|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | FG-AI4H-A-025 | |
| **ITU-T Focus Group on AI for Health** | |
| **Original: English** | |
| **WG(s):** | | N/A | Geneva, 26-27 September 2018 | |
| **DOCUMENT** | | | | |
| **Source:** | | Ada Health GmbH | | |
| **Title:** | | Defining the medical data space - an approach driven by data type | | |
| **Purpose:** | | Discussion | | |
| **Contact:** | | Andreas Kühn Ada Health GmbH Germany | | Tel: +49 30 60031987 Email: [andreas.kuehn@ada.com](mailto:andreas.kuehn@ada.com) |
| **Contact:** | | Henry Hoffmann Ada Health GmbH Germany | | Tel: +49 177 6612889 Email: [henry.hoffmann@ada.com](mailto:henry.hoffmann@ada.com) |

|  |  |
| --- | --- |
| **Keywords:** | AI4H; medical data space; artificial intelligence (AI); health |
| **Abstract:** | This document shows an approach to classify the expected medical data, the Focus Group will face and outlines how AI for Health solutions can be formally described as “data transformators” within the medical system |

## General Idea

The Medical Data itself is various as it is a technical approach to structure a non-human built system. So Medical information can always only cover parts of the actual underlying biology of medicine and the complex dependencies of the clinical process. In this brief document, important types of medical data are shortly described and their interaction is discussed briefly. The second part describes how this formalism helps to design the test frameworks for AI4H and gives some brief examples how AI4H solution can be covered by the created formalism.

## 1. The formalization of medical data

**Figure 1** shows general classification possibilities of data. The named entities are further described here. The Entities are categorized in 3 main categories according to their processing step.

1st category (not interpreted data, free text data):

* **Imaging Data**  
  This data type summarizes all data which includes picture material with its meta-data. Namely in this type is included: Histopathology data, MRI and CT data without interpretation etc.
* **Laboratory Results**  
  This data type may include uninterpreted genetic data, blood counts, antibody titer and similar data
* **Time series data**  
  This type includes data which has it’s information only based on the time-development as ECG, EEG or (in parts) fitness tracking functions as pulse rate
* **Patient Records**This data is to be seen as “uninterpreted” as it needs a transformation to be digitally useful. Therefor it is classified in this space

2nd category (semantically explicit data)

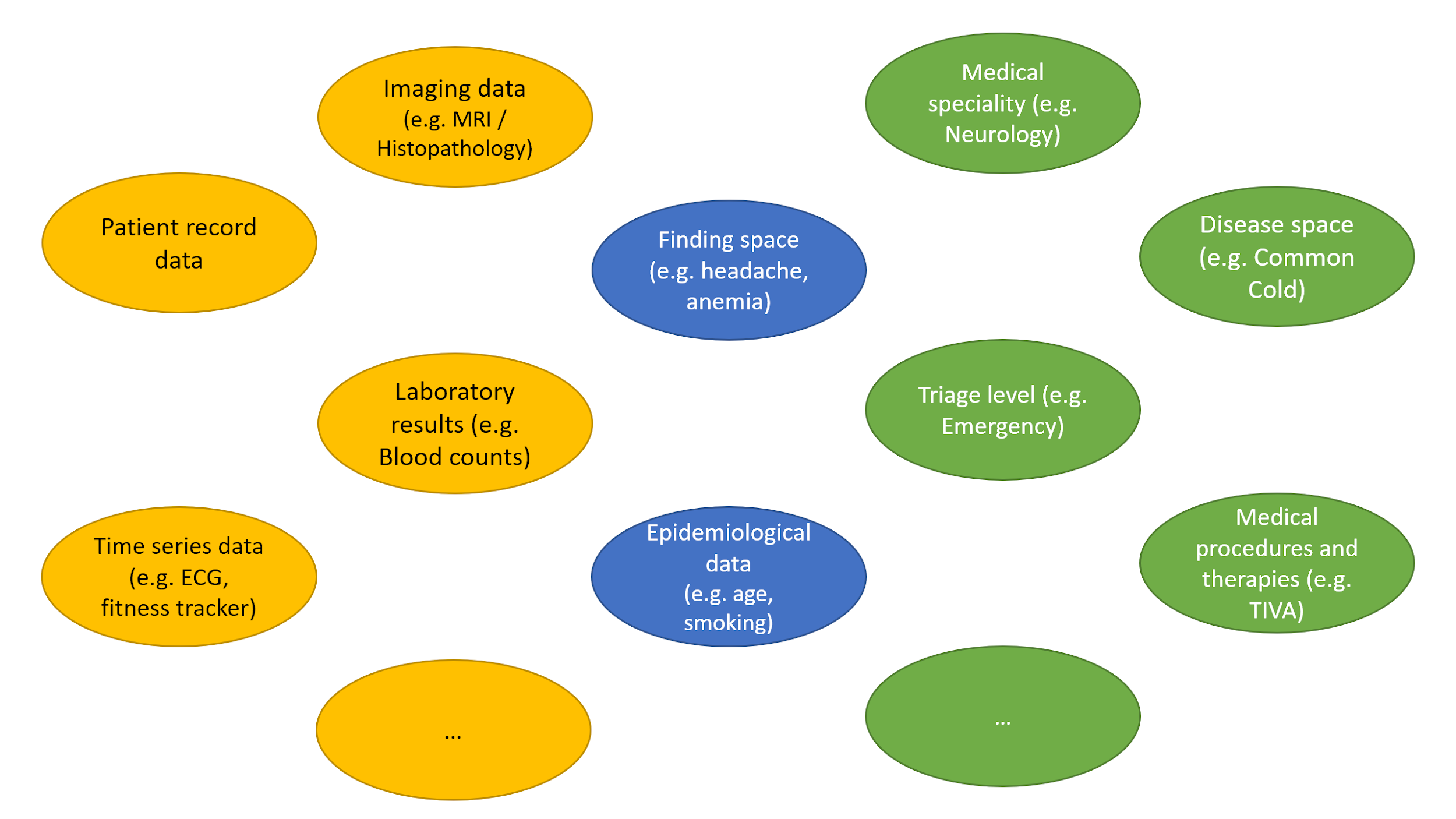
* **Epidemiological data**  
  This data type includes lifestyle variables (smoking, activity), underlying genetic risks (interpreted genetic data) and basic epidemiologic data as sex-type, age, social status
* **Medical term space**This is a wide category covering all terms that are not diagnoses. As it has an inner logic it is necessary to be declared in a complex **ontology** (e.g. SNOMED, HPO). This category is difficult to describe and maybe needs some finer sub-categorization. A complete finding space would at least **include or make use of**:
  + Symptoms as headache, abdominal pain
  + Findings as a skin rash, nodule, but also more complex terms as “nephrotic syndrome”
  + an anatomy
  + medical attributes as laterality, time dimension of symptoms etc.

3rd category (highly interpreted data and clinical course data)

* **Triage level**  
  contains the next-step advices of an AI solution including referral, emergency categorization etc
* **Disease / Condition space**  
  this is, as the “finding space” highly connected data with an inner logic and hierarchy. Therefor a complex **ontology** is required as well (e.g. ICD-11).
* **Medical Speciality**  
  For guidance and process reasons this can be discussed as an own category as AI solutions may have the primary purpose to guide users to the right specialist
* **Medical procedures and therapy**  
  therapeutic data is as well highly complex and hierarchical information and may be subclassified to pharmacologic therapy and interventions. An **ontology** (as ICHI, OPS) is necessary as well

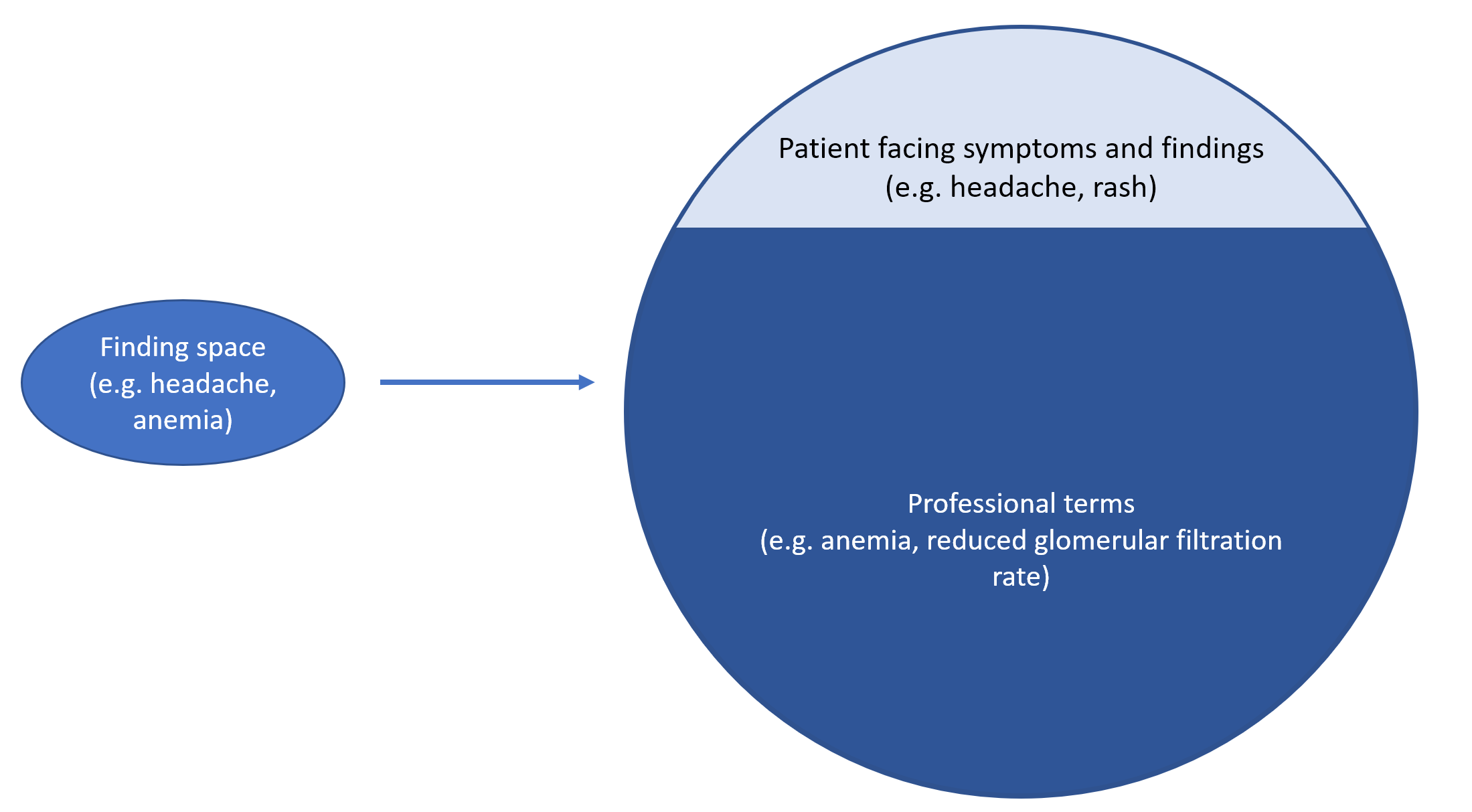
The list in neither category may be complete and the **re- and subclassification** as well as the **detailed definition** will be part of the Focus group work.

**Figure 1**: Overview over categorized data



The Subclassification of data can be done in various ways. One example for the **finding space** would be to split the space into the terms that can be understand by non-professionals and the only-professional terms (**Figure 2**).

**Figure 2**: potential (sub)classification of the finding space



## 

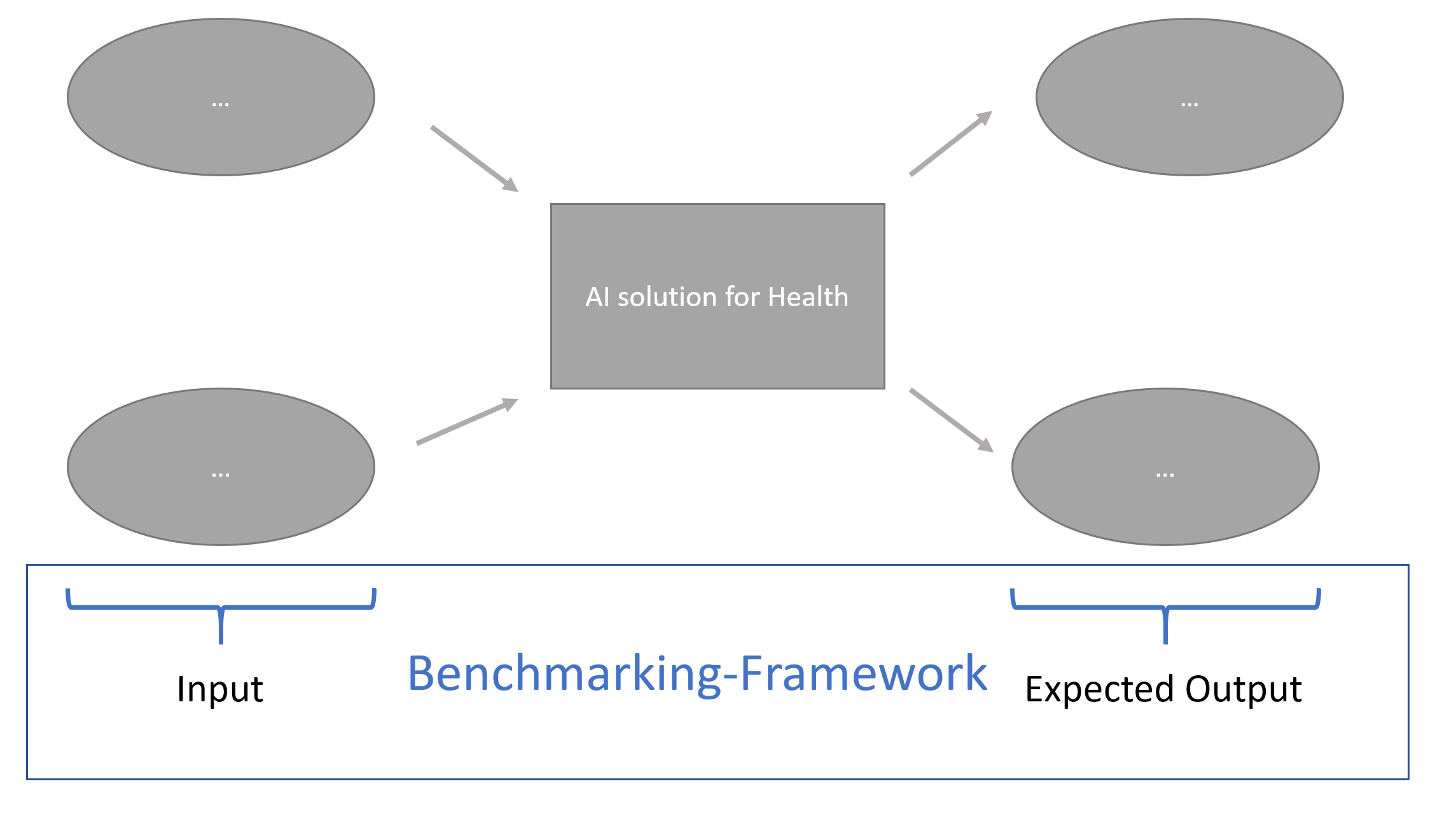
## 2. How AI4H algorithms can be classified using this formalism

The benefit of this categorization of data would be the **formalized description** of Use Cases.

Every AI4H solution can formally be described as a **transforming process of these data types** in the sense of a mapping-function (from data x1,x2,... to data y1,y2,...).

The approach would enable the Focus Group to define clearly, what the **input** and **output** of these solutions should look like and to create a **benchmarking framework** that provides the needed interface (Figure 3). This principle is generalizable to even not-yet designed AI solutions by adding maybe medical data categories if necessary.

**Figure 3**: General principle of Defining the necessary test framework for AI4H solutions by this algorithm

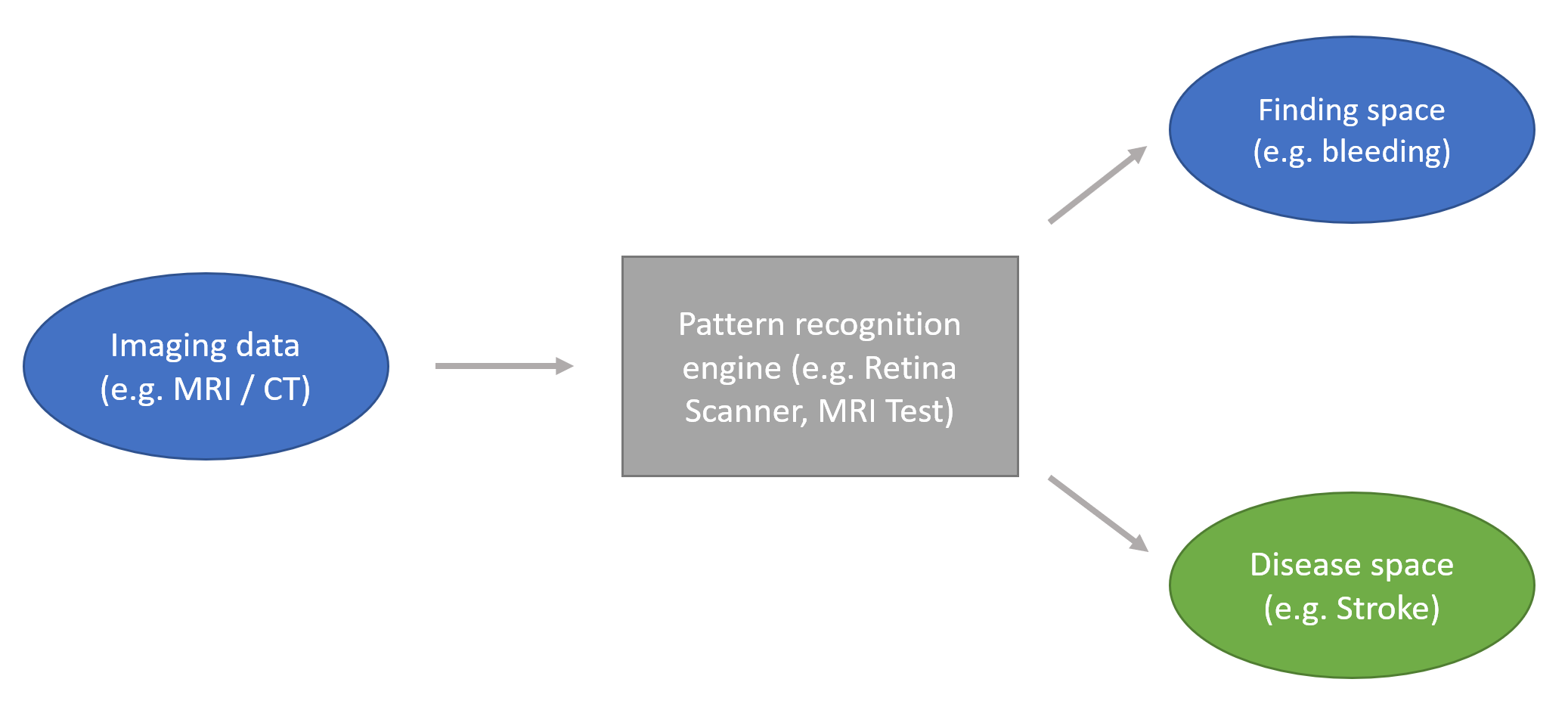


The following Figures show, how this general approach is used in the concrete test setting.

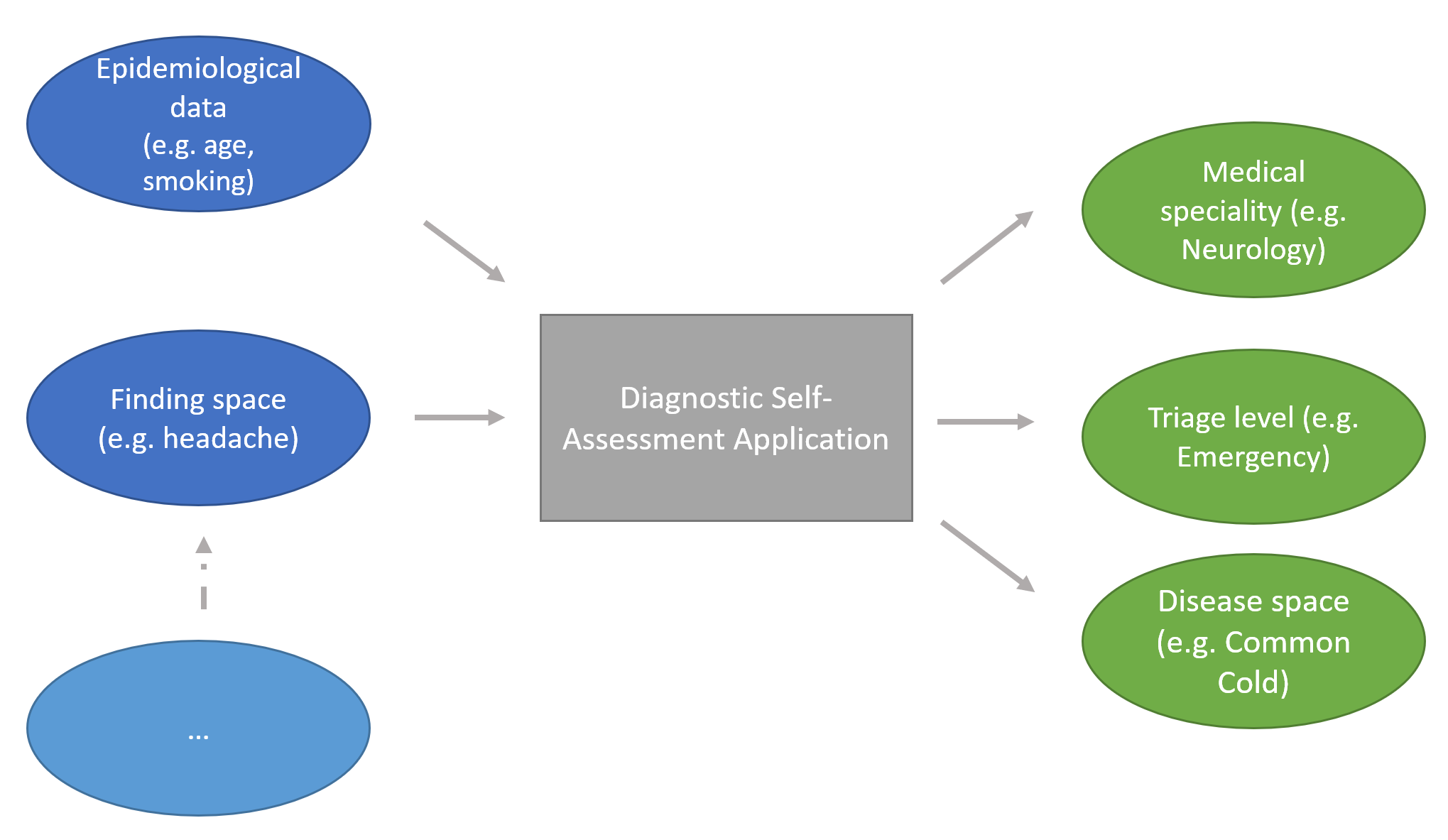
**Figure 4** exemplifies a test for pattern recognition of images (e.g. Retinopathy Scanner or Histopathological assessment). it normally **transforms** the imaging data with or without respect to meta-information into medical terms (e.g. cotton wool spots) or directly into a potential underlying disease / condition.

**Figure 5** gives an example for a diagnostic self-assessment application. This mostly multi-dimension integrating application uses maybe (pre)interpreted data of the 2nd category from the finding space and the epidemiological category. It **transforms** all this information into an advice, a possible responsible speciality and/or a potential underlying condition/disease.

**Figure 4**: Pattern recognition



**Figure 5**: Diagnostic self-assessment application



\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_