Secondary use of IPS FHIR or CDA with Mortality Data (EHDS2)

(De-Identification Handbook Example) to be referenced as a use case example in the Volume 1 of De-Identification Profile

For Inclusion in updates to the IHE ITI De-Identification Handbook(https://build.fhir.org/ig/IHE/ITI.DeIdHandbook/branches/main/index.html)

Scenario:

The region is under threat of outbreak for a novel virus causing influenza-like-symptoms. An approved public health monitoring program wishes to review primary health information from across multiple jurisdictions for impacts of co-morbidities, mortality rates, and occupational health risks. The study also intends to review the population health impact for medication treatments, including vaccination. The purpose of the research request reflects a permitted purpose of use, Public Interest in the area of Public, Occupational Health. Data for this study represents categories from Healthcare, Medicinal products, Areas of Public Health, Areas of Occupational Health, and possibly Serious cross-border threats.

- The Data Discovery leveraging the HealthDCAT Application Profile (HealthDCAT-AP) determines that the information is available within the HDAB Information Resources, indicating data is available in a standard International Patient Summary (IPS), and mortality data in the Vital Records Death Reporting (IHE VRDR) standard formats as indicated in the conformsTo attribute (see https://healthdataeu.pages.code.europa.eu/healthdcat-ap/releases/release-5/).
- The user submits a Data Access Permit application (drawn from example at: https://data-access.dsa.ec.europa.eu/public/hta/data-access) requests data access authorization providing in the Research project-specific information section:
    - Information on the type and format of the data requested, indicates International Patient Summary format for the data, indicating the specific attributes from that standard content needed for the study and applicable de-identification methods should be used to preserve sufficient information to fulfill the study.
    - The researcher also provides a date range of access needed for a 3-year period.

    - access justifications are provided indicating the value of the proposed research
    - The purpose of use is indicated as Scientific Research
    - Data elements requested that are sensitive include de-identification methods to be used that will retain sufficient information for the research study.
- Policy for secondary use includes data minimisation for the use of secondary health data. This includes limiting the amount, type, and granularity of data during data preparation. The application for data access includes a request for the following data and de-identification methods:
    - Pseudonymized demographics with synthetic data for data of birth (age-group), and address
        - Pseudonymized name
        - Patient location is important to the study. Address locations generalized to the initial 3-digits of the postal code for data-minimization
        - Administrative gender is an important metric in the study and will be included
        - Date of birth will be used to determine and convey a generalized age-group. The Date of birth, if included, will be Synthetic Data applied by date shifting within age-group breakdown
        - Preferred language will be omitted for data-minimization
        - Patient identifier sent as a pseudonymized identifier that is applied to the same patient over time
        - Insurance information will be omitted for data-minimization
        - Given the potential for identifying health risks to come subjects of care, the data is to be reversibly pseudonymized to support tracking of patient care related to the event across health care facilities and settings.
    - Problems
        - Problems will be key to determining primary conditions, symptoms, co-morbidities, and clinical outcomes.
        - Conditions will be reviewed for potential identifiable outliers for suppression.
    - Procedures
    - Medications
        - Medications are needed to identify treatment and contraindication impact

        - Dates associated with medications are relative to incident and treatment dates. Data is collected but protected by data-shifting the study records
        - Other attributes associated with Allergies and intolerances are removed for data-minimization
    - Allergies and intolerances will support clinical outcome measures
        - Allergy Agent
        - Onset date is needed relative to incident and treatment dates. Data is collected but protected by data-shifting the study records
        - Reaction
        - Other attributes associated with Allergies and intolerances are removed for data-minimization
    - Results
        - Observation resulted is needed to inform the detection of infectious agents and clinical metrics
        - Observation value is needed for metrics
        - Observation date is needed relative to incident and treatment dates. Data is collected but protected by data-shifting the study records
        - Other attributes associated with Results are removed for data minimization
    - Immunizations
        - Vaccine for type of disease is needed to assess treatments and measure mitigation results
        - Date of immunization is needed relative to incident and treatment dates. Data is collected but protected by data-shifting the study records
        - Other attributes associated with Immunizations are removed for data minimization
    - Medical Devices are not needed for the study
        - This section may be empty with a data absent reason of 'masked'
    - Social History lifestyle factor information is needed to measure potential environmental impact including
        - Occupation – either usual occupation or current occupation is needed to identify incidents and to help identify risk factors associated with identified occupations
        - Industry – either usual industry or current industry is needed to identify incidents and to help identify risk factors associated with identified occupations

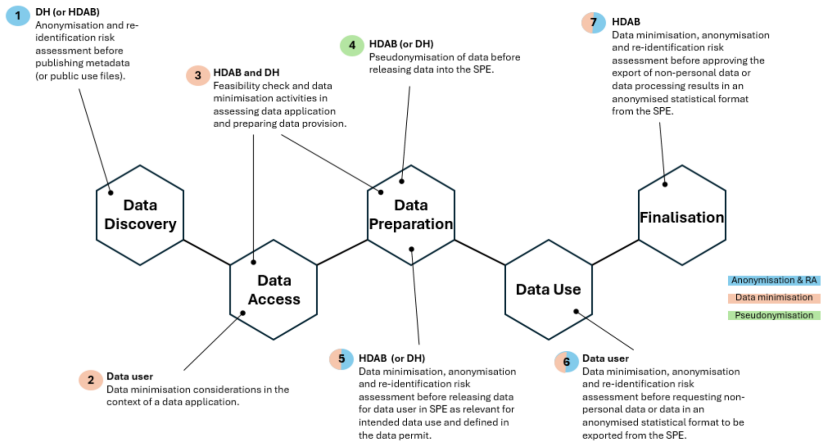
        - Occupation and Industry will be reviewed for outliers and suppressed
        - All other attributes associated with Social History are removed for data minimization
    - Pregnancy History
        - Pregnancy status – pregnancy information is needed for the study to review potential impact of the incident on pregnancy
        - Estimated Delivery date - is needed relative to incident and treatment dates. Data is collected but protected by data-shifting the study records
        - Other attributes associated with Pregnancy History are removed for data minimization
    - Mortality data is available through vital records offices. The mortality data will include:
        - Pseudonymized name
        - Pseudonymized identifier
        - Date of death is needed relative to incident and treatment dates. Data is collected but protected by data-shifting the study records
        - Cause of death
        - All other mortality data is omitted for data-minimization
- The application is reviewed and a Data Access permit is granted
- The Health Data Access Body (HDAB) uses an Intermediation entity for pseudonymizaton, and data preparation to prepare the data using approved anonymization, pseudonymization, generalization, suppression, and randomization methods according to the permit.
- A second pass removes potentially identifying outlier data from the dataset for privacy protection. While the identification of outliers is not required by the EHDS regulations, the HDAB performs this analysis to optimize privacy protection before releasing the data set.

Data flows section

Before the de-identification processing, EHDS 2 regulations specify that the Health Data User conducts data discovery to identify data that is available for the study, and that an application for a data permit be submitted and approved.

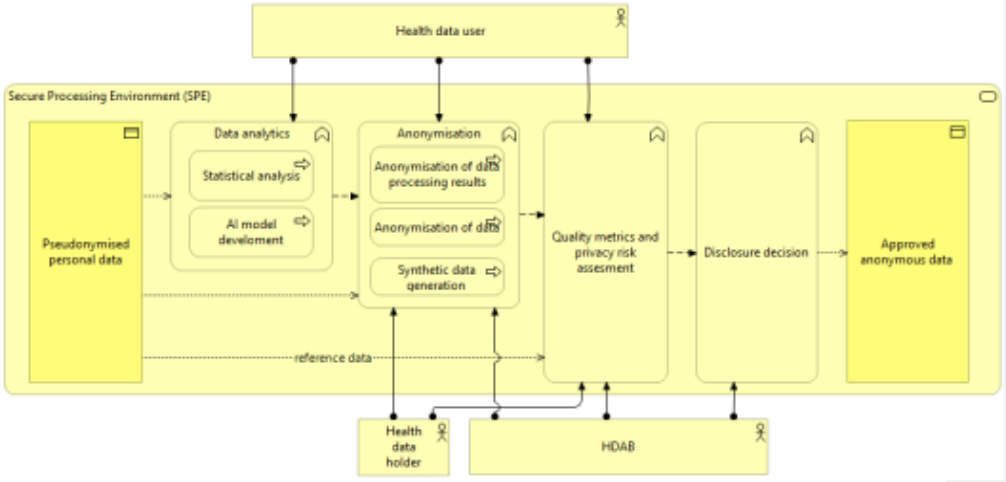
The following diagram from the EHDS2 M7.2 Draft guideline on data minimisation, pseudonymisation, anonymisation and synthetic data depicts he EHDS user journey for

the full process, highlighting the areas within the process that address data minimization, pseudonymization and anonymization:



Abbreviation: DH: Data holder; HDAB: Health Data Access Body; RA: re-identification risk assessment; SPE: secure processing environment.

Once a Data Permit is granted, the third phase, Data Preparation, begins. The HDAB prepares the data set according to the permit content and de-identification to be applied. The EHDS2 M7.2 Draft guideline on data minimisation, pseudonymisation, anonymisation and synthetic data depicts a High-level architecture for safe disclosure of anonymized data, processing results, and synthetic data:



Requirements

Policy

- The HDAB data access application management process is primarily set out in Articles 67–69 of the EHDS regulation, which define the procedural context that any organisational or technical solution for secondary use data applications must align with. Articles 67 and 69 includes the requirements for the common application forms for data access applications and data requests, respectively, to be used by applicants, and which provides the essential information for the processing of applications. Article 68 governs the issuance of data permits and the associated obligations of HDABs, whereas Article 69 provides similar provisions for data requests.
- The secondary use of electronic health data is based on pseudonymised or anonymised data, in order to preclude the identification of the data subjects (Recital 53 in the EHDS regulation)
- Data anonymisation, pseudonymisation, and linkage techniques are addressed in M7.2 Technical specification for Health Data Access Bodies on data minimisation and de-identification, and M7.5 Guideline for Health Data Access Bodies on linkage of health datasets.

Permitted Uses

The Health Information Exchange system has defined the following standard permitted uses (Art 53 a-c). The standard healthcare purposes of use specified by ISO TC215 14265: Health Informatics - Classification of purposes for processing personal health information concepts associated with these EHDS2 defined purposes are provided in italicized sub-bullets:

- Improving the delivery of care, Treatment, Optimization and providing healthcare
    - *Treatment*
        - *Clinical Care Provision to an Individual Subject of Care*
        - *Emergency Care Provision to an Individual Subject of Care*
        - *Support of Care Activities within the Provider Organization for an Individual Subject of Care*
        - *Subject of Care Uses*
    - *Operations*
    - *Health Service Management and Quality Assurance*

- Public Interest in the area of Public, Occupational Health and Policy Making and Regulatory Activities, Statistics, national, multinational, and Union level official statistics
    - *Public Health*
        - *Public Health Surveillance, Disease Control*
        - *Population Health Management*
        - *Public Safety Emergency*
- Scientific Research contributing to public health or health technology assessment with the aim of benefitting the end users
    - *Research*
- Vocational and Higher Education Teaching Activities
    - *Education*

Uses NOT Permitted

The Health Information Exchange system has define the following standard are not permitted uses:

- carrying out advertising or marketing activities;
    - *Market Studies*

No corresponding purpose of use is defined by EHDS2 as either permitted or not permitted, so the assumption is the following ISO-defined standard classifications of use purposes are not permitted uses:

- *Legal Procedure*
- *Enabling the payment of care provision to an individual subject of care*

Commented [LR1]: Not referenced

Additional regulation purposes from Article 54 not permitted:

- Taking decisions detrimental to a natural person or a group of natural persons based on their electronic health data; in order to qualify as 'decisions' for the purposes of this point, they have to produce legal, social or economic effects or similarly significantly affect those natural persons;
- developing products or services that may harm individuals, public health or society at large, such as illicit drugs, alcoholic beverages, tobacco and nicotine products, weaponry or products or services which are designed or modified in such a way that they create addiction, contravene public order or cause a risk for human health;
- carrying out activities in conflict with ethical provisions laid down in national law.

Risk Assessment

Risk and associated mitigation is determined for each data element associated with a given Data access request. For this test case, see the data elements described in the next section.

Data Elements

Data Types: The IPS format requested contains primarily structured data with some attributes containing textual data content. There are no Medical imaging data, Bio-signal data, Genetic data, Textual data, or Multi-modal data. Available to this research study through the IPS structured format.

IPS Section				
IPS Patient				
	Patient Name	Structured Data	Direct Identifier	Reversibly pseudonymized name
	ID	Structured Data	Direct Identifier	Reversibly pseudonymized Identifier
	Date of Birth	Structured Data	Quasi-identifier	Synthetic Data applied by date shifting within age-group breakdown
	Gender	Structured Data	Quasi-identifier	Administrative gender is an important metric in the study and will be included
	Telecom	Structured Data	Direct Identifier	Omitted for data minimization
	deceased indicator	Structured Data	Quasi-identifier	Included
	deceased date	Structured Data	Quasi-identifier	Synthetic Data applied by date shifting relative to shifted date of birth
	Patient address	Structured Data	Quasi-identifier	Address locations generalized to the initial 3-digits of the postal code for data-minimization
	Preferred language	Structured Data	Subject to data minimization. Not requested by the study	Preferred language will be omitted for data-minimization
	General Practitioner	Structured Data	Quasi-identifier	Omitted for data minimization
	Insurance	Structured Data	Quasi-identifier	Omitted for data minimization
Problems				
	Problem Type	Structured Data		

	Description	Textual Data	Quasi-identifier	Omitted due to the possibility of free text privacy issues
	Diagnosis	Structured Data	Quasi-identifier	Diagnoses will be reviewed for potential identifiable outliers for suppression.
	Severity	Structured Data	non-identifying	unchanged
	Onset Date	Structured Data	Quasi-identifier subject to date shifting relative to shifted birth date	
	Problem Status	Structured Data	non-identifying	
	Specialist Contact	Structured Data	Quasi-ientifier	Omitted for data minimization
History of Procedures				
	Procedure code	Structured Data	Quasi-identifier	Procedures will be reviewed for potential identifiable outliers for suppression.
	Procedure description	Textual Data	Omitted due to the possibility of free text privacy issues	
	Body site	Structured Data	non-identifying	Diagnoses will be reviewed for potential identifiable outliers for suppression.
	Procedure date	Structured Data	Quasi-identifier subject to date shifting relative to shifted birth date	
Medication Summary				
	Product Code	Structured Data	non-identifying	Product code will be reviewed for potential identifiable outliers for suppression.
	Product Common Name and strength	Textual Data	non-identifying	Unchanged as required if known, and the coded product code is not required
	Active ingredient substance code	Structured Data	non-identifying	unchanged
	Active ingredient substance strength	Structured Data	non-identifying	unchanged
	Period of medication use	Structured Data	Quasi-identifier subject to date shifting relative to shifted birth date	
	Route of admininstration	Structured Data	non-identifying	Removed for data minimization (not requested)
	Dose quantity	Structured Data	non-identifying	unchanged
	Dose frequency	Structured Data	non-identifying	unchanged
Allergies	Allergy and intollerance description	Textual Data	non-identifying	Removed for data minimization (not requested)

	Clinical status	Structured Data	non-identifying	unchanged
	Onset date	Structured Data	Quasi-identifier	subject to date shifting relative to shifted birth date
	End date	Structured Data	Quasi-identifier	subject to date shifting relative to shifted birth date
	Criticality	Structured Data	non-identifying	unchanged
	Certainty	Structured Data	non-identifying	unchanged
	Type of propensity	Structured Data	non-identifying	unchanged
	Diagnosis	Structured Data	Quasi-identifier for outliers	Omitted for data minimization
	Reaction Manifestation	Structured Data	non-identifying	unchanged
	Reaction Severity	Structured Data	non-identifying	unchanged
	Agent Code	Structured Data	non-identifying	unchanged
	Agent Category	Structured Data	non-identifying	unchanged
Results				
	Date of observation	Structured Data	Quasi-identifier	subject to date shifting relative to shifted birth date
	Observation type	Structured Data	non-identifying	unchanged
	Result description	Textual Data	non-identifying	Removed for data minimization (not requested)
	Result value	Structured Data	non-identifying	unchanged
	Observation resul	Structured Data	non-identifying	unchanged
	Performer	Structured Data	non-identifying	Omitted for data minimization
	Observer	Structured Data	non-identifying	Omitted for data minimization
Immunizations	Vaccine for type of disease	Structured Data	non-identifying	unchanged
	Date of immunization	Structured Data	Quasi-identifier	subject to date shifting relative to shifted birth date
	Number in series of doses	Structured Data	non-identifying	unchanged
	Target disease	Structured Data	non-identifying	Omitted for data minimization
	Product name	Textual Data	non-identifying	Omitted for data minimization

	Vaccine/Prophylaxis	Textual Data	non-identifying	Omitted for data minimization
	Product administration performer	Structured Data	non-identifying	Omitted for data minimization
	Route of admininstration	Structured Data	non-identifying	Omitted for data minimization
Social History			Quasi-identifier	Occupations will be
	Lifestyle factor information (occupation)	Structured Data	subject to date shifting relative to shifted birth date	reviewed for potential identifiable outliers for suppression.
	Lifestyle factor information (industry)	Structured Data	Quasi-identifier subject to date shifting relative to shifted birth date	
Pregnancy History	Pregnancy Status	Structured Data	non-identifying	unchanged
	Estimated Delivery Date	Structured Data	Quasi-identifier subject to date shifting relative to shifted birth date	
Medical Devices	No medical device data is needed for the study	Structured Data	non-identifying	Omitted for data minimization. Data absent reason marked as 'masked'

All unspecifed data that may be available in the IPS are removed for data-minimization

VRDR content

Mortality data	Name	Structured Data	Direct Identifier	Reversibly pseudonymized name
	Identifier	Structured Data	Direct Identifier	Reversibly pseudonymized Identifier
	Date of death	Structured Data	Quasi-identifier	Synthetic Data applied by date shifting within age-group breakdown
	Cause of death	Structured Data	Quasi-identifier	Cause of Death will be reviewed for potential identifiable outliers for suppression.

Add Case example pre and post de-identification

Commented [LR2]: Add example detail after use case review