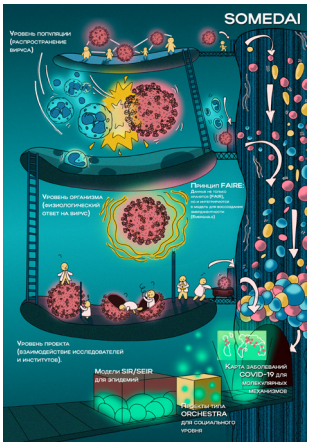
Methodology article

MAKING DATA FINDABLE, ACCESSIBLE, INTEROPERABLE, REUSABLE, AND EMERGEABLE (FAIRE): PROPOSING A SILICON ORGANISM MODEL-ENGROSSED DATA INTEGRATION (SOMEDAI) APPROACH IN SYSTEMS BIOLOGICAL PREPARATIONS FOR THE NEXT EPIDEMICS

A. N. Kolodkin[1, 2]*



[1] Competence Center for Methodology and Statistics,
Luxembourg Institute of Health,
Translational Medicine Operations Hub,
L-1445, Strassen, Luxembourg
[2] LCSB — Luxembourg Centre for Systems Biomedicine,
University of Luxembourg,
L-4367, Belvaux, Luxembourg

* Correspondence: alexeykolodkin@gmail.com

Abstract: This manuscript is dedicated to the concept of emergence in modern systems biology and discusses the essence of this concept, its role, and its implementation using the example of epidemic modeling for COVID-19, where three dimensions of emergence should be integrated: (i) the emergence of the epidemic and the spread of the virus in a population from interactions between susceptible, infected, and recovered individuals; (ii) the emergence of health-related properties of one individual from interactions between biomolecules; (iii) the emergence of the research project from interactions between various institutions collecting different pieces of clinical data.

Systems biological understanding of biological emergence may require changes in data management approaches. By adding the concept of Emergence, classical FAIR data management may be extended to FAIRE, where data become not only Findable, Accessible, Interoperable, and Reusable, but also Emergeable (FAIRE). To implement FAIRE data management, we propose a Silicon Organism Model-Engrossed Data Integration approach (SOMEDAI), where data are not just stored in separate databases, but are integrated into a mathematical model. The latter allows for validating data consistency and using the data to reconstruct the emergent behaviour of the whole system.

Keywords: systems biology; emergence; state dependent component properties; complex adaptive systems; FAIR data management; Silicon Organism Model-Engrossed Data Integration (SOMEDAI); SIER models; disease maps; COVID-19

To cite this article:

Kolodkin A. N. Making data Findable, Accessible, Interoperable, Reusable and Emergeable (FAIRE): Proposing a Silicon Organism Model-Engrossed Data Integration (SOMEDAI) approach in Systems Biological preparations for the next epidemics. *Advanced targets in Biomedicine*. 2025;1(1):51—64. https://10.0.23.34/ATB-2025-1-1-4

Received
17.04.2025
Revised
02.06.2025
Accepted
05.06.2025
Published
15.07.2025

Методологическая статья

КАК СДЕЛАТЬ ДАННЫЕ НАХОДИМЫМИ, ДОСТУПНЫМИ, ВЗАИМНО СОВМЕСТИМЫМИ, ПЕРЕИСПОЛЬЗУЕМЫМИ, А ТАКЖЕ ЭМЕРДЖЕНТНЫМИ (FAIRE): ПРЕДЛОЖЕНИЯ ПО ИСПОЛЬЗОВАНИЮ МОДЕЛЬНО-ОРИЕНТИРОВАННОЙ ИНТЕГРАЦИИ ДАННЫХ (SOMEDAI) В СИСТЕМНО-БИОЛОГИЧЕСКОЙ ПОДГОТОВКЕ К СЛЕДУЮЩИМ ЭПИДЕМИЯМ

А. Н. Колодкин[1, 2]*

[1] Центр Компетенций Методологии и Статистики (CCMS), Люксембургский Институт Здоровья (LIH), Операционный Хаб Трансляционной Медицины, Л-1445, Люксембург, Штрассе
[2] Люксембургский Центр Системной Биомедицины (LCSB), Университет Люксембурга, Л-4367, Люксембург, Бельво

* Автор-корреспондент:: alexeykolodkin@gmail.com

Резюме: Настоящая работа посвящена обсуждению концепции эмерджентности в современной системной биологии и затрагивает применение этой концепции в эпидемиологическом моделировании COVID-19, где можно выделить три направления эмерджентности: (i) эмерджентность эпидемии в популяции, возникающая из взаимодействий между неинфицированными, инфицированными, и переболевшими индивидуумами; (ii) эмерджентность физиологического ответа на инфекцию на уровне инфицированного организма, возникающая из взаимодействий между биомолекулами, и (iii) эмерджентность самого научно-исследовательского проекта возникающая из взаимодействий между различными институтами и частниками проекта.

Системно-биологическое понимание биологической эмерджентности вызывает необходимость некоторых изменений в подходах по дата менеджменту. С добавление концепции эмерджентости, классический подход так называемого FAIR дата менеджмента (аббревиация от Findable, Accessible, Interoperable, Reusable), трансформируется в FAIRE (Findable, Accessible, Interoperable, Reusable, Emergeable), то есть данные становятся не только легко находимыми, доступными, взаимно совместимыми, но также эмерджентными. Для реализации FAIRE дата менеджмента предлагается подход модельно-ориентированной интеграции данных (SOMEDAI), при котором данные не просто сохраняются в базах данных, но сразу же интегрируются в математическую модель, что позволяет одновременно верифицировать качество данных и реконструировать биологическую эмерджентность.

Ключевые слова: системная биология, эмерджентность, свойства компонентов, зависящие от состояния всей системы. комплексные адаптивные системы, FAIR данные, модельно-ориентированная интеграция данных, SIER модели, карты болезней, COVID-19

Для цитирования:

Колодкин А. Н. Как сделать данные находимыми, доступными, взаимно совместимыми, переиспользуемыми, а также эмерджентными (FAIRE): Предложения по использованию модельно-ориентированной интеграции данных (SOMEDAI) в системно-биологической подготовке к следующим эпидемиям. *Современные направления в биомедицине*. 2025;1(1):51—64. https://doi.org/10.5922/ATB-2025-1-1-4

Поступила
17.04.2025 г.
Прошла рецензирование
02.06.2025 г.
Принята к печати
05.06.2025 г.
Опубликована
15.07.2025 г.

Introduction: Emergence, Design and State Dependency of Component Properties in Computer Models

Systems Biology aims to understand how biological function, which is absent from molecules in isolation, emerges when they become components of a system [1; 2]. The philosophical concept of "emergence" is central here. An emergent property can be defined as a property of a system that satisfies three theses about emergence: (i) the thesis of physical monism, (ii) the thesis of synchronous determinism, and (iii) the thesis or notion of being a systemic (organizational) property [3]. If all three theses are satisfied simultaneously, the property may be called an emergent property. The thesis of physical monism restricts the nature of the system's elements, stating that the system consists only of physical entities. The thesis of synchronous determinism restricts the way systemic properties and the system's microstructure are related, stating that there can be no difference in systemic properties without changes in the structure of the system or in the properties of the components: properties of interest (e. g., metabolite concentrations) are underpinned by changes in the system (e. g., changes in component concentrations, reaction rates). The thesis or notion of being a systemic (organizational) property means that a property is not exhibited by elements in isolation.

It has become traditional to distinguish between weak emergence and strong emergence, depending on whether the specific behaviour of the system's components derives from, or does not derive from, the components' behaviour in simpler configurations [3]. This distinction has a deep intuitive background. Let us consider the classic example: a piece of diamond is hard, and a piece of graphite is soft, because atoms of carbon are arranged differently in these two systems; the emergent property (hardness or softness) depends on the interactions between the system's components. If we assume a simpler configuration of the system, e. g., a small piece of diamond, we can deduce how the atoms of carbon should be arranged to give rise to hardness. In other words, systemic properties of a large piece of diamond can be deduced from the components' behaviour in simpler configurations. Consequently, the emergence would be classified as weak emergence. Now let us consider a living cell. Obviously, if we cut this cell into a hundred small pieces, each piece would be dead in isolation. From here, we can intuitively conclude that the property of being alive is a strongly emergent property.

However, we may also take a look at the issue from a different angle. The properties of carbon atoms in a diamond are the same in both larger and smaller pieces; component properties do not depend on the state of the system (e. g., the size and geometrical shape of the diamond). By contrast, the properties of macromolecules in the cell — e. g., how they interact with each other — do depend on the state of the cell. For instance, assume we alter the concentration of just one component in the cell, such as an enzyme. This will change the concentrations of that enzyme's substrates and products. It is very likely that one of these substrates or products will serve as an activator, inhibitor, or substrate for another protein. This second protein may, in turn, be a transcription factor for a different enzyme or an activator of a third protein, and the effect of the initial change will propagate until, eventually, the properties (behaviour, concentrations, activity) of all components in the cell are altered [4; 5]. Components are fit to the system as a whole; their properties depend on the presence of other components, on boundary conditions, and on the initial conditions of the cell [6]. The component properties of pieces of diamond are not state-dependent. On the other hand, the component properties of parts of a cell are, to a large extent, state-dependent. If we imagine a small piece of a cell with the same composition of amino acids, lipids, ribonucleotides, and other molecules as the intact cell, some state-dependent information about molecular interactions — e. g., the purposeful ar-

rangement of these molecules as it existed in the whole cell — would be missing. The reconstruction of the emergent property of the whole cell would require information about these state-dependent properties of molecules. The emergence in the diamond differs from that occurring in the cell in the degree to which component properties in these two systems are state-dependent.

There is no emergence that is purely weak or purely strong; rather, emergence can be stronger or weaker, depending on how much we need to know about the state-dependency of the component properties — e. g., hysteresis, boundary conditions, flux of mass and energy through the system, and the number of other components affecting a given component property. The stronger the emergence, the more information about state-dependency is required. This also implies that, even if some properties might seem very strongly emergent, their emergence can still be reconstructed from the knowledge of component properties — we simply need to consider more information about their state-dependency [4; 6].

This state-dependency of system component properties can be described in terms of mathematical equations. It is, in fact, a central feature of systems biological models (Fig. 1). It makes the systems biology approach both reductionist and holistic at the same time. It is reductionist because the emergent behaviour is explained in the model in terms of system components, and it is holistic because the reconstruction of emergent behaviour requires information about the system as a whole [4; 6].

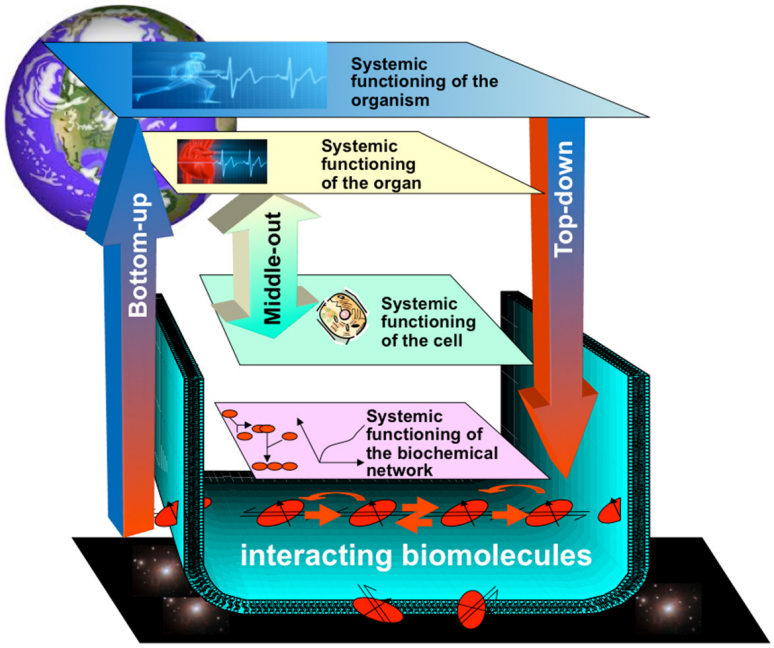


Fig. 1. Reconstruction of emergent behavior in a systems biological model.
Systemic functioning emerges from interactions between biomolecules.
There are three approaches to link the layer of interacting biomolecules to systemic functioning: the bottom-up approach begins with experimentally characterized interactions between biomolecules; the top-down approach starts with systemic functioning; and the middle-out approach begins with an entity of intermediate complexity, such as a cell, and extends both up to organismal functioning and down to interacting biomolecules

The reconstruction of emergent behaviour in a computer model from the knowledge of state-dependent component properties is closely linked to the concept of systems design. By definition, design is "the constellation of system components, their specific properties, and their pattern of interactions that together determine the inte-

grated behaviour of the system" [7]. This understanding — that the arrangement of system components is responsible for a generic set of functions — can be formalized as a more general concept of a design principle.

A design explanation builds a bridge between the underlying mechanism of interacting components and a specific emergent property associated with the design. In this respect, the design explanation resembles a causal explanation, but it also adds a new dimension by requiring the emergent property to be functional — i. e., to serve a specific role for the organism — and by addressing the question: "Why does the system consist of the parts it does, and why are those parts organized in this particular way?" [8]. Design implies a purposeful arrangement of parts [9]. In a design study, we assume that the system is structured to give rise to the emergence of a useful biological function.

Going further, according to Anokhin's theory of functional systems (TFS), "a system can be called a complex of selectively involved components where all interactions are presented by mutual cooperation of components focused on obtaining a useful result" [10]. The idea that a goal precedes its realization by the organism was also recognized in Kant's doctrine on the end of all things, illustrated in his example of the flute: "If Peter carves a flute, then his will (or his movement of his hands in a certain way) causes the flute (O), but his will (his movement of his hands) is determined by Peter's idea of the flute (S)" [11]. Interestingly, this topic has also found its way into design studies of artificial systems, and is discussed by various authors — sometimes in direct analogy with the ancient "noema" concept of Plato (428/427 BC) [12]. According to this concept, a future noema (not yet realised) becomes a current noema in the process of noesis (realisation through interaction with the environment). Since the current noema will differ from the originally planned future noema, the future noema must be readjusted, and the loop continues. The theory of functional systems does not use the term noema, but it effectively implies noesis when discussing reverse afferentation.

To summarize, when building a systems biological model, we imply design of the system, and reconstruct emergent behavior *in silico* by incorporating information about the state-dependency of component properties into mathematical equations. The stronger the emergence, the more information about the state-dependency of system component properties we need to include in our model. Let us explore this in the context of COVID-19 studies, where, in order to learn from the pandemic and prepare for potential future epidemics, we must consider three dimensions of emergence.

Three Dimensions of Emergence in COVID-19 Research

First Dimension of Emergence: Dynamic Model of Virus Spread at the Population Level

With the outbreak of the COVID-19 pandemic, the initial focus of research was to build a predictive model of how SARS-CoV-2 spreads within the human population. This was modeled using the SIR (Susceptible, Infectious, and Recovered) framework, where the spread of the virus emerged from interactions between susceptible, infected, and recovered individuals. One of the earliest COVID-19 SIR models, developed by Westerhoff & Kolodkin [13], considered 14 components, with transitions between them described by a mass-action law (Fig. 2). The model, available at FAIRDOMHub (https://fairdomhub.org/models/693), enabled predictions of various measures' effects on the epidemic and allowed for comparisons of different lockdown-release strategies across countries. Although the model proved useful, its predictive power was limited by the simplistic mass-action reactions used to describe transitions between system components. This limitation was partially addressed in a delay-diffe-

rential SEIR model for improved infection dynamics, developed by Kiselev, Akberdin, and Kolpakov [14]. In this model, Susceptible, Exposed, Infected, and Recovered (SEIR) individuals were treated as model components, with transitions modeled using more sophisticated mathematical equations that incorporated time delays and instantaneous transitions. Furthermore, the model accounted for different population groups, symptom progression from mild to critical, vaccination, the duration of protective immunity, and new virus strains.

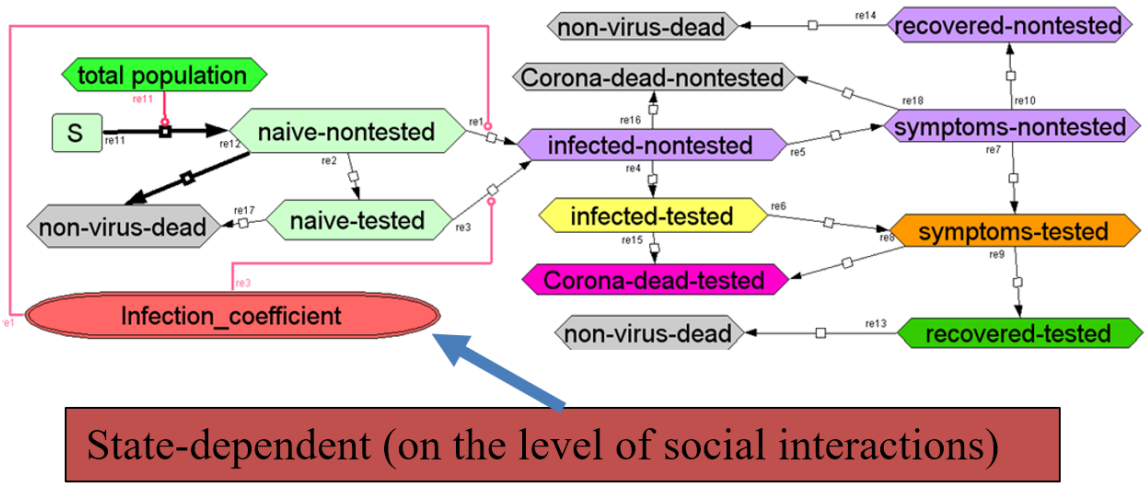


Fig. 2. Systems biology model of the COVID-19 epidemic.
Species are represented in boxes, and reactions
(irreversible, mass-action) are shown as arrows.
The infection coefficient is modeled as a linear function of the numbers
of infected-nontested, infected-tested, symptoms-nontested,
and symptoms-tested individuals.
It is also state-dependent on the level of social interactions.
The full model, including all parameter values, is available at
https://doi.org/10.15490/fairdomhub.1.model.693.1
(Adapted from Westerhoff & Kolodkin [13])

Still, even more sophisticated mathematical equations describing transitions, such as between Exposed and Infected, or Infected and Recovered, remain too simplified and tend to underestimate the strength of emergence. For instance, the transition from Exposed to Infected emerges from multiple interactions between the human organism and the virus, underpinned by numerous physiological processes within the organism that are highly state-dependent on individual genetic and environmental factors. This represents another dimension of emergence.

Second Dimension of Emergence: Virus-induced Physiological Behaviour

The response of the organism to the virus and the progression of COVID-19 at the individual level emerges from interactions between the virus and a range of physiological processes. For instance, SARS-CoV-2 affects multiple molecular networks and triggers several pathophysiological mechanisms. Viral RNA and proteins have been found in tyrosine hydroxylase-immunoreactive neurons in the substantia nigra of deceased COVID-19 patients. The presence of the virus was associated with microgliosis and CNS platelet-enriched microthrombi in small parenchymal vessels [15]. There are multiple interconnected mechanisms through which the virus may induce both microgliosis and thrombosis, such as via NF-kB signaling, which could, in

turn, be linked to oxidative stress and mitochondrial dysfunction [16]. In attempting to reconstruct, *in silico*, the influence of SARS-CoV-2 on the organism, we must account for increasingly complex molecular mechanisms. Even more intracellular networks will need to be considered when developing COVID-19 precision medicine, such as delivering specific drugs to infected cells [17]. Ultimately, we may need to extend the model to encompass a whole-body silicon cell model.

The COVID-19 Disease Map initiative, organized by the University of Luxembourg, aimed to integrate and continuously update all available information on the mechanisms by which SARS-CoV-2 interacts with the human organism. The initiative involved more than 230 scientists from diverse backgrounds, representing 120 institutions across 30 countries. As a result of the project, a knowledge repository of the molecular mechanisms underlying COVID-19 was created — approaching the complexity of the whole body (https://covid.pages.uni.lu/).

However, only several fragments of the knowledge integrated within the COVID-19 Disease Map were ultimately transformed into mathematical models. One successful example is the flux balance model by Renz, Widerspick, and Dräger [18]. The COVID-19 Disease Map proved useful [19] and helped, to some extent, reconstruct the emergence of diseased behaviour at the level of the human organism. Still, the emergence was represented as weaker than it actually is, and much of the state-dependency of component properties was missing. It is crucial that both the emergence of individual, patient-specific responses to the virus and the emergence at the epidemiological level (e. g. modelling how various virus strains spread through society) should be integrated into a single model.

Third Dimension of Emergence: Emergence of Research Projects

There is also a third, explicitly social, dimension of emergence, where interactions between various participants and institutions — each collecting different pieces of information — should lead to the emergence of an integrated systems biological understanding. In principle, this would involve merging the modelling of epidemics with the modelling of individual physiological responses to a virus. On this layer of emergence, the system's components are researchers, data managers, medical doctors, nurses, research groups and institutions, hospitals, and others. They interact within the framework of a project, and various project outputs — such as insights, scientific concepts, and proposed medical treatments — emerge. Once again, the concept of design is applicable: the goal serves as the system's formative factor, and the system is designed for useful emergence — e. g., new scientific understanding. However, this useful output (the project goal) evolves continuously throughout the project (a continuous adjustment of the future noema), making the project itself a Complex Adaptive System.

An example of how the goal — as the system's forming factor — may adapt during a project is found in research on the long-term consequences of SARS-CoV-2 infection. Initially, researchers would define "Long COVID", a consequence of COVID-19, in varying ways. In some cases, it was simply defined as the presence of any symptom persisting 12 months after infection [20]. In others, more specific, model-driven definitions were proposed. For instance, in one mathematical model of the immune response to SARS-CoV-2, Long COVID corresponded to a stable stationary or periodic solution that persisted in the system after the acute infection phase [21].

Subsequently, the World Health Organization (WHO) unified the understanding of Long COVID by providing the following definition: "the continuation or develop-

ment of new symptoms 3 months after the initial SARS-CoV-2 infection, with these symptoms lasting for at least 2 months with no other explanation" (www.who.int/europe/news-room/fact-sheets/item/post-covid-19-condition).

This WHO definition came into conflict with ongoing research projects in which data collection schedules had already been established based on different timelines. For example, the Luxembourg-based CON-VINCE study (COVID-19 National survey for assessing VIral spread by Non-affected CarriErs) [22], later followed by ORCHESTRA project, an EU Horizon 2020 initiative (https://orchestra-cohort.eu/), which aimed to standardize and harmonize COVID-19-related data collection methodologies across multiple participating institutions to ensure consistency and interoperability in data sharing among researchers and public health authorities throughout Europe. Its scope encompassed epidemiological modeling, vaccine development, therapeutic interventions, and the evaluation of public health strategies. The project also served as a platform for rapid data analysis and dissemination of findings, supporting informed decision-making and public health responses during the COVID-19 pandemic [22].

The CON-VINCE — ORCHESTRA data collection began as early as March 2020, at the very start of the pandemic, with the observation of non-infected participants. One fraction of participants remained non-infected and formed the control group. Another fraction became infected at some point during the study, thus providing samples both before and after infection. However, the project had been launched before the WHO Long COVID definition was published. As more knowledge was becoming available (e. g., the WHO definition of Long COVID), the project settings — such as sampling intervals, test types (blood, stool, saliva, and SWAB tests), and even the questionnaires — were adjusted accordingly. Initially, when asked about their symptoms, the participants would answer questions, like: "Did you have any symptoms until today?" Later, this question was modified to: "Did you have any symptoms during the previous week?" or "Have you been having continuous symptoms since the last visit?" These changes, while small, imposed limitations on the interpretation of Long COVID data [22—24]. In addition, some symptoms — such as loss of smell — which were initially not monitored, later turned out to be central in Long COVID studies [22—24].

Many challenges in project setup were unavoidable, given that both the spread of infection in the population (the first dimension of emergence) and the exact mechanisms by which the virus affects the organism on an individual level (the second dimension of emergence) were unknown at the outset. If a whole-body computer replica of the human organism (silicon / virtual / digital) had existed, the data collection and analysis processes could have been organized differently. The COVID-19 disease map could have been projected onto this whole-body model, which could then be parameterized for each individual patient (digital twin).

Live Data: Findable, Accessible, Interoperable, and Reusable (FAIR), and also Emergeable (FAIRE) Data Management

We often think of data as something given to us. The very word 'data' in all Indo-European languages means 'given,' originating from the Sanskrit ददाति [dadAti] and दत्ते [datte], which literally translate as 'give' or 'given.' This evolved into various forms across languages: дать [dat'] (Russian), देना [dena] (Hindi), dare (Latin and Italian), dar (Spanish), donner (French), donate (English), and so on.

It can be argued, however, that data is something that we take in depending on the way our perception is set up. We might see water as a liquid with particular physical properties, as a molecular formula (H_2O), or as a system of elementary particles. With

systems biological modelling, we often treat the properties of system components as given, fixed inputs and focus on quantifying data; on making data findable, accessible, interoperable, and reusable (FAIR). However, these component properties are actually state-dependent — and in many cases, emergeable, since they *emerge* from interactions between other components in the system. Linking one component's property to others and reconstructing the system's overall emergence would, in effect, bring Emergence into the FAIR principles. In this sense, FAIR becomes FAIRE — to borrow the French verb *faire*, meaning "to do" or "to make." FAIRE is not just about organizing data for access, but about integrating them into a model where one piece of data can be derived from others. In the words of Ernest Rutherford, who once said that "all science is either physics or stamp collecting," we move from "stamp collecting" to "physics" in biology — computing emergent properties from known components, and linking them together in one equation. What sets biology apart, however, is that we do not rely on a handful of universal equations. Instead, we reconstruct emergence through a more complex mathematical model that accounts for higher levels of state dependency.

The next question is how to organize this emergence-oriented FAIRE data management. We propose an approach called Silicon Organism Model-Engrossed DAta Integration, abbreviated as SOMEDAI (Fig. 3). We use the term "model-engrossed data integration" to emphasize that a mathematical model, rather than a traditional database, is used to store data — and not only to store it, but also to integrate different data points by linking them within the model itself. The "Silicon Organism Model" refers to a widely accepted concept of whole-body mathematical models in which emergent behaviour is reconstructed *in silico* through the interactions of system components (biomolecules).

The first step in the SOMEDAI approach is blueprint modelling of the organism. Instead of modelling individual organisms or addressing isolated research questions, we propose creating a single, generic blueprint model with various "version control" instantiations. In fact, all cells and organisms are remarkably similar: the basic building blocks (e. g., amino acids) and metabolites (ATP, glucose, pyruvate, etc.) are largely the same, and the biochemical reactions between them follow similar patterns. All living cells have very much the same processes coded by the same genes, with differences arising mainly from which genes are switched on or off during cell differentiation. Therefore, we can conceptualize a blueprint organism model as a kind of master copy that encompasses all potential instantiations. We may build such a blueprint model where small changes in gene expression bring organism-specific, or tissue-specific, or condition-specific, or disease-specific, or other specialized versions of the model.

There are two main strategies for building a blueprint model. One is a bottom-up approach: starting with knowledge-based, mechanistic interactions between components — e. g. biomolecules such as genes, RNA, proteins, and metabolites — and then generating organism-, cell-, tissue-, or other context-specific instantiations. Measured parameters are inserted into these instantiations to observe emergent behaviour, which can then be compared with experimental observations from specific cells or organisms. The other strategy is a top-down approach, where the network of interactions between components is inferred by observing the behaviour of the entire system in response to perturbations. A third option is a middle-out strategy, which integrates fragmentary knowledge to model the behaviour of a single organ or functional system. This is done by focusing on interactions between entities at lower, but not necessarily molecular, levels of organization (Fig. 1).

Promisingly, the outputs of all three approaches can be compared — with each other and with experimental observations (Fig. 3). Step by step, both top-down

and bottom-up models should converge with experimentally measured behaviour. Estimating this convergence will form a kind of controller for data integration — namely, the model itself — enabling the reconstruction of all three dimensions of emergence discussed earlier.

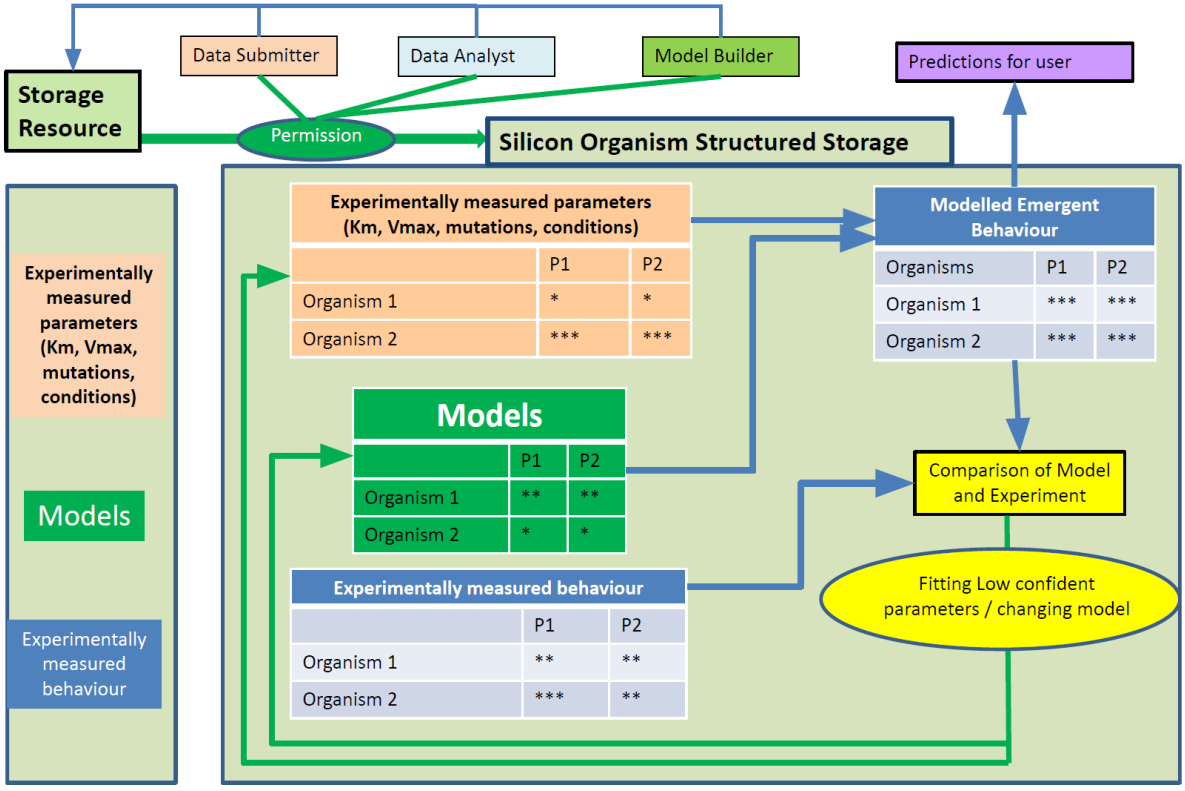


Fig. 3. Silicon Organism Model-Engrossed Data Integration (SOMEDAI). Some data consist of specific experimentally measured parameters, such as physical properties of certain enzymes (e. g., Vmax, Km). Other data capture experimentally observed emergent behaviour in response to perturbations. These parameters and emergent behaviours can be integrated into a mathematical model. The model, based on state-dependent component properties, aims to reconstruct the emergent behaviour *in silico*. The simulated behaviour can then be compared with the corresponding experimental data. Parameters that have not yet been measured, or those measured with low confidence, can be estimated (fitted). This creates a feedback loop, improving the model with each new data input

After model-driven integration of separate data pieces into the "right places," data will become not only FAIR (Findable, Accessible, Interoperable, and Reusable), but also *emergeable* and "intelligent," at least in some aspects of intelligence [25]. For example, the integrated system will show robust adaptation, such that all data together form a homeostatic system, and each new "small piece" of data will be adjusted or rejected based on the integrity of the larger model encompassing all datasets as a whole. Thus, the system will become capable of proposing the next set of experiments and guiding researchers — in a way that humans may eventually lose their grasp on the research process and delegate the "understanding" to more advanced integrated systems, such as a robot scientist (https://en.wikipedia.org/wiki/Robot_Scientist).

This integrated "intelligent" model will make data *live* — in the sense that the knowledge embedded in the data in the data about component properties of the sys-

tem and about the behaviour of the system will continuously adjust in response to the system's evolution. In the study of epidemics, this will enable the integration of all three dimensions of emergence: the emergence of an epidemic from interactions between individuals in a population; the emergence of individual physiological responses to the virus from intracellular mechanisms; and the emergence of the project aimed at understanding and controlling the epidemic.

Concluding Remarks

We have discussed how addressing unknown epidemics depends on successfully integrating three dimensions of emergence: emergence at the population level, emergence at the level of organism-specific response to the virus, and emergence of project organization. We have proposed the Silicon Organism Model-Engrossed Data Integration (SOMEDAI) system as a way to embrace the high state dependency of component properties and, thereby, to integrate all three dimensions of emergence — making data collection and analysis both live and intelligent.

A system like SOMEDAI cannot be implemented at the level of a single researcher, research group, or even several "self-organizing" groups. Instead, it requires a designed infrastructure that enables the consolidation of professional efforts, and the unification and standardization of various research components (e. g., modeling and data collection) for better integration into a unique system.

In reality, the first steps toward SOMEDAI data integration have already been taken at the level of research infrastructures. One example is the pioneering work of Make Me My Model (M4) by Infrastructure Systems Biology Europe [26; 27], within the EU Horizon 2020 project, Coordinated Research Infrastructures Building Enduring Life-science Services (CORBEL) (www.corbel-project.eu), which was further developed into the European Open Science Cloud for Life Sciences (EOSC-Life) (www.eosc-life.eu).

The starting point was the JWS Online model repository (https://jjj.bio.vu.nl/), where pre-built models — essentially small blueprint models — could be run online using either default parameters curated and approved by the JWS team or parameters chosen by the user based on their own experimental data. Users would often manually select parameters from curated databases, such as SABIO-RK (www.sabio.h-its.org), so the first step was to automate parameter selection. Now, users can access the database, choose the appropriate organism and conditions, click a button, and all selected data are automatically retrieved for simulations on JWS Online. This is just the first small building block of the SOMEDAI system. The next step would be to create a universal blueprint model linked with organism- or cell-specific instantiations. After fitting a cell-specific model to data and comparing the emergent behavior between the model and experiment, the feedback should be used to adjust the blueprint whole-body model.

At the same time, it should be acknowledged that relatively small projects like CORBEL or EOSC-Life are not equipped to fully implement the SOMEDAI approach. The strength of emergence that needs to be reconstructed in a blueprint model is substantial and requires the emergence of an infrastructure with the appropriate complexity — potentially comparable to the scale of Space programs.

Funding. This study received no external funding.

Conflict of Interest. The author declares no conflict of interest.

Финансирование. Это исследование не получало внешнего финансирования.

Конфликт интересов. Автор заявляет об отсутствии конфликта интересов.

References

1. Alberghina L., Westerhoff H. V. Systems Biology: Definitions and Perspectives. Springer: Berlin, Germany. 2005.

2. Westerhoff H. V., Kolodkin A. N., Conradie R., Wilkinson SJ., Bruggeman F. J., Krab K., et al. Systems biology towards life in silico: mathematics of the control of living cells. J. Math. Biol. 2009, 58(1-2), 7—34, DOI: 10.1007/s00285-008-0160-8.

3. Stephan A. The dual role of 'emergence' in the philosophy of mind and in cognitive science. Synthese. 2006, 151(3), 485—498, DOI: 10.1007/s11229-006-9019-y.

4. Kolodkin A. N., Simeonidis E., Balling R., Westerhoff H. V. Computing life: Add logos to biology and bios to physics. Progress in Biophysics & Molecular Biology. 2012, 111(2-3), 69—74, DOI: 10.1016/j.pbiomolbio.2012.10.003.

5. Kolodkin A. N., Bruggeman F. J., Plant N., Mone M. J., Bakker B. M., Campbell M. J., et al. Design principles of nuclear receptor signaling: how complex networking improves signal transduction. Mol. Syst. Biol. 2010, 6, 446, DOI: 10.1038/msb.2010.102.

6. Kolodkin A. N., Simeonidis E., Balling R., Westerhoff H. V. Understanding complexity in neurodegenerative diseases: in silico reconstruction of emergence. Frontiers in Systems Physiology. 2012, 3, 291, DOI: 10.3389/fphys.2012.00291.

7. Wall M. E., Hlavacek W. S., Savageau M. A. Design of gene circuits: Lessons from bacteria. Nature Reviews Genetics. 2004, 5(1), 34—42, DOI: 10.1038/nrg1244.

8. Wouters A. G. Design explanation: Determining the constraints on what can be alive. Erkenntnis. 2007, 67(1), 65—80, DOI: 10.1007/s10670-007-9045-2.

9. Behe M. J. Darwin's black box : the biochemical challenge to evolution. Free Press: New York. 2006, ISBN 0684834936.

10. Анохин П. К. Очерки по физиологии функциональных систем. Медицина: Москва. 1975, ISBN 978-5-458-38439-11975.

11. Heidemann D. (Ed.). Teleology. Kant Yearbook. 2009.

12. Лосев А. Ф. Собрание сочинений. Том 1. Бытие. Имя. Космос. Издательство «Мысль», Российский открытый университет: Москва. 1993, ISBN 5-244-00717-3.

13. Westerhoff, H.V., Kolodkin, A. N. Advice from a systems-biology model of the corona epidemics. NPJ Syst. Biol. Appl. 2020, 6, 18, DOI: 10.1038/s41540-020-0138-8.

14. Kiselev I. N., Akberdin I. R., Kolpakov F. A. Delay-differential SEIR modeling for improved modelling of infection dynamics. Sci. Rep. 2023, 13(1), 13439, DOI: 10.1038/s41598-023-40008-9.

15. Emmi A., Rizzo S., Barzon L., Sandre M., Carturan E., Sinigaglia A., et al. Detection of SARS-CoV-2 viral proteins and genomic sequences in human brainstem nuclei. NPJ Parkinsons Dis. 2023, 9(1), 25, DOI: 10.1038/s41531-023-00467-3

16. Kolodkin A. N., Sharma P. R, Colangelo A. M., Ignatenko A., Martorana F., Jennen D., et al. ROS networks: designs, aging, Parkinson's disease and precision therapies. NPJ Syst. Biol. Appl. 2020, 6(1), 34, DOI: 10.1038/s41540-020-00150-w.

17. Zubarev I., Vladimirtsev D., Vorontsova M., Blatov I., Shevchenko K., Zvereva S., et al. Viral Membrane Fusion Proteins and RNA Sorting Mechanisms for the Molecular Delivery by Exosomes. Cells. 2021, 10(11), 3043, DOI: 10.3390/cells10113043.

18. Renz A., Widerspick L., Dräger A. FBA reveals guanylate kinase as a potential target for antiviral therapies against SARS-CoV-2 Click here to expand. Bioinformatics. 2020, 36(2), i813—i821, DOI: 10.1093/bioinformatics/btaa813.

19. Ostaszewski M., Niarakis A., Mazein A., Kuperstein I., Phair R., Orta-Resendiz A.,et al. COVID-19 Disease Map community. COVID-19 Disease Map a computational repository of SARS-CoV-2 virus-host interaction mechanisms. Mol. Syst. Biol. 2021, 17(10), e10387, DOI: 10.15252/msb.202110387.

20. Fischer A., Zhang L., Elbéji A., Wilmes P., Oustric P., Staub T., et al. Long COVID Symptomatology After 12 Months and Its Impact on Quality of Life According to Initial Coronavirus Disease 2019 Disease Severity. Open Forum Infect. Dis. 2022, 9(8), ofac397, DOI: 10.1093/ofid/ofac397.

21. Khristichenko M., Nechepurenko Y., Grebennikov D., Bocharov G., Computation and analysis of stationary and periodic solutions of the COVID-19 infection dynamics model. Journal of Bioinformatics and Computational Biology. 2025, 23(1), 2540001, DOI: 10.1142/S0219720025400013.

22. Tsurkalenko O., Bulaev D., O'Sullivan M.P., Snoeck C., Ghosh S., Kolodkin A., et al. CON-VINCE consortium and the ORCHESTRA working group. Creation of a pandemic memory by tracing COVID-19 infections and immunity in Luxembourg (CON-VINCE). BMC Infect. Dis. 2024, 24(1), 179, DOI: 10.1186/s12879-024-09055-z.

23. Conde P. M., Bulaev D., Rauschenberger A., Ohnmacht J., Fritz J. V., O'Sullivan M. P., et al. Distinguishing Post COVID from Pandemic-Related Health Effects: Insights from a 4-Year Prospective Cohort Study. Journal of Infection, 2025, in press.

24. Novelli S., Reinkemeyer C., Bulaev D., O'Sullivan M.P., Snoeck C., Rauschenberger A., et al. Waning of anti-SARS-CoV-2 antibodies after the first wave of the COVID-19 pandemic in 2020: a 12-month-evaluation in three population-based. PloS One. 2025, 20(5), e0320196, DOI: 10.1371/journal.pone.0320196.

25. Westerhoff H. V., Brooks A. N., Simeonidis E., Garcia-Contreras R., He F., Boogerd F. C., et al. Macromolecular networks and intelligence in microorganisms. Frontiers in microbiology. 2014, 5, 379, DOI: 10.3389/fmicb.2014.00379.

26. Kolodkin A., Alberghina L, Snoep J. L., Piccialli G., Borsani M., Westerhoff H. V. Infrastructure Systems Biology Europe (ISBE): Emergence of Innovative Systems Biology Servicing Abstract book of 4th Dutch Bioinformatics & Systems Biology Conference Congrescentrum De Werelt, Lunteren. 71, DOI: 10.18699/BGRSSB-2018-108.

27. Dos Santos M. V., Anton M., Szomolay B., Ostaszewski M., Arts I., Benfeitas R., et al. Systems Biology in ELIXIR: modelling in the spotlight. F1000Research. 2022, 11(ELIXIR), 1265, DOI: 10.12688/f1000research.126734.1.

The authors

Alexey N. Kolodkin, PhD, clinical data manager, Competence Center for Methodology and Statistics (CCMS), Luxembourg Institute of Health (LIH), Translational Medicine Operations Hub; visiting researcher, Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Luxembourg.

ORCID: 0000-0002-7466-5027

Об авторах

Алексей Николаевич Колодкин, PhD, менеджер клинических данных, Центр компетенций методологии и статистики (CCMS), *Люксембургский институт здоровья* (LIH), Операционный хаб трансляционной медицины; приглашенный исследователь, *Люксембургский центр системной биомедицины* (LCSB), Университет *Люксембурга*, *Люксембург*.

ORCID: 0000-0002-7466-5027