**101132545 EUVABECO**	**Deliverable Title**	**DRAFT IMPLEMENTATION PLAN FOR A LINKAGE TOOL**		
Deliverable N°	D5.1/2	Type	Document, report	Dissemination Level	PU-Public
Due date	2024-10-31	Actual submission date	2024-10-28		
Lead Beneficiary	SCI – SCIENSANO				
Contributor(s)	AES : Bonsang Joé; Delporte maud; Vaccaroli Raffaella	SCI : Billuart Matthieu; Nasiadka Leonore; van Loenhout Joris			
	CIM: Cimino Alain; Cimino Mariane; Jackson Sally	SYA: Kaag François; Koeck Evarita; Koeck Jean Louis; Laporte Mathieu			
	DGS: Alves Bruno; Pereira Natália; Arriaga Miguel; Cardoso Susana	UoC: Anastasaki Marilena; Angelaki Agapi; Biggs Sarah; Chlouverakis Gregory; Galanakis Emmanouel; Galenianos Myron; Kochiadakis Georgios; Kofteridis Diamantis; Lionis Christos; Marketou Maria; Papadakis Sophia; Petelos Elena; Roumeliotaki Theano; Stavroula Tsinorema; Vasilaki Eirini			
	FRT: Collart Gaëtan; Désirant Christophe; Meyers Philippe; Pollet Isabelle; Vandenberghe André	UoT: Anagnostopoulou Lemonia; Georgalis Leonidas			
	JU: Duplaga Mariusz; Grysztar Marcin; Halik Rafał; Jakubowski Szczepan; Krężel Gabriela; Sikorska Magdalena; Stropalova Olena	USAAR: Becker Sören; Bragazzi Nicola; Lehr Thorsten; Ngbede Emmanuel			
	LIU: Faresjo Ashlild; Faresjo Tomas; Iredahl Fredrik; Lebana Andrea	VE: Czwarno Anna; Quilici Sibilia; Weindorfer Ingrid			
	RSU: Ozolina Kristine; Zavadska Dace	WMU: Kurpas Donata; Lomper Katarzyna; Manulik Stanisław; Soll - Morka Aneta; Stefanicka - Wojtas Dorota; Uchmanowicz Bartosz; Uchmanowicz Izabella			

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Document History

Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or HaDEA. Neither the European Union nor the granting authority can be held responsible for them.

Version	Date	Reason
0.0	2024-05-27	Initialization
1.0	2024-10-25	Draft finalised

Table of Contents

List of Abbreviations and Acronyms

Abbreviation / Acronym	Meaning
DPIA	Data protection impact assessment
DPO	Data protection officer
GDPR	General data protection regulation
HCW(s)	Healthcare worker(s)
ID	Identifier
IT	Information technology
MFA	Multifactor authentication
PPRL	Privacy-preserving record linkage
RBAC	Role-based access control
SCRA	Small-cell risk assessment
TTP(s)	Trusted third party(ies)
UPI	Unique patient identifier

List of Stakeholders

Stakeholders	Meaning
Health Authorities	Competent bodies focusing on the creation and application of health-related policies
Database Owner	Entity or organisation legally controlling a database
Trusted Third-party	Independent, neutral intermediary entity trusted by both sending and receiving parties to guarantee the security of a data exchange
Legal Authority	Regulatory body controlling and authorising the implementation of the data linkage. Depending on the implementer's situation, it might overlap with the definition of the information security regulatory body
Social Security	Public healthcare insurance organisation
Ethics committee	Independent group of individuals ensuring that research is conducted ethically and in accordance with legal and moral standards
DPO	Person ensuring, in an independent manner, that an organization applies the laws protecting individuals' personal data
Information security regulatory body	Competent entity to deliver legal authorisation for the electronic exchange of personal data

List of Figures

The purpose of EUVABECO is to deliver to Member States implementation plans for several tools able to support existing or future vaccination practices.
These implementation plans are practical guides for a Member State to decide upon the launch of an implementation project, assign adequate resources, deploy the given tool and keep it operational after deployment.
They are structured with three main sections:

- Description of the tool is a functional analysis of the tool with an overview, the stakeholders using or contributing to the use of the tool, their respective functional requirements, the non-functional requirements, and a collection of use cases illustrating the desired functions.
- Prerequisites lists the contextual conditions that must be met before the project is launched, and a few workarounds that could be used to anticipate upon their fulfilment.
- Implementation addresses the actual implementation, with the architecture, resources, workflow and planning for the build phase, and the missions to be ensured during the run phase to keep the tool operational.

1 Description of the tool

This section describes functionally the intended tool and its usage. It does not correspond to any specific implementation of the tool.

1.1 Objectives

This section is the overall rationale for the tool.

The Linkage tool relies on the establishment of data linkage between several pre-existing data sources, on an individual level, aiming for several potential applications related to vaccinations (in routine or crisis context). One such application consists of screening the population based on specific health and demographic criteria to issue priority vaccination invitations. This process ensures that individuals who are defined at higher risk receive timely notifications and invitations to get vaccinated. Additionally, data linkage can be used to monitor the pillars of vaccination surveillance (coverage, effectiveness and safety). By keeping track of these elements, health authorities can ensure the effective monitoring of the vaccination campaign.

1.2 Involved stakeholders and their expectations

These are all the actors within the implementing Member State using or contributing to the use of the tool once it has been implemented. Their expectations are requirements for any implementation of the tool.

The key stakeholders are:

- Health authorities
- Database owners
- Trusted third party
- Legal authorities
- Citizens

1.2.1 Health authorities

Health authorities are aiming for a safe and effective implementation of their vaccination policy, be it for routine vaccination or targeted campaign in a crisis context. To assess its success, a precise and detailed monitoring of the different components of a vaccination roll-out, and the evaluation of the impact and outcomes of vaccination are needed. Health authorities are responsible for inviting the right persons according to pre-defined criteria. As such, they expect information that can help defining the vaccination strategies, determine the prioritized population and improve public health communication strategies, both for routine vaccination as well as during public health emergencies.

1.2.2 Database owners

Database owners require a legal framework that ensures and permits the transfer of their data, as well as clear guidelines on data handling responsibilities and protocols. They expect a secure and robust data exchange. Additionally, a financial compensation is expected for their involvement, to support the provision and maintenance of the services, as this will require staff time from the side of the database owner.

1.2.3 Trusted third party

Trusted third parties (TTPs) require legal authorisation in order to process and handle data. They aim to set up a solid data flow and guarantee the reliability and security of data exchange. They are in charge of the pseudonymisation of the datasets.

1.2.4 Legal authorities

Legal authorities are responsible for approving data linkage requests, according to a certain need. They require full compliance with the General Data Protection Regulation (GDPR) and related legislations.

1.2.5 Citizens

Citizens should be informed, alerted and recommended vaccinations against diseases they are defined as the most at risk for. They expect to receive a relevant and timely invitation / notification for vaccination. Additionally, they should be aware on how their data is being used, what information is being collected and have access to publications/reports built with it. They expect that their data are processed securely, and their privacy is guaranteed.

1.3 Constraints

Constraints are the non-functional requirements on the tool. They do not correspond to a function to be performed by the tool, but not respecting them would impair the viability of the tool.

Availability of databases and data quality

The number and diversity of relevant databases define the data linkage range of applications. The quality of the data will determine the performance of the linkage tool. The quality can be split up in three criteria:

- Completeness: occurrence in the real world is exhaustively recorded

- Correctness: the recorded values are a faithful reflection of the true values
- Timeliness: changes in the real world are reflected on the data in a timely manner

Computer literacy among HCW for operational feasibility

For the data linkage process to be used effectively, healthcare workers should have a basic level of computer literacy to interact with their recording/reporting system (primary data collection system). A lack of computer literacy can lead to user errors, data entry mistakes, and improper use of the system, which can compromise the integrity and accuracy of the linked data. The system may not be usable, regardless of its technical capabilities. Healthcare workers with adequate computer skills can use the system more efficiently and are more prone to accept and use the system.

IT infrastructure

The performance of the linkage tool is dependent on the performance of the IT infrastructure. It directly impacts the capacity to provide real-time insight and accurate, continuous monitoring. Several aspects should be considered.

- Timeliness: the frequency and the speed of the data transfers.
- Reliability and availability: the uptime (percentage of time it is operational), redundancy (backup systems ensuring continuity) and resiliency (capacity to recover in case of major failure) of the infrastructure.
- Scalability: the capacity to accommodate changes in the volumes and structures of data, as well as an increased workload.

Available funding

Adequate funding is necessary for the resources (infrastructure, technology, personnel and expertise) required for developing and maintaining the system. Sufficient funding is essential for the initial development and implementation phases, ensuring the system is built to meet the required specifications. Ongoing funding is required to maintain the system, implement new tools, perform data updates and keep the system secure, efficient and up to date.

Culture of data-driven policies among the stakeholders

Data-driven culture among stakeholders impacts the willingness and ability of stakeholders to effectively implement and utilize the data linkage process. A supportive culture is necessary to ensure that the linked data is actively used for policy- and decision-making processes. Without this cultural foundation, the data linkage process may not achieve its intended impact. Stakeholders who value data-driven insights are more likely to prioritize and invest in the necessary infrastructure, training, and resources needed to implement and sustain a data linkage process. More specifically, participation and contribution to the European Health Data Space (EHDS) ecosystem will facilitate implementation of the linkage tool.

Institutional trustworthiness regarding health (data)

The acceptability of the tool depends on citizen confidence that government and other stakeholders will act in their best interests. If public opinion on the tool is positive, there is more willingness to share personal data, and more acceptance of communication regarding outputs

All involved stakeholders should adopt a conduct promoting citizen's acceptance of health data sharing, linkage, and reuse, as well as adhesion to the resulting decisions and communication.

Behaviours fostering trust are competency (demonstrating expertise and knowledge) transparency (being open and honest on operations and decisions), and fairness.

A practical application is effort in the protection of sensitive personal data: protection measure, effective risk management, compliance with standards, etc. In addition, there should be clear and transparent communication about this to the general public.

Multidisciplinary team

For the optimal utilization of the processed data, a collaboration is required between data scientists and public health experts. The first will extract, prepare, and analyse the data while the second will work on interpreting the results interpretation, and translating it in real-world recommendations.

1.4 Use cases

The use cases are illustrative scenarios representing how the actors identified above could use the tool to meet their expectations. There are as many use cases as needed to describe every desired function of the tool.

1.4.1 Vaccination surveillance

Post-authorization vaccine surveillance consists of monitoring various outcomes related to the vaccine. It is used to inform policy decisions, optimize vaccination strategies and allocate resources effectively to improve overall public health outcomes.

The healthcare sector is an information-intensive environment, where the transmission of information can be altered in the event of overload, such as during a public health emergency or the introduction of new vaccine.

Establishing a link between the national vaccination registers and existing databases of national health registers, all of which contain a national UPI, aims to create a prospective cohort of vaccinated people. The data linkage of pre-collected data avoids the need to set up a new prospective data collection system, which add to the burden already imposed on healthcare staff. Such linkage make it possible to monitor vaccination coverage, safety, and effectiveness among the general population as well as in specific subgroups (e.g. elderly people, healthcare workers, nursing home residents,…).

Potential data sources linked to vaccination registry and output associated:

DATA SOURCE	(POSSIBLE) CONTENT	OUTPUT
Laboratory test results database	Data on tested patients Information on test prescriptions, test results (including rapid tests), symptoms, variant, suspected false negatives and false positives	Identification of breakthrough cases Calculation of **vaccine effectiveness against symptomatic infection**
Hospitals clinical database	Data on hospitalized patients (e.g. comorbidities, symptoms, complications, length of stay, treatments, outcome of hospitalization, entry and discharge of intensive care unit, etc.)	Identification and characterization of hospitalized breakthrough cases **Vaccine effectiveness against hospitalization**

Healthcare professional database	Data allowing identification of healthcare workers (HCWs)	Determination of **vaccination coverage among healthcare workers**
National statistics databases	Socio-economic information (family composition, nationality/origin, employment status, income, …)	Differences in **vaccine uptake by:** – **Underlying medical conditions** – **Socio-economic status** – **Socio-demographic groups** Confounders for **vaccine effectiveness** calculations
Insurances databases	Data on reimbursed care and medicines of citizens insured in the country (e.g. pseudo pathologies as comorbidities, nursing home resident status, medications, etc.)	

The different information collected and analysed through the linkage of databases can be used for infographics or communication support for stakeholders involved in policy decisions, as well as for the general population, regarding the almost real-time vaccination coverage during a vaccination campaigns, the effectiveness of the vaccines administered.

1.4.2 Screening for vaccination invitation

Vaccination requires preparation in order to target the people for whom vaccination is the most necessary or effective (specific to an age group, medical condition, profession, risk of exposure, etc.). Linkage of existing databases can help to identify individual for an invitation for vaccination based on chosen characteristics and ensure the protection of those in need.

Example of screening for priority invitation for vaccination based on specific characteristics.

For disease X, a vaccine is available, and the complete population is eligible to get vaccinated. However, certain categories of people have been identified by the national authority as prioritized for vaccination:

- *Individuals presenting certain underlying medical conditions, identified with an increased risk of severe complication in case of infection;*
- *Healthcare workers, identified with an increased risk due to close exposure to patients;*
- *Older age groups.*

Selection means that a person is selected according to a prioritization on the basis of established criteria to be allowed to be vaccinated from a certain moment.

Medical prioritization criteria have been established by national authorities and individuals have been selected either centralized within the healthcare insurances (public (social security) or private insurances) databases or decentralized through the Electronic Medical Record by their general practitioners or specialists. Data allowing identification of healthcare workers are registered in a dedicated HCW database. Individual identification information, from which the date of birth is extracted for age selection, is recorded in a national citizen register.

A dedicated environment is created to host the linkage and the screening procedure. The different databases go through the TTP for the pseudonymization procedure using deterministic encryption, before being imported in this environment. The linkage is thus performed based on the pseudo-UPI. Based on the vaccination recommendation, the priority patients are flagged. Those who died or are already vaccinated are filtered out and a list of pseudo-UPI is extracted.

The deterministic encryption makes it possible to send this list is then back to the TTP for de-pseudonymization. Thanks to this process, the competent authorities is able to contact the prioritized patients without ever knowing the reason of prioritization, thus protecting their privacy.

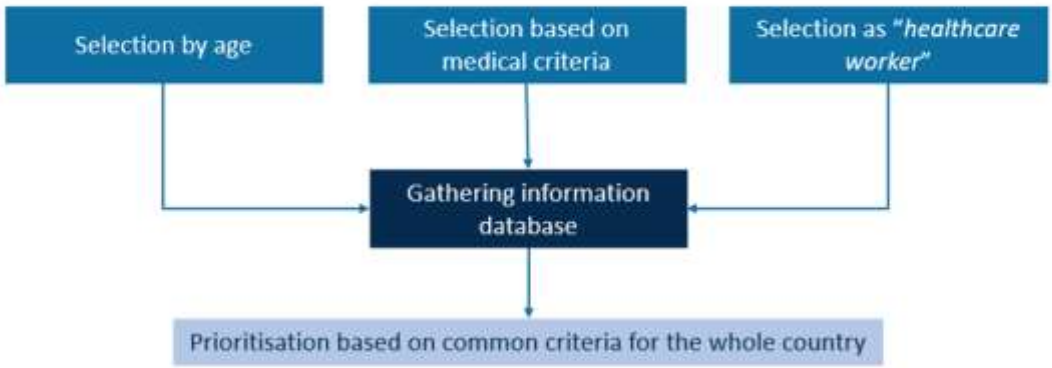

Figure 1. Example of the process of the selection of individual based on specific characteristics for invitation / notification to priority vaccination

2 Prerequisites

Prerequisites represent a context or resources that are not specific to the tool but needed for its implementation or operation. They form a general background that should exist to guarantee the correct operation of the tool, once deployed.

2.1 Assessment of prerequisites

2.1.1 Legal and ethical

Legal authorisation

Implementation of the tool must comply with the national and European legal framework. If it is not the case, a DPO must assess if the implementer benefits from a specific mandate or legal exemption allowing the deployment of the tool.

If there are no pre-existing authorisation for the use case of the data linkage, the implementer must obtain approval from a MS official decision body or patients consent regarding personal data usage.

Ethical agreement by ethics committee (information protection)

Depending on the objective and tool and country specificity, the approval and oversight from an ethics committee might be necessary. They guarantee ethical standards and principles to ensure that data reuse, processes, and goals comply with ethical standards and practices (public interest and individual rights).

Specific legal framework

Specific set of laws, regulations and/or guidelines are in place to address circumstances as sharing and processing individual health-related data, to ensure actions are compliant, ethical and secure. The legal framework should define data owners and users 'specific roles. This transparency can help foster greater citizen trust in the initiative of linking individual data.

2.1.2 Political

Practical implementation depends on access to data from multiple sources, which requires agreements and cooperative frameworks among the different health authorities and data owners. Collaboration is essential for technical and operational feasibility of the data linkage process.

2.1.3 Technical

IT infrastructure

The tool requires the IT infrastructure and network to enable the entry, collection, transfer, storage, and access of data, all this in a reliable, secured, and automated fashion.

Patient-level databases

The data linkage process does not encompass the collection of data. As such, its implementation is dependent on the pre-existence of at least two databases, using a common unique personal identifier (UPI).

Trusted third party

To protect patient interest, the UPI must be pseudonymized. A TTP with the knowledge and technology to organize the pseudonymization is thus needed.

2.2 Filling the gaps

Meeting the prerequisites is generally a long-term action that goes far beyond the scope of the implementation plans. This section presents workarounds that could help to initiate the implementation despite the lack of some prerequisites, although background effort will be required to catch up.

2.2.1 Legal and ethical

Legal and ethical authorisations, listed in "*2.1 Assessment of the prerequisites*", might be a long-term process but mandatory in the case of the usage of individual data. The implementers should obtain an approval related to data handling from the information security regulatory body .If the implementer

considers conducting research beyond basic monitoring and screening, securing an approval from the Ethics Committee is most likely necessary.

2.2.2 Political

Authorisation from the data owners is required for access, and collaboration is essential for technical and operational feasibility of the data linkage process. These authorisations define the objectives that can be addressed through access to linked data. If one of the entities does not authorise the sharing of its data, this does not prevent the data linkage from being implemented, but it does redefine the scope that can be reached by the linkage and the subsequent analysis. Depending on the content of the databases and their accessibility, different objectives can be achieved (ref. *Uses cases – 1.4.1 Vaccination surveillance*).

An overview of the different data owners in relation to the objective sought can be set up, in order to be able to find potential alternatives. Identifying the political entities in charge of health decisions, preparation, vaccination and any other subject related to the data used can be also essential to engage and support collaboration of the data owners. A decision-making committee can be set up to establish clear, shared objectives and processes, i.e. the questions that can be answered or the surveillance that can be undertaken by linking the data.

2.2.3 Technical

Unique Patient Identifier

Ideally the UPI should be a national identification number (e.g. Social Security number, National Register Number, Civil Identification Number) to enable linkage of data coming from different data holders. If all data are collected by the same organisation, creating a unique patient identifier at the organisation level is acceptable.

Privacy-Preserving Record Linkage is also an alternative if sufficient personally identifiable information is shared across databases. Although it decreases the quality of the linkage, this method allows the coupling of data without UPI and preserve privacy.

Pseudonymisation

If no TTP could be identified and all data are held by the same organisation, it is possible to set up the pseudonymisation process internally.

3 Implementing

3.1 Build

This section is the core of the implementation plan. It details how the tool is constituted, which roles should be present in the project team, the tasks they will have to perform, and a typical planning for implementation. It can be complemented with further supporting resources elaborated during the EUVABECO project.

3.1.1 Architecture

The Linkage tool is a process, more than a product or service. As such, the exact architecture will vary across the implementers to meet their own constraints and needs. However, some key elements and principles should be present in all implementations: a secure data server, a distinct technical and operational environment and a pseudonymization step (example Figure 2.).

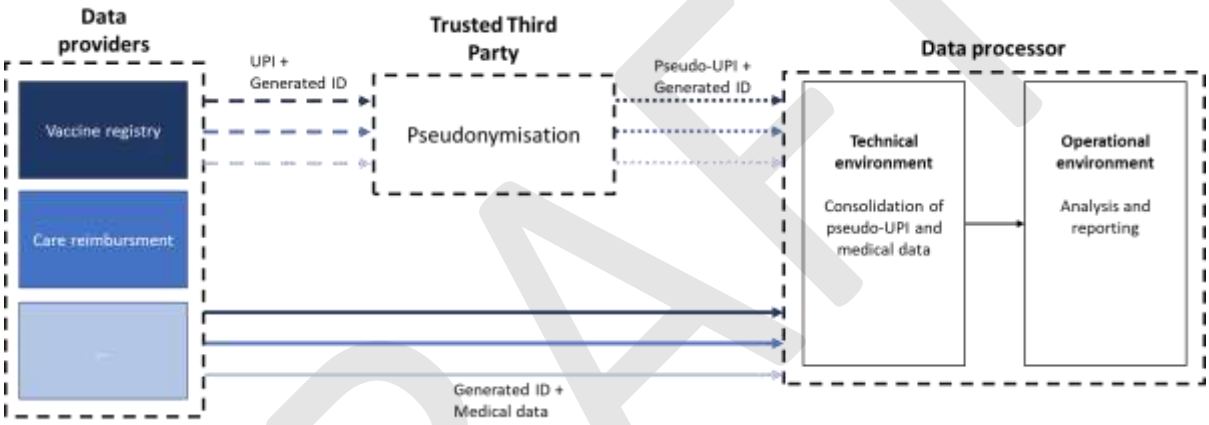

Figure 2. Example of the architecture of data linkage and dataflow associated.

Secure data server

The collected data contain medical and personal information. For privacy protection and legal compliance, they must be hosted on a secure data server. In addition, a data breach would be harmful regarding citizens trust and continuity of the health surveillance operation.

Some key features of a secure server are:

- *Encryption*: the data should be encrypted both at rest and during transit (using protocols like TLS/SSL).
- *Access control*: having a strong password policy and using multifactor authentication (MFA) helps creating a strong authentication process. This goes along with role-based access control (RBAC) to limit the number of users having access to the most sensitive information.
- *Security monitoring*: to identify and respond to potential threats, a continuous monitoring, the use of a logging system, and regular security audits such as vulnerability assessment and penetration tests.

Distinct technical and operational environment

An important feature is the separation between the technical environment and the operational environment. The former is dedicated to all technical processes such as the actual linkage, data consolidation and data validation. The latter contains only the processed, pseudonymised datasets

and is where the data analysis to monitor the situation and answer policy/research questions are conducted.

Building on the access control characteristic of a secure server, both environments should meet the same requirement in terms of authentication. A role-based access is important as well, meaning that people conducting the technical processes should be different from the ones conducting the analysis. Despite this, there is still a risk of reidentification of the patient even with pseudonymized data. This is particularly the case when multiple databases are cross-linked, since many indirect identifiers (e.g. age, zip code, profession, etc.) and unique data (e.g. rare chronic condition) are combined. If this is the case, persons working with the operational data should be asked to sign a declaration of 'good clinical practice', which stipulates the purposes of working with the data.

In addition, for the operational environment more specifically, access to the data should be limited to people mandated to monitor the situation and support policy, or having operational goals aligned with the objectives and motivations of the data linkage. These motivations and objectives, as well as the evaluation of a research project, should be discussed within a decision board (which could include representatives of the different data providers).

Also, to mitigate the reidentification risk, only aggregated data should be exported from the operational environment.

Pseudonymization

Ideally, the medical data are never shared along with the UPI or to the same party. To achieve this, the pseudonymization step is handled by a TTP and in such a way that medical data are never transferred along with the UPI.

One way to achieve this is to ask the data providers to split their data in two parts. In the first part, the UPI will be replaced by a generated ID. This dataset containing generated ID and medical data will be sent directly to the data processor. The second part contains the list of UPI with their matching generated ID. This list goes through the TTP to replace the UPI by a pseudo-UPI using deterministic encryption[1]. Once the data processor receives both part of the data, they can consolidate the message by reuniting the pseudo-UPI and the medical data based on the common generated ID. After consolidation and data validation, this generated ID is definitively deleted, and the consolidated dataset is made available in the operational environment.

Depending on the number of data providers and their relationship, the rest of the architecture, and local regulation, the pseudonymisation process can vary. It is nonetheless essential that it occurs before the data reach the research environment and are analysed.

[1] *The same pseudonymisation key is used for every data provider, ensuring that a UPI is always transformed to the same pseudo-UPI.*

3.1.2 Project team

Setting up and then exploiting the data linkage requires experts with competencies in IT, data science, health/epidemiology, as well as a data protection officer (DPO) with solid understanding of GDPR. General and comprehensive awareness about the European Health Data Space (EHDS) can help.

The first step is to draw up an agreement on the objectives of the data linkage and the way to meet those. In practice, it could mean reaching a compromise between the legal and technical constraints, and the ideal needs in terms of health surveillance. To do so, experts from each field will collaborate in an iterative process:

- Health/epidemiology experts identify the needs to be met by the linkage, inventory the data required to do so, and compare it to the existing resources.
- Legal experts and DPO ensure compliance with the legal framework and establish guidelines in terms of data protection.

- IT experts offer guidance on the technologies and resources that meet the information needs of the linkage, the protection regulations, and the security of the data.

Once the agreement is reached, the IT professional can start to set up the technical infrastructure bringing all the data together. When it is up and running, the data scientists and epidemiologists collaborate to analyse the data and derive surveillance indicators, recommendations, and scientific knowledge from it.

All parties remain in collaboration during the process to assess and answer to new use cases.

3.1.3 Workflow

- Identify the relevant dataset and ensure their interoperability (data format compatibility, data standardisation formats used, semantic interoperability);
- Once interoperability is assessed, define a detailed protocol and description for a secure and pseudonymized data transfer, which safeguards the data during transmission and maintains privacy;
- Datasets are gradually added to the data flow, starting with vaccination data, allowing for a phased integration process;
- As datasets are incorporated, indicators of data quality (i.e. plausibility, completeness compliance) and the robustness of the automated processes (e.g. % of successful transfer, % of records transmitted, comparison of aggregated values between the original database and data in the operational environment) are monitored;
- Once the data flow is established, continuous maintenance is performed to ensure smooth operation, data are made available for use;
- Effective communication of the data is maintained to ensure that relevant stakeholders are informed and engaged.

3.1.4 Typical planning

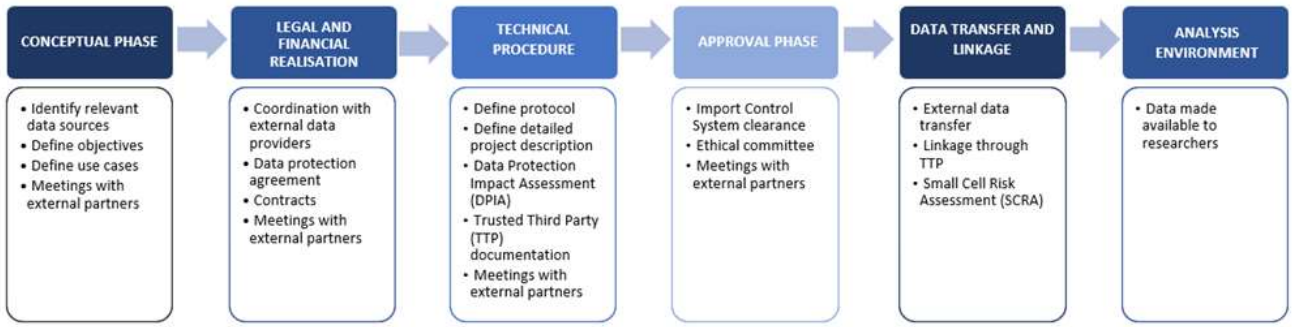

Figure 3. Steps for the implementation and use of data linkage

Small Cell Risk Assessment (SCRA) *is the process of evaluating the potential risk of re-identifying individuals within a dataset, particularly when dealing with small populations and/or high granularity. This risk arises when data is aggregated or presented in a way that small groups (or "small cells") contain very few individuals, making it easier to deduce their identities. SCRA is crucial to protect individual privacy, especially if the dataset contains sensitive information such as health or demographic data. Conducting a SCRA before making data available in an analysis environment ensures that individual privacy is preserved, sensitive data is protected, and legal obligations are met, all while maintaining the data's usefulness for operational purposes.*

Task	M1	M2	M3	M4	M5	M6
Conceptual Phase						
Identify relevant data sources						
Define objectives						
Define use cases						
Legal and Financial realisation						
Coordination with external data providers						
Data protection agreement						
Contracts						
Technical procedure						
Define protocol						
Define detailed project description						
Data Protection Impact Assessment (DPIA)						
Trusted Third Party (TTP) documentation						
Approval phase						
Ethical committee						
Data transfer and linkage						
External data transfer						
Linkage through TTP						
Small Cell Risk Assessment (SCRA)						
Analysis environment						
Data available to researcher						

3.1.5 Build resources

List of useful tools:

- Protocol/software for data transfer
- Protocol/software for pseudonymisation
- Protocol/software for data storage
- Protocol/software for operational environment access
  - Citrix Gateway
- Protocol/software for data analysis/management
  - SAS Enterprise Guide
  - R and Rstudio
- Software for data reporting
  - Shiny app
  - Power BI
  - Looker Studio

3.1.6 Verification

- **Time component:** *How often and how fast are the data transferred?*
- **Comprehensiveness:** *Do we have access to all the variables we need?*
- **Completeness:** *Do we have data about all the population?*
- **Reliability:** *How certain are we that the values in the dataset and linkage are accurate?*

3.2 Run

Once the tool has been deployed, there is still a need for lasting resources to support its adoption and ensure its maintenance. This section details these further actions.

3.2.1 Governance

Once the data linkage is set up, governance actions must be taken to ensure it remains aligned with vaccination policies and upcoming challenges. It includes:

- **Progress of ongoing activities**: progress toward the objectives set during the conception of the data linkage or added afterward requires continuous follow-up on the technical and operational level.
- **Evaluation of new research projects**: new projects or objectives related to the data linkage must be assessed to decide if they are a) technically feasible (i.e. data are available or can be obtained) and b) compliant with the legal framework and authorisations obtained.
- **Expanding data catalogue**: all the required databases will most likely not be available since the beginning. Continuous effort is needed to integrate agreed-upon data sources and update them when needed. Looking actively for new relevant databases is also encouraged to keep the tool relevant for new use cases.

3.2.2 Monitoring

To ensure optimal performance of the data linkage, two key technical aspects must be monitored:

- **Dataflow continuity**: a recurrent exchange of data (be it daily, weekly, monthly, …) is possibly foreseen for some of the databases. For those, the dataflow should be monitored to ensure that the new data are coming in or be able to quickly fix any interruption.

- **Data quality**: when a new database is added to the linkage, and periodically afterwards, the quality of the data must be assessed as it will impact the reliability of the generated outputs.

3.2.3 Communication

Two area of communication are identified for the tool. The first one is information about the data linkage itself: its process, purpose and performance, as well as the benefits and how the risks – such as privacy issues – are mitigated. The second area covers information generated thanks to the tool, such as scientific knowledge and public health insights. This type of information, and who it is communicated to, will vary depending on the use case.

Three communications targets are identified and should be reported to :

- **Health authorities**: must receive continuous and punctual reporting on knowledge impacting the public health decisions (e.g. coverage or vaccine effectiveness), as well as resources to help them in the application of public health policies (e.g. list of prioritized patients for vaccination).
- **Data owners**: must be informed of how their databases are used and, as experts on their data, be involved in how they are interpreted.
- **Citizens**: must be informed on the usage of their data and the measures taken to protect their privacy. If public health insights are generated, it also is preferable that they are made available through punctual (e.g. press release) and continuous reporting (e.g. dashboard or regular bulletin).

Additionally, depending on the use case, scientific knowledge building thanks to the data linkage should be communicated to the scientific community through e.g. conferences or journal articles.