Nationales Forschungsnetzwerk der Universitätsmedizin zu COVID-19 (NaFoUniMedCOVID19)

National research network of university medicine on COVID-19

German Corona Consensus Data Set (GECCO)

Prof. Dr. med. Christof von Kalle
Prof. Dr. med. Sylvia Thun
Prof. Dr. med. Janne Vehreschild

In cooperation with experts from the Nationales Forschungsnetzwerk der Universitätsmedizin (NFN)

Table of content

1. Management summary

This document describes the creation of the "German Corona Consensus Data Set" (GECCO), a nationally uniform data set for the systematic collection of scientific data on COVID-19 in Germany.

In a transparent process involving experts from university hospitals, professional associations, and research initiatives, data elements relevant for research on SARS-CoV-2 were collected, prioritized, and consolidated in a consensus data set. This core of the data set consists of approximately 80 data elements and 330 response options for the collection of information e.g. on demography, anamnesis, symptoms, therapy, and laboratory values.

The use of international standards and terminology ensures the syntactic and semantic interoperability of the data set. The dataset is published on a national platform and can be successively extended by subject-specific extension modules.

2. Background

To cope with the current pandemic and the associated treatment of patients, the Federal Ministry of Education and Research (BMBF) is funding a national network of university medicine in the fight against COVID-19. Among other things, the network will systematically collect and bundle the data of the treated COVID-19 patients. The researchers are to collect, track, and analyze the treatment of COVID-19 patients in a standardized way.
The high threat level has led to intensive scientific activity on COVID-19, including numerous regional, national, and international epidemiological surveys and register studies. Without coordinated action and the assurance of semantic and syntactic interoperability, a common language of the collected data, it is precisely the multitude of efforts that threatens to generate a dangerous segmentation of information. The result could delay or even prevent urgently needed scientific knowledge.

The consensus data set gives the science around COVID-19 a common language and working basis.

3. Application purposes of the components

The data concepts are to be made available as interoperable components for various application purposes.
With the internationally usable components (medication, laboratory, personal data, questionnaires, diagnosis, and symptoms, procedures, case data) the optimization of collection, merging, and analysis of COVID-19 related data in different software systems and scientific institutions is possible. With uniform components, clinical studies as well as patient parameters in patient records or e.g. APPS for patient -related outcomes can be semantically standardized.

Figure 1 Components of medical information

4. Basis of the data model and approach

Basis of the data set are both the ISARIC protocol of the WHO[1] as well as the data concepts of the LEOSS-Register[2]. The purpose of ISARIC is to prevent disease and death from outbreaks of infectious diseases such as COVID-19. ISARIC brings together clinical research networks worldwide to provide the fastest possible research response to an outbreak of an infectious disease.

The ESCMID Emerging Infections Task Force (EITaF), the German Centre for Infection Research (DZIF), and the German Society for Infectiology have initiated the LEOSS network to establish a clinical patient registry for patients infected with SARS-CoV-2. The aim of LEOSS is to provide a comprehensive and cross-sectoral survey of the clinical epidemiology of patients with COVID-19 with the aim of better understanding the disease, identifying prognostic factors, and evaluating common interventions in order to improve patient care.

These data sets were merged and sent to the university hospitals for comment and prioritization. As a staring point, the most important generally valid data concepts were defined in a so-called core. In further steps, subject-specific modules will be created in the coming months.

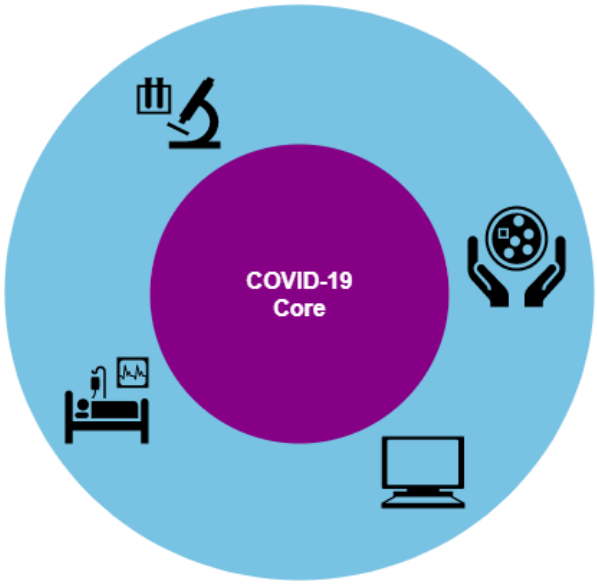


Figure 2 COVID-19 Core

Additional planned subject-specific modules are currently:

- Virology
- Radiological diagnostics
- Oncology
- Intensive care
- Pediatrics
- Epidemiology
- Pregnancy

[1] https://isaric.tghn.org/clinical-characterisation-protocol-ccp/
[2] https://leoss.net/

5. Collection of concepts and relevant response elements

To achieve optimal professionalism and high acceptance of the COVID-19 consensus data set, first of all, proposals and concepts for the data set were collected by members of an expert panel. This expert panel is composed of specialists from university hospitals, professional associations, and other relevant organizations such as the interoperability working group of the Medical Informatics Initiative. Until April 10, 2020, all members of the expert board had the opportunity to informally submit proposals for further data elements. These were reviewed by the coordination committee at the Charité and collected in a structured form by April 23, 2020. The collected data elements were then made available for further coordination and prioritization. The collected data elements were checked for duplications of data elements using unique semantic codes of international coding systems - where available and necessary - and adjusted for these duplications (see Realization with international standards).

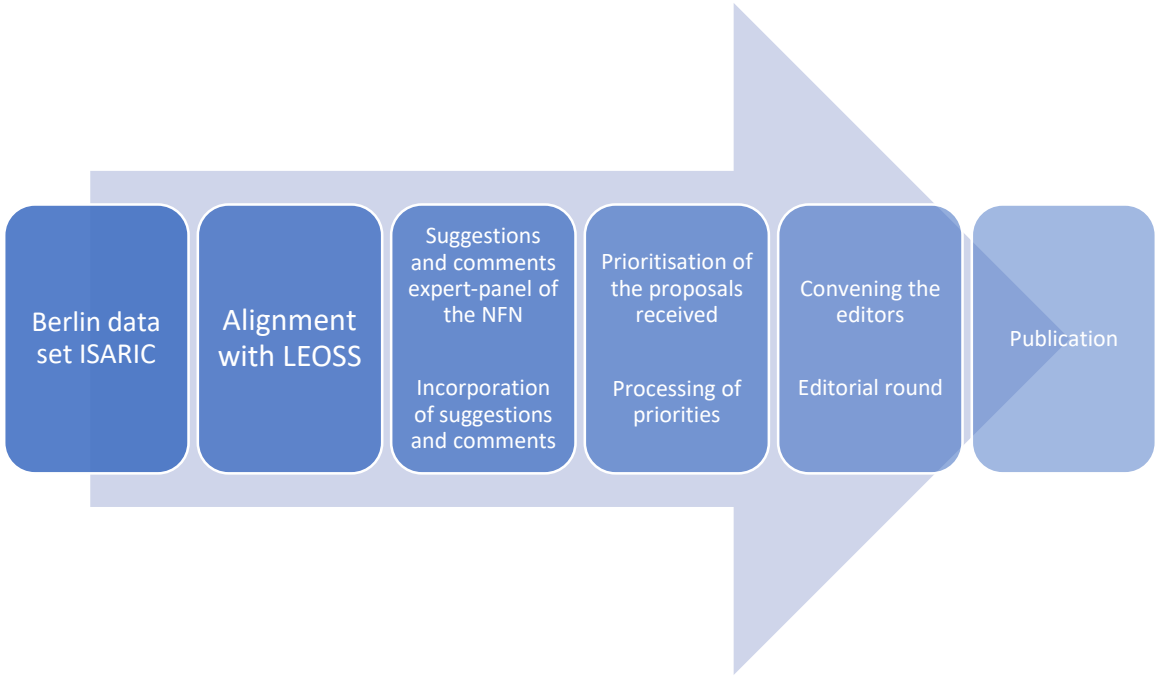


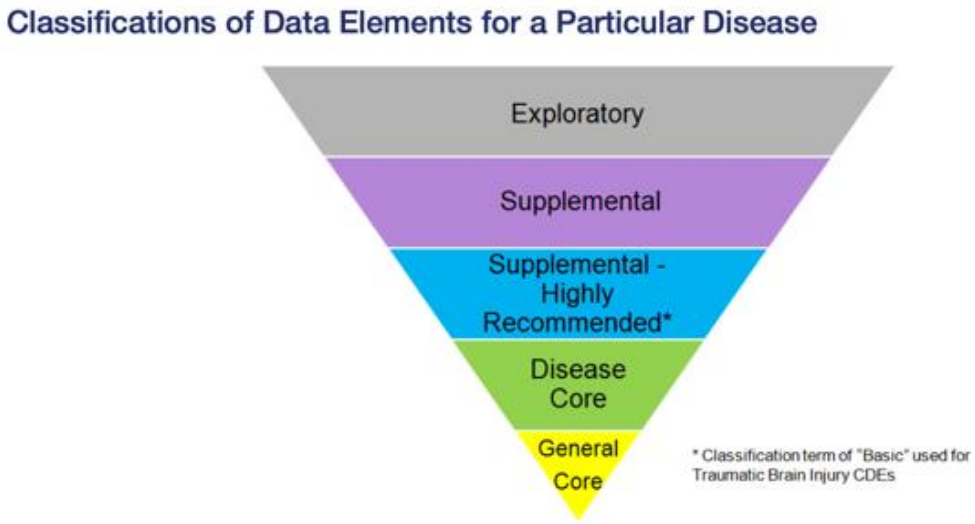
Figure 3 Workflow of consensus building

The following variables were recorded for each data element:

Variable	Explanation
Area	Area/domain of the parameter to be collected (e.g. demography, medication, vital signs, laboratory values)
Parameter case report form	Parameter / data element
International coding of the parameter	Coding of the parameter with international terminology (e.g. ICD-10, ATC, LOINC, SNOMED CT)
Response options	Response options for the parameter (e.g. yes/no)
International coding of the response options	Coding of the response option with international terminology (e.g. ICD-10, ATC, LOINC, SNOMED CT)
Prioritization	Importance of the parameter from 1-5 (5: highly relevant; 4: very relevant; 3: relevant; 2: slightly relevant; 1: not relevant)
Organization	University hospital that suggested the parameter (e.g. Charité Berlin)
Comment	Possibility for comments

6. Prioritization of data elements

To assess the relevance of the data elements, the experts were asked to assign priority values to their proposals for data elements. The priority of each data element was indicated on a 5-level scale - based on the NIH model for 'Common Data Elements' (the scale value 1, "not relevant", was not used here at first, but was used for the later, joint reconciliation and consolidation of the final core data set).



https://www.commondataelements.ninds.nih.gov/glossary#collapse-1-3

Scale value	Priority	NIH-Classification	Definition
5	highly relevant	General Core / Disease Core*	Data element with essential general or specific information relevant to COVID-19
4	very relevant	Supplemental – Highly Recommended	Data element that is essential under certain conditions or for certain study types and is therefore strongly recommended
3	relevant	Supplemental	Data element that is often collected in clinical studies, but whose relevance depends on the study design or type of research
2	slightly relevant	Exploratory	Data element that requires further validation, but which can fill current gaps in the data elements and/or replace an existing data element
1	not relevant	-	Data elements that are not considered relevant to the data set

** Since this is a disease-specific data set (i.e. COVID-19), both the general and disease-specific core categories of the NIH ("Core") were assigned to the highest priority level.*

After all the proposed data elements had been combined, the members of the Expert Board were able to vote on the individual data elements using the 5-level scale described above. Based on these prioritizations of the experts, the data set was consolidated and a final consensus data set was created.

7. Decision on concepts in the National Task Force in the editorial team

Based on the priority values provided by the members of the expert panel a decision was made on the inclusion of the data elements in the final COVID-19 core data set.

The aim was to define a data set on time that contains as many relevant data elements as possible while remaining as manageable and practical as possible. To define this COVID-19 core data set, the data elements with high priority were selected.

These elements were reviewed by an editorial team of experts from various fields and supplemented by other elements considered important. Elements that were not considered essential were deleted or moved to the subject-specific extension modules so that in the end about 80 concepts for questions and 330 possible answers for the core data set were defined.

8. Realization with international standards

To ensure syntactic and semantic interoperability of the data set, the data elements are mapped to international standards and terminology. Using these standards also helps to avoid duplication of elements when creating the data set (see above). To describe the data elements semantically unambiguously, the following terminologies and code systems are used, for example: ICD-10 (for diagnoses); LOINC (for laboratory values or other measurements); UCUM (units of measurement); ATC (active ingredients); SNOMED CT (for diagnoses and many other medical concepts).

The "Fast Healthcare Interoperability Resources" (FHIR) standard is used for syntactically standardized transmission. FHIR can be used to define uniform data structures, so-called profiles (e.g. for laboratory tests), which enable interoperable data exchange via different IT systems. When defining the FHIR profiles, it is possible to partly use preliminary work, especially the profiles of the Medical Informatics Initiative and the basic profiles of HL7 Germany. This facilitates the manufacturer-neutral filling of the data set, especially from the data integration centers of the NFN members.

The following FHIR resources represent the main part of all concepts of the core data set:

- Patient (e.g. DEMOGRAPHICS)
- Consent (e.g. STUDY ENROLLMENT / INCLUSION CRITERIA)
- Observation (e.g. LABORATORY VALUES)
- Condition (e.g. ANAMNESIS / RISK FACTORS)
- Procedure (e.g. THERAPY)
- Encounter (e.g. DISCHARGE)
- Medication/MedicationStatement (e.g. MEDICATION)

9. Reconciliation through HL7 Germany with the participation of the Medical Informatics Initiative

Specialists for standardization are involved in the final reconciliation of the data set. These include the standardization organization Health Level 7 (HL7) / Integrating the Healthcare Enterprise (IHE) Germany as well as members of the Medical Informatics Initiative (MII), who have already offered a great deal of support in the context of their definition of a Germany-wide core data set for university hospitals.

After the announcement of the vote, experts will register for the ballot procedure and provide comments. The balloting is subject to the rules of HL7 Germany and HL7 International.
The procedure for voting on a specification is internationally established and is also used, for example, in the Medical Informatics Initiative.

10. Publication of the components and extensions on a national platform

The final data set will be published on a national platform. Temporarily the platform will be provided by HL7 Germany, the collaboration platform cocos.team and the tools commonly used in Europe (Simplifier/Art-Decor). The coordination and implementation of the medical-technical requirements of GECCO will remain with the National Research Network NFN.

The COCOS platform was initiated by the Federal Association of Statutory Health Insurance Physicians (KBV), the health innovation hub (hih) of the Federal Ministry of Health, Health Level 7 (HL7) Germany, the Robert Koch Institute (RKI), the Federal Institute for Drugs and Medical Devices (BfArm) together with the German Institute for Medical Documentation and Information (DIMDI), the Medical Informatics Initiative (MII), the University Medicine Network and the Federal Association for Health IT (bvitg) as well as supporting organizations such as gematik, IHE Germany and the Interoperability Forum. COCOS initiative aims to establish uniform data formats and standards for interoperability for COVID-related data and their combination, thus providing users, developers, and researchers with a basis on which they can work together in an interoperable manner.

Via the cocos.team platform, recommendations for developed standards are presented and new standards are proposed, which can be used for the development of solutions against corona and facilitate the consolidation of data, among others for research purposes.

Among the first definitions available via COCOS are the core profiles of the KBV, modules of the core data set of the Medical Informatics Initiative as well as COVID-specific data e.g. from GECCO or FHIR profiles of HL7 Germany and the profiles of the RKI for the electronic SARS-CoV-2 pathogen detection.

- Further definitions and extensions may follow if necessary. Proposals can be submitted for this purpose, and a decision on their inclusion will be taken in a voting procedure.

Governance for the introduction of further modules and elements will be agreed with all COCOS actors.

A kick-off for GECCO took place on June 6, 2020, within the framework of the COCOS initiative cocos.team of the Health Innovation Hub (hih) of the Federal Government.

Figure 4 COCOS Component Standards

11. Results of the selection of data concepts for the core data set

Approximately 80 data set concepts with 330 possible answers were selected according to defined criteria in a transparent process. These are presented below as simple terms. The semantic annotation in SNOMED CT, LOINC, UCUM, ICD, and ATC as well as the beginning work on FHIR modeling can be accessed online at any time.

Response options, reference ranges, units of measurement, and semantic annotations were listed in the detailed description in ART-DECOR.

ID	CATEGORY	PARAMETER CASE REPORT FORM	RESPONSE OPTIONS
1	ANAMNESIS / RISK FACTORS	Does the patient suffer from a chronic lung disease?	Asthma | COPD | Fibrosis of lung | Pulmonary hypertension | Extreme obesity with alveolar hypoventilation | Sleep apnea | Obstructive sleep apnea syndrome | Cystic fibrosis | Other | No | n. a.
2	ANAMNESIS / RISK FACTORS	Does the patient suffer from a cardiovascular disease?	High blood pressure | History of myocardial infarction | Cardiac arrhythmia | Heart failure | Peripheral arterial occlusive disease | History of heart revascularization | Coronary arteriosclerosis | Carotid artery stenosis | Other | No | n. a.
3	ANAMNESIS / RISK FACTORS	Does the patient suffer from a chronic liver disease?	Steatosis of liver | Cirrhosis of liver | Chronic infectious hepatitis | Autoimmune liver disease | Other | No | n.a.
4	ANAMNESIS / RISK FACTORS	Does the patient suffer from at least one rheumatological/immunological disease?	Inflammatory bowel disease | Rheumatic arteritis | Disorder of connective tissue | Vasculitis | Congenital immunodeficiency disease | Other | No | n. a.
5	ANAMNESIS / RISK FACTORS	Is the patient infected with HIV?	Yes | No | n. a.
6	ANAMNESIS / RISK FACTORS	Does the patient have a history of being an organ transplant recipient?	Heart | Lung | Liver | Kidneys | Intestine | Skin | Cornea | Ossicle | Heart

			valve | Blood vessel | Cerebral meninges | Bone tissue | Cartilage tissue | Tendon | No | n. a.
7	ANAMNESIS / RISK FACTORS	Does the patient suffer from diabetes?	Diabetes mellitus type 1 | Diabetes mellitus type 2 | Insulin treated type 2 diabetes mellitus | Secondary diabetes mellitus | No | n. a.
8	ANAMNESIS / RISK FACTORS	Does the patient suffer from at least one active tumor/cancer disease?	Active | Remission | No disease | No | n. a.
9	ANAMNESIS / RISK FACTORS	Has the patient ever smoked cigarettes?	Yes | Nonsmoker | Former smoker
10	ANAMNESIS / RISK FACTORS	Does the patient suffer from at least one chronic neurological disease or psychiatric illness? Which one?	Parkinson's disease | Dementia | Multiple sclerosis | Combined disorder of muscle and peripheral nerve | Epilepsy | Migraine | History of a cerebrovascular accident with residual deficit | History of a cerebrovascular accident without residual deficits| Psychotic disorder | Depressive disorder | Anxiety disorder | No | n. a.
11	ANAMNESIS / RISK FACTORS	Did oxygen or respiratory therapy already exist before the current illness?	Yes | No | Unknown
12	ANAMNESIS / RISK FACTORS	Does the patient suffer from chronic kidney disease?	With hemodialysis | Without hemodialysis | No | Unknown (Severity)
13	ANAMNESIS / RISK FACTORS	History of travel	Destination of travel
14	ANAMNESIS / RISK FACTORS	Gastrointestinal ulcer	Yes | No | Unknown
15	ANAMNESIS / RISK FACTORS	Immunization status	Influenza | Pneumococcal | BCG | COVID-19 | Other
16	ANAMNESIS / RISK FACTORS	Do-not-resuscitate-order	Yes | No | Unknown
17	IMAGING	Imaging procedures of the lung	CT | Radiography | Ultrasonography

18	IMAGING	Radiological findings	Undetermined finding | COVID-19-characteristic finding | Normal finding
19	DEMOGRAPHICS	Pregnancy	Yes | No | Unknown | n.a.
20	DEMOGRAPHICS	Ethnic group	Caucasian | African | East-/West-Asian | Arabs | Latein-American
21	DEMOGRAPHICS	Sex assigned at birth	Male | Female | Not answered| Diverse
22	DEMOGRAPHICS	Date of birth	
23	DEMOGRAPHICS	Frailty score before admission	
24	DEMOGRAPHICS	Body weight	kg
25	DEMOGRAPHICS	Body height	cm
26	EPIDEMIOLOGICAL FACTORS	Did the patient knowingly have contact with a person with probable or proven COVID-19 disease within 14 days before the onset of his/her symptoms?	Yes | No | Unknown
27	COMPLICATIONS	Thromboembolic events	Venous thrombosis | Pulmonary embolism | Stroke | Myocardial infarction | Other
28	COMPLICATIONS	Infectious disease of lung	Yes | No | n. a.
29	COMPLICATIONS	Infectious agent in bloodstream	Yes | No | n. a.
30	ONSET OF ILLNESS / ADMISSION	Stage at diagnosis	Uncomplicated phase | Complicated phase | Critical phase | Recovery phase | Dead | Unknown
31	LABORATORY VALUES	CRP	nmol/L | mg/L | mg/dL
32	LABORATORY VALUES	Ferritin	ng/mL | /{cell} | pmol/L | ug/L
33	LABORATORY VALUES	Bilirubin	mg/dL | umol/L | /[HPF] | mL/min
34	LABORATORY VALUES	D-dimer	ug/mL | ng/mL | ug/L | ug/mL {FEU} | ng/mL{FEU} | ug/L {DDU}
35	LABORATORY VALUES	Gamma-GT	U/L
36	LABORATORY VALUES	GOT/AST	U/L

37	LABORATORY VALUES	LDH	U/L
38	LABORATORY VALUES	Cardiac troponin	ug/L | ng/L | ng/mL
39	LABORATORY VALUES	Hemoglobin	mmol/L | g/L | g/dL |mg/L | mg/dL
40	LABORATORY VALUES	Creatinine	mg/dL | umol/L | mmol/L
41	LABORATORY VALUES	Lactate	mg/dL | mmol/L
42	LABORATORY VALUES	Leukocytes absolute	10*3/uL | /[HPF] | /uL |
43	LABORATORY VALUES	Lymphocytes absolute	10*3/uL
44	LABORATORY VALUES	Neutrophils absolute	10*3/uL
45	LABORATORY VALUES	PTT	s
46	LABORATORY VALUES	Platelets absolut	10*3/uL
47	LABORATORY VALUES	INR	{INR}
48	MEDICATION	COVID-Therapie	Antipyretics | Corticosteroids | Atazanavir | Darunavir | Chloroquine phosphate | Hydroxychloroquine | Ivermectin | Lopinavir/ritonavir | Ganciclovir | Oseltamivir | Remdesivir | Ribavirin | Camostat | Favipiravir | Convalescent plasma | Steroids (> 0.5 mg/kg prednisone equivalents) | Steroids (<= 0.5 mg/kg prednisone equivalents) | Tocilizumab | Sarilumab | CNI or mTor inhibitors (e.g. cyclosporin A, tacrolimus, sirolimus, everolimus) | Anti-TNF-alpha inhibitors (e.g. adalimumab, etanercept) | Il1-receptor antangonists | Ruxolitinib | Colchicine | Interferone

			(any) | 25-Hydroxyvitamin D | Zinc | Other
49	MEDICATION	ACE inhibitors	Yes | No | Unknown | stopped
50	MEDICATION	Immunoglobulins	Yes | No | Unknown
51	MEDICATION	Anticoagulation	Unfractionated heparin | Low molecular weight heparin | Argatroban | Platelet aggregation inhibitor | Danaparoid | Phenprocoumon | direct oral anticoagulants | Other (Thrombosis prophylaxis | Therapeutic anticoagulation)
52	OUTCOME AT DISCHARGE	Respiratory outcome	On ventilator | Not on ventilator
53	OUTCOME AT DISCHARGE	Type of discharge	Alive | Admission to hospital | Referral to other institution | Death | Palliative discharge | Unknown
54	OUTCOME AT DISCHARGE	Follow-up swab result	Pos | Neg
55	STUDY ENROLLMENT / INCLUSION CRITERIA	Confirmed COVID-19 diagnosis as main reason for admission	Yes | No | Unknown
56	STUDY ENROLLMENT / INCLUSION CRITERIA	Is there a signed consent form available?	Study consent | MII Broad Consent | LEOSS | CAPNETZ | None | Other | Withdrawn
57	SYMPTOMS	Loss of taste or smell	Yes | No | Unknown (Severity)
58	SYMPTOMS	Abdominal pain	Yes | No | Unknown (Severity)
59	SYMPTOMS	Clouded consciousness	Yes | No | Unknown (Severity)
60	SYMPTOMS	Diarrhea	Yes | No | Unknown (Severity)

61	SYMPTOMS	Vomiting	Yes | No | Unknown (Severity)
62	SYMPTOMS	Cough	Yes | No | Unknown (Severity)
63	SYMPTOMS	Dyspnea	Yes | No | Unknown (Severity)
64	SYMPTOMS	Nausea	Yes | No | Unknown (Severity)
65	SYMPTOMS	Fever	Yes | No | Unknown (Severity)
66	SYMPTOMS	Headache	Yes | No | Unknown (Severity)
67	THERAPY	Dialysis / Hemofiltration	Yes | No | n. a.
68	THERAPY	Apheresis	Yes | No | Unknown
69	THERAPY	Prone position	Yes | No | n. a.
70	THERAPY	ECMO therapy	Yes | No | n. a.
71	THERAPY	Ist he patient in the intensive care unit?	Yes | No | n. a.
72	THERAPY	Type of ventilation	Nasal High-Flow-Oxygen-Therapy| Non-invasive ventilation | Invasive ventilation | Tracheotomy | No | n. a.
73	VITAL SIGNS	PaCO2	mm[Hg]
74	VITAL SIGNS	PaO2	mm[Hg]
75	VITAL SIGNS	FiO2	%
76	VITAL SIGNS	pH-value	[pH]
77	VITAL SIGNS	SOFA-Score	
78	VITAL SIGNS	Respiratory rate	{breaths}/min
79	VITAL SIGNS	Blood pressure diastolic	mm[Hg]
80	VITAL SIGNS	Blood pressure systolic	mm[Hg]
81	VITAL SIGNS	Heart rate	{beats}/min

82	VITAL SIGNS	Body temperature	Cel
83	VITAL SIGNS	Peripheral oxygen saturation	%

12. Results in ART-DECOR (work in progress)

The data set is created in the ART-DECOR® tool based on international standards. ART-DECOR® is an open-source tool suite that supports the creation and maintenance of data set descriptions, data use scenarios, value sets, and HL7 templates. The tool has cloud-based federated repository modules for this purpose. It supports the comprehensive collaboration of team members within and between governance groups. The usage is free of charge.

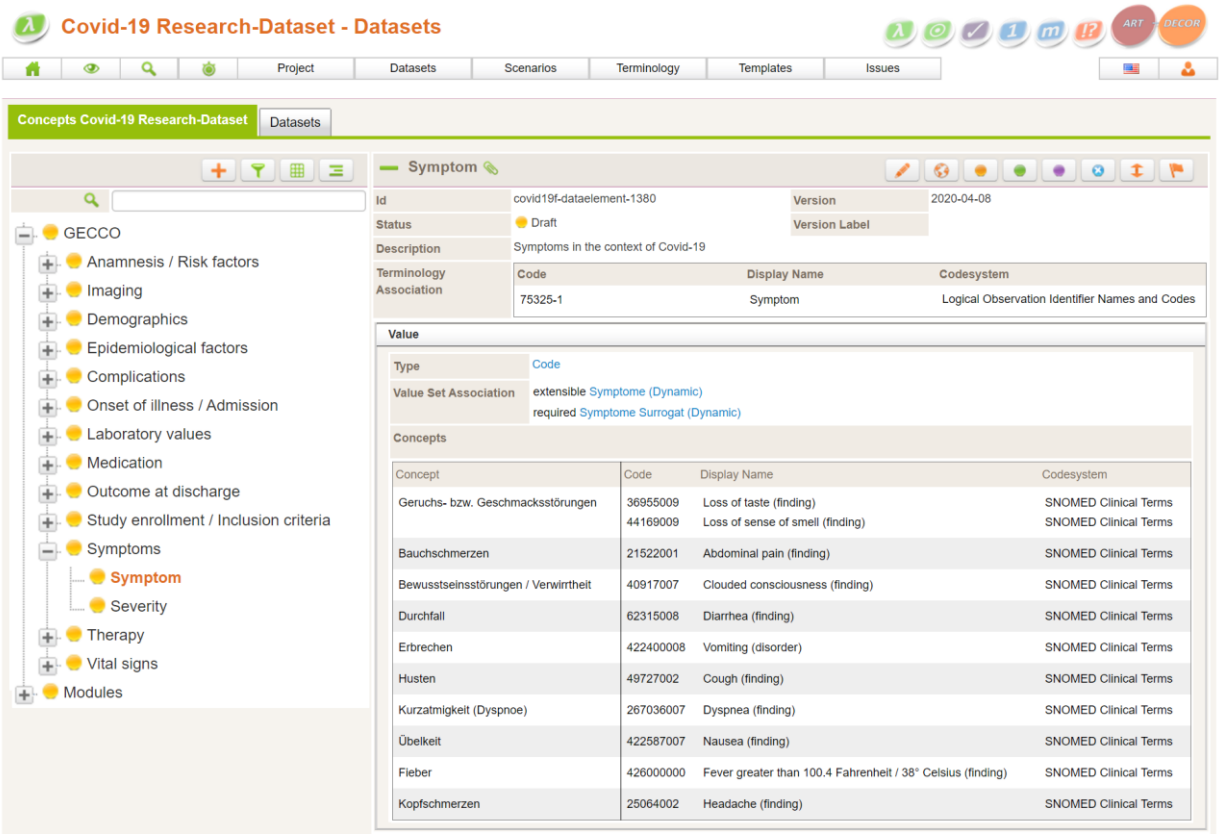


Figure 5 Data set in ART-DECOR (https://art-decor.org/art-decor/decor-datasets--covid19f-)

13. Further measures

Further modules will be developed successively and depending on the application scenario via the coordination office of the NFN. Quality control takes place continuously. Submission of LOINC concepts for missing data elements to the Regenstrief Institute, Indianapolis, is ongoing. An implementation guideline is created for the core data set in SIMPLIFIER.

Versioning, voting, and technical governance platform are planned in ART-DECOR and SIMPLIFIER.

The implementation of the data set in the NFN locations by the data integration centers using decentralized data entry structures will be promoted. The data quality requirements are defined centrally by the NFN coordination office.

Implementations of the open-source specification of the Federal German Data Set (Core) for COVID-19 in LEOSS and other software systems, registers, eCRFs, and APPs are the prerequisites for the funding of projects within the National Research Network of University Medicine on COVID-19.