

# Change request / New concept proposal

## Data Provider Institute / Data Provider

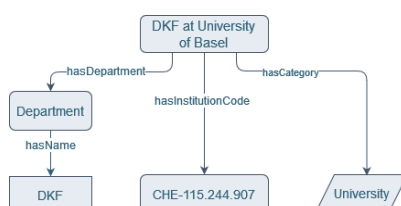
<b>Author</b>	<b>Kristin Gnodtke, Harald Witte</b>	<b>Date last updated</b>	<b>16.11.2023</b>
<b>Project</b>	General interest	<b>Contact</b>	DCC
<b>Dataset release</b>	2024.1	<b>Consulted expert</b>	Katie Kalt

### 1 Change request input / rationale

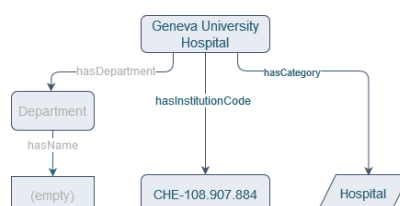
Various stakeholders have expressed interest in collecting and sharing the provenance of data. Information about data provenance i.e., meta-data that explains how the data were generated and where they came from, gives an indication about data quality and allows the researcher to assess whether the data are suitable for his or her type of data science (fit for purpose).

The current concept of *Data Provider Institute* represents information only at the general level of an organization, which could be a hospital or a university, for example. The newly proposed concepts of *Data Provider* (change request from Data Provider Institute) and *Department* (new concept proposal) offer additional details by adding more fine-grained information, e.g., on departments and the category of the providing organization. This additional detailed information is relevant if an organization has more than one single point of exit for data, or if the category of data origin (e.g., from a research department vs. routine clinical data) is of interest for liability reasons. The two new concepts will help to increase data FAIRness by enriching provenance meta-data. The following three examples illustrate the required information capability.

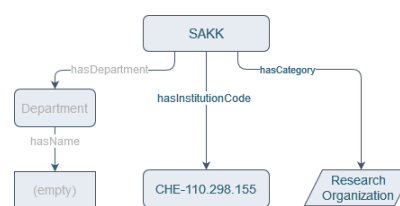
**Example 1:** The department of clinical research (DKF) at the University of Basel acts as a Data Provider.



**Example 2:** The University Hospital of Geneva acts as a Data Provider.



**Example 3:** The Swiss Group of Clinical Cancer Research (SAKK) acts as a Data Provider.



## 2 Comparison to other standards/data models

There is no exact match to Data Provider Institute or Data Provider in other common models and standards.

### 2.1 NCI

The NCI term “[Institution](#)” (C41206) describes “An established society, corporation, foundation or other organization founded and united for a specific purpose, e.g. for health-related research; also used to refer to a building or buildings occupied or used by such organization.” The mention of health-related research implies that data is collected and thus also possibly shared, i.e., provided. This corresponds to the intention behind the SPHN concept Data Provider Institute.

### 2.2 FHIR

FHIR uses the term “[Organization](#)” to describe “A formally or informally recognized grouping of people or organizations formed for the purpose of achieving some form of collective action.”. This term can be used in the scope of healthcare and research but is not a fully corresponding match to Data Provider or Data Provider Institute due to its rather general nature.

### 3 Concept information

#### 3.1 Currently released concept

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	Data Provider Institute	institute providing the data	Data Provider	institute providing the data					
composedOf	code	code, name, coding system and version describing the concept	code	code, name, coding system and version assigned to the data provider institute	Code	UID			1:1

#### 3.2 Proposed new concept

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 composed Of
Concept	Data Provider	provider of data of interest	Data Provider	provider of data of interest					

composedOf	institution code	coded information specifying the institution of the concept	institution	coded information specifying the institution of the data provider	Code	UID			1:1
composedOf	department	department of the concept	department	department, division, or subunit of an organization	Department				0:1
composedOf	category	category associated to the concept	data provider category	category of the data provider	qualitative		Company; External Laboratory; Federal Office; Health Insurance; Hospital; Pharmacy; Private Practice; Research Organization; Service Provider; University		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
Data Provider				

Leave field empty if not applicable.

## 4 Pros and cons

### 4.1 Advantages

The new concept Data Provider offers additional metadata compared to Data Provider Institute in the context of data provenance. It provides more details on data origin and thereby increases the FAIRness of data.

### 4.2 Disadvantages

Data Provider Institute is one of the three core concepts of the SPHN ontology (current release: 2023.2), along with Administrative Case and Subject Pseudo Identifier. Changing such SPHN core concepts may imply the need for significant changes to existing data pipelines.

## 5 Impact on the SPHN Dataset

Data Provider is supposed to replace the existing core concept Data Provider Institute.

## 6 Discussion

A change to the concept Data Provider Institute represents a significant change of the overall architecture of the SPHN schema. However, the benefits will increase the data quality (enriched metadata, increased data FAIRness) and outweigh the effort required.

The composedOf “category” with the value set {Company; External Laboratory; Federal Office; Health Insurance; Hospital; Pharmacy; Private Practice; Research Organization; Service Provider; University} 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 the hospital primary system may have been removed (<https://sphn.ch/network/data-coordination-center/de-identification/>). The composedOf “category”, however, will allow an estimation of the type of data, data quality standards, accreditation level etc. in such cases. It was initially considered to propose a new concept Institution featuring a code of standard UID and a composedOf department of type [Department](#). Institution would be extendible and was meant to inherit to Data Provider. Institution would therefore be superordinate to Data Provider (Institute), one of the three core concepts of the SPHN schema. It may be difficult to implement this hierarchy in the SPHN Connector, therefore this idea was put on hold for the moment.

## 7 Examples

### 7.1 Example 1

#### Data Provider

institution code:  
 identifier: CHE-115.173.213  
 name: University of Zurich  
 coding system and version: UID  
 department:  
 name: UZH Microbiology Lab  
 category: University

### 7.2 Example 2

#### Data Provider

institution code:  
 identifier: CHE-115.173.213  
 name: University of Zurich  
 coding system and version: UID  
 department:  
 name: Sequencing Lab  
 category: University

### 7.3 Example 3

#### Data Provider

institution code:  
 identifier: CHE-110.298.155  
 name: Schweizerische Arbeitsgemeinschaft für Klinische Krebsforschung SAKK  
 coding system and version: UID  
 department: -  
 category: Research Organization

### 7.4 Example 4

#### Data Provider

institution code:  
 identifier: CHE-115.250.351  
 name: Universität Bern  
 coding system and version: UID  
 department:  
 name: ISPM  
 category: University