New concept proposal

Source System, Healthcare Primary Information System

Author	**Kristin Gnodtke, Harald Witte**	**Date last updated**	**13.11.2023**
Project	General interest	**Contact**	DCC
Dataset release	2024.1	**Consulted expert**	Katie Kalt

1 Rationale

Following the FAIR principles (Wilkinson et al., 2016, DOI: 10.1038/sdata.2016.18) maximizes the value of research data by making it findable, accessible, interoperable, and reusable. This includes providing information on data origin, source data, and source system to researchers as metadata along with the actual dataset. Source system information can include details where data originates from, e.g., a radiology information system, a cohort or a registry, or a clinical report form. Additionally, it can also be provided on the concept level, e.g., where data of measurements or assessment such as Body Surface Area or Body Height is taken from.
The proposed concept Source System describes the type and purpose of the source system, e.g., a hospital information system for clinical routine data (Figure 1). Source System also holds an attribute Healthcare Primary Information System which describes the primary source system of the healthcare data, e.g., a clinical laboratory information system (Figure 1).
Providing this additional information makes data FAIR, reliable, and suitable for reuse by third parties.

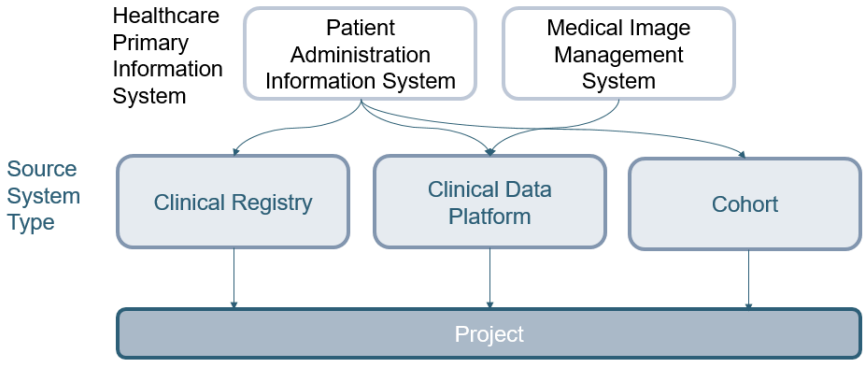

Figure 1: *Relationship between Source System, Healthcare Primary Information System, and data delivery to a project.*

2 Comparison to other standards/data models

2.1 OMOP (Observational Medical Outcomes Partnership)

OMOP features various terms related to data provenance spread across several tables, e.g., OBSERVATION_PERIOD, VISIT_DETAIL, VISIT_OCCURRENCE, or PERSON (https://ohdsi.github.io/CommonDataModel/cdm54.html#Clinical_Data_Tables).

2.2 SNOMED CT

SNOMED CT does not feature the terms source systems, source data, or data provenance. It does offer, however, a list of information systems (SNOMED CT: 706593004 | Information system (physical object) |) which covers potential clinical primary source systems ranging from anesthesia information systems through medical image management systems to radiology information systems.

2.3 FHIR

FHIR mentions source systems and data provenance information in the resource "Provenance": "Provenance of a resource is a record that describes entities and processes involved in producing and delivering or otherwise influencing that resource. Provenance provides a critical foundation for assessing authenticity, enabling trust, and allowing reproducibility. Provenance assertions are a form of contextual metadata and can themselves become important records with their own provenance. Provenance statement indicates clinical significance in terms of confidence in authenticity, reliability, and trustworthiness, integrity, and stage in lifecycle (e.g. Document Completion - has the artifact been legally authenticated), all of which may impact security, privacy, and trust policies."
Scope and usage of the provenance information is defined as follows: "The Provenance resource tracks information about the activity that created, revised, deleted, or signed a version of a resource, describing the entities and agents involved. This information can be used to form assessments about its quality, reliability, trustworthiness, or to provide pointers for where to go to further investigate the origins of the resource and the information in it."

2.4 NCIT (National Cancer Institute Thesaurus)

NCIT features the term "Provenance" with the definition "Where something originated or was nurtured in its early existence." (NCI Thesaurus Code: C43581). There are no terms directly describing source system or source data. NCIT also defines, however, terms for several types of data sources, e.g., "Cohort" (Cohort NCI Thesaurus Code: C61512); "A group of individuals, identified by a common characteristic."), "Registry" (NCI Thesaurus Code: C61393; "An official written record of names or events or transactions."), and various information systems, e.g., "Laboratory Information Management System (NCI Thesaurus Code C48209; "Integrated computer systems that are used in the laboratory for the management of samples, laboratory users, instruments, standards and other laboratory functions such as invoicing, plate management, and work flow automation.")", "Picture Archiving and Communication System" (NCI Thesaurus Code: C17624; "Picture Archiving and Communication System: A network established to provide access, storage and retrieval of images from medical imaging instrumentation."), or "Health Information System" (NCI Thesaurus Code: C18229).

3 Concept information

3.1 Source System

Concept or concept compositions or inherited	General concept name	General description	Contextualized concept name	Contextualized description	Type	Standard	Value set or subset	Meaning binding	Cardinality for composedOf
Concept	Source System	electronic system the data has been retrieved from	Source System	electronic system the data has been retrieved from					
composedOf	category	category associated to the concept	source system category	category of the source system from which the data is extracted for the recipient	qualitative		Biobank; Case Report Form; Clinical Data Platform; Clinical Registry; Cohort; Data Repository; Healthcare Information System; OMICS Facility; Research Laboratory		0:1
composedOf	name	name associated to the concept	source system description	source system description expressed using free text where a coded category is not applicable	string				0:1

composedOf	purpose	objective of the concept	purpose	purpose the source system has been set up for	qualitative		Billing; Patient Care; Quality Control; Research		0:1
composedOf	primary information system	primary system providing input to the concept	healthcare primary information system	primary system providing input to the source system	Healthcare Primary Information System				0:1

Concept cardinalities:

General concept name	cardinality for concept to Administrative Case	cardinality for concept to Data Provider	cardinality for concept to Subject Pseudo Identifier	cardinality for concept to Source System
Source System		1:1		

Leave field empty if not applicable.

3.2 Healthcare Primary Information System

Concept or concept compositions or inherited	General concept name	General description	Contextualized concept name	Contextualized description	Type	Standard	Value set or subset	Meaning binding	Cardinality for composedOf
Concept	Healthcare Primary Information System	primary source system of the healthcare data	Healthcare Primary Information System	primary source system of the healthcare data				SNOMED CT: 706593004 | Information	

								system (physical object) |	
composedOf	code	coded information specifying the concept	code	coded information specifying the category of healthcare primary information system		SNOMED CT	descendant of: 706593004 |Information system (physical object)|		0:1
composedOf	name	name associated to the concept	primary system description	healthcare primary information system description expressed using free text where a coded category is not applicable	string				0:1

Concept cardinalities:

General concept name	cardinality for concept to Administrative Case	cardinality for concept to Data Provider	cardinality for concept to Subject Pseudo Identifier	cardinality for concept to Source System
Healthcare Primary Information System		0:1		

Leave field empty if not applicable.

4 Impact on the SPHN Dataset

The concept Source System will be linked to most data concepts in the SPHN Dataset on the concept level, akin to the current three core concepts Administrative Case, Data Provider, and Subject Pseudo Identifier (SPHN ontology release 2023.2).
This primarily concerns concepts with a relationship to Subject Pseudo Identifier (except for example Code or Body Site). In addition, concepts are also linked to Source System if information on it can be of specific interest, including the concepts Care Handling, Drug, Drug Article, Interpretation, Reference Interpretation, Substance, and Therapeutic Area (release 2024.1).

5 Discussion

It has been discussed whether to link the Source System to data concepts on composedOf level or concept level. It has been decided to start on the concept level (similar to OMOP), as a tradeoff between amount of effort and benefit. In the future it needs to be assessed how frequently different composedOfs of one concept actually originate from different source systems.
A direct link between the Source Data concept and Source System had initially been discussed but ruled out to avoid duplicating information as the Source System is linked to most data concepts in the SPHN Dataset on the concept level.
After an evaluation using comprehensive data from CDWHs, however, the link of Source Data to Source System has been introduced with a cardinality of 1:n. The lack of a direct link may cause problems in the context of other concepts, e.g., Semantic Mapping. When data originates from different source systems (e.g., KISIM and PDMS), codes of the Source Data are different but the output code of a semantic mapping is identical. This renders the original code used in the source data non-traceable.
The cardinality 1:n for the link between single concepts and the Source System concept reflects that data elements represented by one concept can come from different source systems. For example, multiple source systems of Diagnosis imply that DiagnosisCode may be from source system A and DiagnosisSubjectAge from source system B.

Instead of “Healthcare Primary Information System”, the concept name “Hospital Primary System” has initially been considered. However, the broader term “Healthcare Primary Information System” has been chosen as it also applies under circumstances where the data provider delivers clinical data yet is not a hospital, e.g., in case of data delivered by a private practice.

The value set for the composedOf 'category' of Source System is represented well in other clinical data models or systems (see Table 1).

Table 1: *Representation of values for Source System-category in other clinical data models or systems*

term in value set	code (approximate)	code label	URL
Bio Bank	EDAM:3337	Biobank	http://edamontology.org/topic_3337
Bio Bank	NCIT:C48800	Biorepository	https://ncit.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=C48800
Case Report Form	32809	case report form	https://athena.ohdsi.org/search-terms/terms/32809
Case Report Form	OMOP4976882	case report form	https://athena.ohdsi.org/search-terms/terms/32809
Case Report Form	NCIT:C40988	Case Report Form	https://ncit.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=C40988
Clinical Data Platform	*[No specific code available.* ***Intended as an overarching term for data warehouses, data lakes etc. in a clinical context.****]*		
Clinical Registry	32879	Registry	https://athena.ohdsi.org/search-terms/terms/32879
Clinical Registry	OMOP4976952	Registry	https://athena.ohdsi.org/search-terms/terms/32879
Clinical Registry	LOINC:74735-2	Health data repository [Identifier]	https://loinc.rg/74735-2/
Cohort	44819246		https://athena.ohdsi.org/search-terms/terms/44819246
Cohort	NCIT:C61512	Cohort	https://ncit.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=C61512
Cohort	ExO:0000120	cohort	http://purl.obolibrary.org/obo/ExO_0000120
Data Repository	LOINC:74735-2	Health data repository [Identifier]	https://loinc.rg/74735-2/
Data Repository	NCIT:C114457	Digital Data Repository	https://ncit.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=C114457
Healthcare Information System	SNOMED CT:706593004 and descendants thereof	Information system (physical object)	https://browser.ihtsdotools.org/?perspective=full&conceptId1=706593004&edition=MAIN/2023-11-01&release=&languages=en
Healthcare Information System	OMIT:0008383	Information Systems	http://purl.obolibrary.org/obo/OMIT_0008383
Healthcare Information System	OMIT:0028178	Health Information Systems	http://purl.obolibrary.org/obo/OMIT_0028178
Healthcare Information System	OMIT:0007894	Hospital Information Systems	http://purl.obolibrary.org/obo/OMIT_0007894
OMICS Facility	ENVO:01001475	omics observatory system	http://purl.obolibrary.org/obo/ENVO_01001475
OMICS Facility	OBI:0001891	sequencing facility organization	http://purl.obolibrary.org/obo/OBI_0001891
Research Laboratory	NCIT:C37984	Laboratory	https://ncit.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=C37984
Research Laboratory	SNOMED CT:261904005	Laboratory (environment)	https://browser.ihtsdotools.org/?perspective=full&conceptId1=261904005&edition=MAIN/SNOMEDCT-CH/2023-06-07&release=&languages=en,fr,de,it

The composedOf 'purpose' of Source System with the value set {Billing; Patient Care; Quality Control; Research} is of general interest but can be particularly beneficial in the case of de-identified data and potential re-use by third parties. There, details on data provider or primary system may have been removed (https://sphn.ch/network/data-coordination-center/de-identification/). The composedOf "purpose", however, will allow an estimation of the data quality standards, accreditation level etc. in such cases.

Moreover, it will provide information on the objective or reason for setting up the source system. Table 2 below provides example situations and the corresponding value. The value set is aligned between the concepts Semantic Mapping and Source System.

Table 2: Usage examples for 'purpose' of 'Source System':

Example situation the source system is primarily in place for	**Corresponding value for purpose**
billing purposes, for example software used for assigning ICD-10- or CHOP-codes	Billing
diagnostic, screening, or therapeutic purposes, e.g., a radiology information system or an oncology information system	Patient Care
quality control purposes, e.g., holding data for a registry	Quality Control
a research project, e.g., capturing data for a clinical trial	Research

The composedOf 'category' of Source System will provide information on the type of source system from which the data is extracted for the recipient. Table 3 below provides example situations and the corresponding value.

Table 3: Usage examples for 'category' of 'Source System'

Example situation for data delivery to the project	**Corresponding value for 'category'**	**Definition (Chat GPT)**
directly from the database or information system of a biobank	Biobank	Biobank: Facility or institution with a focus on the collection, storage, and distribution of biological samples and associated data.
directly from a data capture system, e.g., a REDCap database, set up by the project	Case Report Form	Case Report Form (CRF): A standardized document used in clinical research to collect data on each participant in a study.
from the clinical data platform, e.g., a clinical data warehouse, a clinical data lake, etc. *"clinical data platform" is used as an overarching term for data warehouses, data lakes etc. in a clinical context.*	Clinical Data Platform	Clinical Data Platform: An integrated system for collecting, managing, and analyzing clinical data, often used in healthcare settings.

directly from a clinical registry database, e.g., a rare disease registry database	Clinical Registry	Clinical Registry: A database of information about individuals with a specific condition or disease, typically used for research and monitoring.
directly from a cohort database, e.g., the Swiss Spinal Cord Injury Cohort Study (SwiSCI)	Cohort	Cohort: A group of individuals with shared characteristics, often studied over time to understand health outcomes.
directly from a data repository, e.g., data available from the Swiss Federal Office of Statistics	Data Repository	Data Repository: A centralized location for storing and managing diverse datasets, including health data.
directly from a clinical information system, for example: - if a hospital does not have a clinical data platform and data is delivered directly from the clinical information system to the recipient - if an oncology information system is not yet connected to the clinical data platform and therefore data is delivered directly from the oncology information system to the recipient - if a private practice without clinical data platform delivers data	Healthcare Information System	Healthcare Information System: An integrated information system designed for managing healthcare data within a hospital or a healthcare organization or provider
directly from an OMICS facility information system, e.g., a genome center (example? like the Health 2030 Genome Center)	OMICS Facility	OMICS Facility: Facility or institution dealing with high-throughput technologies like genomics, proteomics, etc., generating large-scale molecular data.
directly from a research information system (non-clinical)	Research Laboratory	Research Laboratory: Physical space equipped for conducting scientific experiments and analyses, generating various types of research data.

The composedOf ‘code’ of the concept ‘Healthcare Primary Information System’ features the descendants of SNOMED CT: 706593004 |Information system (physical object)| as value set . This offers coded specifications for a broad variety of primary source systems ranging from anesthesia information systems through medical image management systems to radiology information systems. Nevertheless, several relevant codes are missing, for example “Intensive Care Information System” or “Oncology Information System”. Further candidate terms are currently collected (Nov. 2023) in collaboration with the Swiss university hospitals and shall be requested from SNOMED CT. They should be added as additional descendants of code 706593004 |Information system (physical object)|. Such an extension would not require any updates of the concept as the parent code (706593004) referred to in the value set remains the same.

6 Examples

6.1 Example 1

Source System

category: Registry
name: National Cancer Registry (NCR)
purpose: Quality Control
healthcare primary information system:
code:
identifier: 467455004
name: Clinical laboratory information system (physical object)
coding system and version: SNOMED-CT-2023-07-31
name: Onkostar

6.2 Example 2

Source System

category: Clinical Data Platform
name: -
purpose: Patient Care
healthcare primary information system:
code:
identifier: 463055006
name: Patient-administration information system (physical object)
coding system and version: SNOMED-CT-2023-07-31
name: -

6.3 Example 3

Source System

category: Healthcare Information System
name: -
purpose: Patient Care
healthcare primary information system:
code:
identifier: 463498001
name: Physician-practice-management information system (physical object)
coding system and version: SNOMED-CT-2023-07-31
name: -