Introduction
It became clear during the first wave of the COVID-19 pandemic that the disease progression is highly variable in different people. For some patients an infection results in minor symptoms, while others suffer from severe to critical symptoms that require hospitalisation and intensive care. It is currently recognised that the severe disease is a complex, multifactorial interplay that includes, but may not be restricted to, aberrant immune responses, endothelial and vascular problems, as well as thrombotic events. However, the manifestations differ widely and multiple disease models will be needed to understand the phase in which a particular patient is. ZonMW will fund a series of projects related to COVID-19 and many of those will need ‘rationalisation’ of their hypotheses in a commonly agreed set of disease models in order to develop a consistent approach to the disease in Dutch points of care.

The COVID-19 HOTEL@ DTL will be set up under auspices of Health RI and will aim to serve all awarded projects in the Bottom Up ZonMW call, with a FAIR data enabled in silico rationalisation workflow to rapidly rationalise phenomena observed in real world situations, generate hypotheses that can be rapidly laboratory tested (for instance at LACDR) and subsequently serve decision making in the clinical setting (for instance the LUMC dashboard). The Hotel facility is explicitly not meant to actually make data generated in other ZonMW Projects FAIR, which is covered as part of the individual projects, and FAIR data points are available and can be easily installed in the institutions. In some cases, external source data needed for data analysis may need to be made FAIR ‘in the hotel’ and health-RI, DTL and the GO FAIR Foundation will offer support through the Hotel and in dedicated M4M workshops¹ for such key external data to be FAIRified, according to community endorsed FAIR Implementation Profiles (FIP)² that are reusable. It is of particular importance that the current COVID-19 efforts are not carried out in splendid isolation and that preventive and therapeutic decisions as well as policies are made on optimally available and trusted data, obviously with full compliance to GDPR and other regulations.

DTL uniquely positioned to offer this overarching disease modelling and analytics service, which will benefit all other projects and ‘connect them at the data’.

Background on in silico disease modelling and workflow in the COVID-19 Hotel
The rationalisation of our insights into the complex disease that SARS-Cov-2 causes needs many different data sources. Gradually, more formally published information becomes available³, unfortunately mostly in textual format (not FAIR). Next, Case Report Forms are in many cases also not yielding FAIR data. In a preparatory project, cofounder by ZonMW and the Philips foundation, the GO FAIR foundation produced a FAIR (RDF), extendable COVID-19 data model⁴, initially mainly based on the WHO CRF, that is now offered by our consortium for re-use and extension in all research and clinical settings, but for instance also to the popular self reporting apps, such as COVID Radar and COVID-check.

Collectively, we qualify such recently published articles, recent clinical studies, direct clinical observations and measurements as well as citizen self reporting as ‘Real World Observations’ (RWO). Obviously, full provenance of origin and methods needs to be stored to make optimal use of these different categories of data and to allow their inclusion and exclusion from analysis at will.

It is further important to combine these RWO with the largest possible integrated (Fully AI Ready) collection of established prior knowledge available. This typically comes from the literature (after error-prone text mining), but preferably also from a large collection of curated databases (> 200) such as maintained by EBI and NCBI and integrated by one of our DTL partners, Euretos. The combination of these data sources based on particular hypotheses or clinical patterns observed leads to in silico disease models that can be used to rationalise insights and support the decision to either first validate the assumptions in vitro, in vivo and/or mover straight to clinical implementation, in case of already validated interventions (for instance drug repurposing).

The COVID-19 Hotel will arrange academic licenses with commercial tools, and supported access to preconceived disease models, and offer access to the data analytics environment as a service to each project supported (initially by ZonMW). Some projects may have budgeted for the data analytics and rationalisation and they can choose to either use that budget for internal data analytics and disease modelling or to pay for services in the Hotel, which may save considerable costs and allow to combine internally generated date with relevant, available FAIR data from other projects and beyond to prior, curated knowledge bases. In this project, we therefore do not budget for these already funded services, but we could offer free at the point of use services for all projects based on a core subsidy for generic disease modelling.

¹ link to M4M documentation
² FIP: reference
³ substantially more than 10,000 articles to date
⁴ link to data model
Description of the methodology of \textit{in silico} disease modelling and workflow in the COVID-19 Hotel

\textbf{A1}: Publications relevant for the biological process under study (here COVID-19) are collected and the collective biomedical concepts mentioned in the sources are listed manually as a ‘set’, which can be loaded in semantics-aware systems such as the Euretos AI platform (EP). Next to mentions in multiple articles and other sources, this is partly a human curation process. The resulting set is visualised in the ‘Relation Map’ functionality of the EP (supplementary example figure 1), which automatically maps all combinations of subjects and their objects as cardinal assertions with one or multiple predicates connecting them. This process also maps each term mentioned in the resources to a proposed globally unique identifier (mapping also between different technology systems and adding the definition -for example from BioPortal, which connects the Unified Medical Language System, UMLS or UniProt- etc.). We call this set of connected subjects, predicates and objects as mapped and visualised in the relation map, a ‘Connectome’, in order to acknowledge that not all connections may be biologically meaningful as is supposedly the case in an annotated ‘Interactome’ between for instance genes, proteins and small molecules (such as established pathways). However, although the EP also recognises co-occurrences in text and some sets may contain many of these unqualified, potential associations, the Relation Map only shows subject-predicate-object cardinal assertions that are already present in minimally one peer reviewed, credible resource (currently link).
COVID-19 Hotel

Health-RI/DTL/GO FAIR Foundation

more than 200 resources covered). Multiple articles and concept sets can be evaluated in the EP to reveal the conceptual overlap (intersect) between the resources and the relevant concepts (expert annotation) can be retained in the ‘greatest common denominator’ list

A2: In addition to formal publications, many data are collected that represent various categories of RWO, including Case Report Forms (CRF), clinical measurements (for instance read outs of cytokine markers) and self reported assertions about for instance symptoms in Smartphone Apps. These all need to be made FAIR (preferably at the source, and prior to submission in the Hotel) in order to allow seamless linking to established knowledge as well and to detect cardinal assertions (multiple assertions stating the same association). It should be emphasised that such data should be stored with rich enough provenance to reason about the source, its credibility and to enable selective filtering of data sources.

B: From the perspective of a given hypothesis, relevant, new and existing, articles and databases can also be used to generate additional concept sets that would be relevant to either strengthen or falsify a given hypothesis. We avoid the term ‘validate’ here as we want to emphasise that this is an in silico modeling approach that always needs further expert annotation and experimental clinical validation. Therefore we use the term ‘in silico rationalisation’ for this step in the overall validation process.

C: Next, we create the greatest common denominator (GCD) of the relevant sets (usually including all concepts that are present in minimally two, independent relevant sets) from collection A and B. And this GCD-set is reviewed by the collaborative team for potentially irrelevant concepts (for instance too generic, like ‘disease’ or concepts that are wrongly mined from the source (many terms used in text are highly ambiguous). This set is now qualified as a ‘core -sub- hypothesis’ set and can be delivered as a set as well as further analysed and visualised in a relation map. In many cases the graphic version would qualify as a ‘ridiculogramme’ as it is far too complicated for the human mind to comprehend, like biology itself, (example fig. 2). But there are multiple ways to filter and simplify subsets of the graph so as to use the GCD set for ‘in silico experiments’. A quick overview can be generated for instance of all genes/proteins with which a particular chemical proposed for intervention will interact (supplementary example figure 3).

D: Subject experts, subsequently try to further categorize GCD, for instance through grouping concepts by semantic types (genes and proteins together, diseases together, processes, pathways etc.) In addition, we can for instance make a proxy ‘timeline’ of phasing of a disease process. For example, in the hypothesis under study at the moment6, the supposed progression from mild cytokine release syndrome and temporary symptoms to cytokine storm, vascular and thrombotic process, to organised pneumonia, acute respiratory distress syndrome, heart failure, stroke and finally fatal outcome can be used to ‘rationalise’ phase specific biomarkers, intervention targets and drugs.

E: once this ‘consolidated hypothesis’ GCD-set is created and published, experts from the field (in fact anyone with approved access to the Hotel for COVID-19 related evaluations) can bring in their own disease model and/or suggest additional concepts of which they want to study the ‘additional connectome’ (i.e. revealing the connections the concept (or the set of concepts) between the newly added concept and the consolidated hypothesis model. For COVID-19, and for novice users, this will usually be done in the Hotel with guiding consultation to avoid the need for a steep learning curve under high time pressure. After an agreed number of ‘Hotel sessions’, academic users can take a license on the tools they like ‘at home’ and or extend their hotel service for a fee.

F: Once a rationalisation of a particular hypothesis (for instance a proposed drug for intervention) is positive, the net step will usually be in vitro and experimental validation. This is not an integral service of the Hotel, but several DTL partners can offer collaborative services, such as -omics measurements, organ on a chip (organoids) or other n vitro assays. Health RI can also supervise the transition from hypothetical intervention to clinical implementation, for example the reuse of an existing drug approved for another disease. For instance, administering heparin is no longer even really ‘drug repurposing’ if the virus-induced pathology includes hyper coagulation and thrombosis. In the case of COVID-19 we will likely first evaluate existing drugs, several of which are already in early trials for repurposing towards COVID related syndromes. We will attempt to not only evaluate their potential molecular interactions in the connectome of the COVID-19 induced thrombosis and vascular damage related disease outcomes, but also try to rationalise the time and phase at which particular drugs may be most effective versus likely to induce adverse effects.

G: One more added value of the hotel is that experts from many different institutes and subdisciplines will contribute their specialist knowledge to improved several sub-models describing the disease. We imagine several models, focused on the immune induced side effects, one the effect of age, smoking, specific comorbidities, environmental factors, vac cine development and smart diagnosis. The basic rule of the Hotel is that all guest contributions are open and reusable for all other hotel guests. If ‘private rooms’ are needed, the hotel will refer the user to the external providers that offer such services.

H: There will also be a facility for community annotation (for instance for interested patients),

6 link to OSF article
A preferred Health-RI publishing channel?

FAIR, machine readable Disease models can be connected to a ‘supplementary article’ in the Open Science Platform of Frontiers (authors may also decide to publish in other environments offering the same features). However, the Frontiers Open Science Platform offers unique annotation services [in collaboration with DTL, Health-RI, Euretos and the GO FAIR Foundation], to couple machine readable disease models, in which ‘in silico experiments’ can be performed, and disease models (here restricted to COVID-19) can be continuously refined and updated (versioned). Each article (soon all articles in Frontiers and beyond?) has a FAIR digital abstract, which is an author-approved collection of all relevant concepts mentioned in the article and their annotated relationships as derived from the Euretos platform (> 250 data sources in the Life Sciences). FAIR digital abstracts are fully interoperable with the in silico disease models, and any other FAIR resource. Euretos has retrospectively created FAIR digital abstracts of all abstracts in PubMed (> 40 M, not yet curated by authors), which means that in principle, each article can serve as a source for additions to in silico models (in [1] creation of manual, curated FAIR digital abstracts of new articles took around one hour per article).

For COVID-19 (in the scope of VODAN) we will publish the first in silico disease model with a supplementary foundational paper in [Frontiers in Medicine?] and we will also create FAIR digital abstracts of all novel articles published on COVID-19 (> 10,000), as part of the TWOC project [reference]. Annotation of new articles and other textual resources will be integrated in the FAIR digital abstracts. The COVID-19 disease model foundational paper will be community peer reviewed and annotated, and future versions may include co-authors who have significantly contributed to the model (also versions) and/or the article. All annotations and contributions in TWOC as well as the public CIOVID-19 hotel will be immediately published in open access as nanopublications. In case academic or private parties wish to set up annotation rooms with restricted access and without automatic publication, they will need to purchase a separate license with the respective companies participating in the COVID-19 Hotel.