

## Review

# High-quality Reconstruction of Disease Mechanisms: Planning, Development and Maintenance

Alexander Mazein<sup>1\*</sup>, Marcio Luis Acencio<sup>1</sup>, Irina Balaur<sup>1</sup>, Adrien Rougny<sup>2</sup>, Danielle Welter<sup>1</sup>, Anna Niarakis<sup>3,4</sup>, Diana Ramirez Ardila<sup>5</sup>, Ugur Dogrusoz<sup>6</sup>, Piotr Gawron<sup>1</sup>, Venkata Satagopam<sup>1,7</sup>, Wei Gu<sup>1,7</sup>, Andreas Kremer<sup>5</sup>, Reinhard Schneider<sup>1,7</sup>, Marek Ostaszewski<sup>1,7\*</sup>

<sup>1</sup> Luxembourg Centre for Systems Biomedicine, University of Luxembourg, 6 Avenue du Swing, L-4367 Belvaux, Luxembourg.

<sup>2</sup> Independent researcher, Massy, France.

<sup>3</sup> Université Paris-Saclay, Laboratoire Européen de Recherche pour la Polyarthrite rhumatoïde - Genhotel, Univ Evry, Evry, France.

<sup>4</sup> Lifeware Group, Inria Saclay-Ile de France, Palaiseau, France.

<sup>5</sup> ITTM Information Technology for Translational Medicine, Esch-sur-Alzette, Luxembourg.

<sup>6</sup> Computer Engineering Department, Bilkent University, Ankara, Turkey.

<sup>7</sup> ELIXIR Luxembourg, 6 Avenue du Swing, L-4367 Belvaux, Luxembourg.

\* Correspondence: Alexander Mazein (alexander.mazein@uni.lu), Marek Ostaszewski (marek.ostaszewski@uni.lu).

**Abstract:** As a conceptual model of disease mechanisms, a disease map integrates available knowledge and is applied for data interpretation, predictions and hypothesis generation. It is possible to model disease mechanisms on different levels of granularity and adjust the approach to the goals of a particular project. This rich environment together with requirements for high-quality network reconstruction makes it challenging for new curators and groups to be quickly introduced to the development methods. In this review, we offer a step-by-step guide for developing a disease map within its mainstream pipeline that involves using the CellDesigner tool for creating and editing diagrams and the MINERVA Platform for online visualisation and exploration. We also describe how the Neo4j graph database environment can be used for managing and querying efficiently such a resource. For assessing the interoperability and reproducibility we apply FAIR principles.

**Keywords:** Systems Biology; Systems Medicine; Disease Mechanisms; Disease Maps; Network Biology

## Introduction

Disease maps have been introduced in systems biomedicine as conceptual models of disease mechanisms [1, 2]. In a disease map, interconnected pathway diagrams depict events from the level of molecular processes to higher levels like communication between cells and organs [1–3]. Approaches to the construction of disease maps advanced during past years, supported by the activity of the Disease Maps community (<https://disease-maps.org>) and formalised into best practices (<https://disease-maps.org/events>) [4]. However, there is still a lack of clear, formal and well-structured guidelines for main steps of developing a disease map in collaborative projects.

The need for such guidelines became evident during the development of the COVID-19 Disease Map [5], a large-scale project supported by an extensive community with more than 200 volunteer participants. The efforts of COVID-19 Disease Map Community would be greatly advanced with step-by-step guidelines to accelerate learning about the approach and creating high-quality resources.

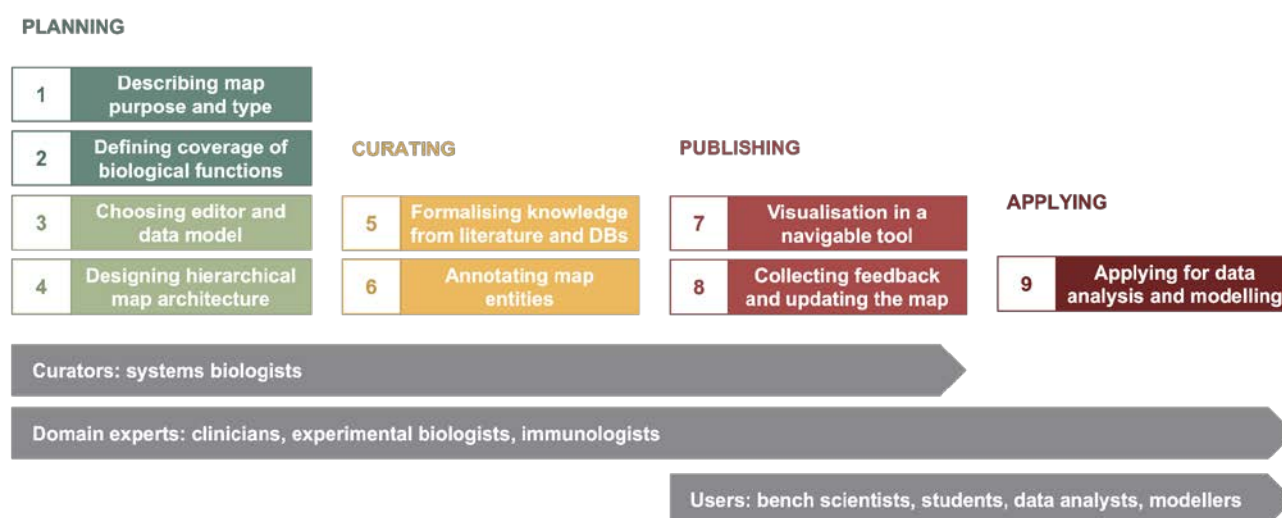
A number of publications can be used as a reference when constructing disease maps [1, 4]. Moreover, brief and informative guideline articles about biocuration [6] and path-

way construction [7] offer valuable introductory material for constructing disease-oriented resources. In particular, an article on construction of a Reactome-based specific knowledge repository [8] proposes instructions on reuse of pathway diagrams in a specific context.

In this article, we review the literature and examples for a particular area of disease map construction, structuring them into steps of a reference workflow. The methods described in this review are built on the experience of the previously developed disease maps (<https://disease-maps.org/publications>) including recently published resources [3, 9–11]. Definitions of domain-specific terms are available at <https://disease-maps.org/glossary>.

### Disease map development workflow

The major phases of map development are illustrated in Figure 1. The **planning phase** includes i) determining map purpose and type, ii) defining the coverage of biological functions, iii) choosing a graphical standard (data model) and an editor, and iv) designing map architecture (Figure 1, Steps 1 - 4). While these steps are shown in a sequence, they are often realised in parallel. Next, in the **curation phase**, available knowledge from literature and databases is encoded into a diagram with accompanying annotation of map entities (Figure 1, Steps 5 and 6). When the map is ready for publication, it is made available online in a navigable tool and updated based on feedback from domain experts – the **publishing phase** (Figure 1, Steps 7 and 8). The **application phase** depends on the goals of the project, expertise and resources available, and can include data visualisation, drug target exploration, computational modelling and hypothesis generation (Figure 1, Step 9).



**Figure 1.** The map development workflow (based on Kondratova et al, 2018 [4]). The workflow includes nine steps in four phases (please see in the text). Curators and domain experts are involved in planning and designing the map, then curators develop the map using a diagram editor of choice, advised by domain experts. After publishing online, data analysts, modellers and other users have access to the map for exploration and applications in translational projects, as well as domain experts suggest further evolving the map.

#### 1. Determining map purpose and type.

At the initial step, it is important to define goals of the project and to determine the map type and its applications. Later, it will influence map's format, design, size and coverage. The key questions at this stage are:

- What is the main focus? Is it one disease or multiple related disorders?
- What is the intended use of the map?

- What resources are available or needed?
- What is the timeline?
- Which domain experts (clinicians and biologists) can advise and revise the curation?

Examples of motivation, objectives and primary users for several disease map projects are shown in Table 1. Often they are aligned with the focus of a research centre so the developed map supports related research. It also can be part of a large-scale translational project for integrating knowledge and future applications in interpreting data generated by the project.

**Table 1.** Examples of disease map purpose, intended use and expected primary users.

Primary purpose	Details	Examples	Primary users
Integrating knowledge	Visualisation, identifying gaps, educational purposes	Atlas of Cancer Signalling Networks [12], Parkinson's Disease Map [13], RA-Map [14]	Research centres, students, modellers
Combining knowledge with external databases and tools	Exploring drug targets, text mining	COVID-19 Disease Map [5]	COVID-19 research community, data analysts
Data analysis	Network-based analysis, omics and multi-omics data visualisation, interpretation and hypothesis generation	AsthmaMap [3]	IMI U-BIOPRED (No 115010, <a href="https://www.imi.europa.eu">https://www.imi.europa.eu</a> ) community, IMI eTRIKS (No 115446, <a href="https://www.etriks.org">https://www.etriks.org</a> ) community, data analysts
Computational modelling	Using maps as templates for building computational models, refining hypotheses and making predictions	Atherosclerosis Map [15]	Research centres, modellers
Computational modelling combined with data analysis	Using maps as templates for building Boolean and hybrid modelling and data analysis	RA-Atlas [10, 16, 17]	Research centres, modellers, data analysts

Intended map purpose might include one or several of the following objectives:

- integrating available knowledge:
  - a. reviewing and visualising disease mechanisms,
  - b. identifying gaps in knowledge,
  - c. educational purposes;
- combining knowledge with external databases and tools:
  - a. exploring drug targets,
  - b. text mining;
- analysis:
  - a. network-based analysis,
  - b. omics and multi-omics data visualisation, interpretation and hypothesis generation;
- computational modelling:
  - a. using maps as the basis/template for different types of models,
  - b. refining hypothesis,
  - c. making predictions.

Clarifying the motivation, objectives and intended applications is needed for the following steps about the coverage of biological functions. This can be documented in a brief project description document (<https://disease-maps.org/template>), or in a web page (e.g. <https://disease-maps.org/projects>), or in a project management platform such as FAIR-DOMHub (<https://fairdomhub.org>).

## 2. Defining coverage of biological functions

Exchange with the disease domain experts allows to define the scope and boundaries of the map. Also, key reviews and figures that can help to create an initial skeleton of the map. The main tasks at this stage include:

- Identify priority pathways;
- Identify cell types, tissues and organs relevant to the pathology mechanisms;
- List the molecules known to be involved and drugs used in a given condition
- Define disease outcomes on the molecular, cellular and tissue levels, including subdivision into disease stages and subtypes, if applicable.

Such lists can be stored in a document shared and discussed with the domain experts, and maintained throughout the map development, e.g. <https://disease-maps.org/template>.

Often this information is available in relevant databases, like disease-focused pathways of Reactome (Reactome:R-HSA-1643685) including cancers (Reactome:R-HSA-5663202) and infectious diseases (Reactome:R-HSA-5663205, Reactome:R-HSA-9694516). Disease pathways of KEGG (<https://www.genome.jp/kegg/disease>), include Parkinson's disease (KEGG:hsa05012) and COVID-19 (KEGG:hsa05171). The Open Targets Platform (<https://platform.opentargets.org>) [18] lists molecules relevant for various diseases, e.g. targets associated with allergic asthma ([https://platform.opentargets.org/disease/MONDO\\_0004784/associations](https://platform.opentargets.org/disease/MONDO_0004784/associations)).

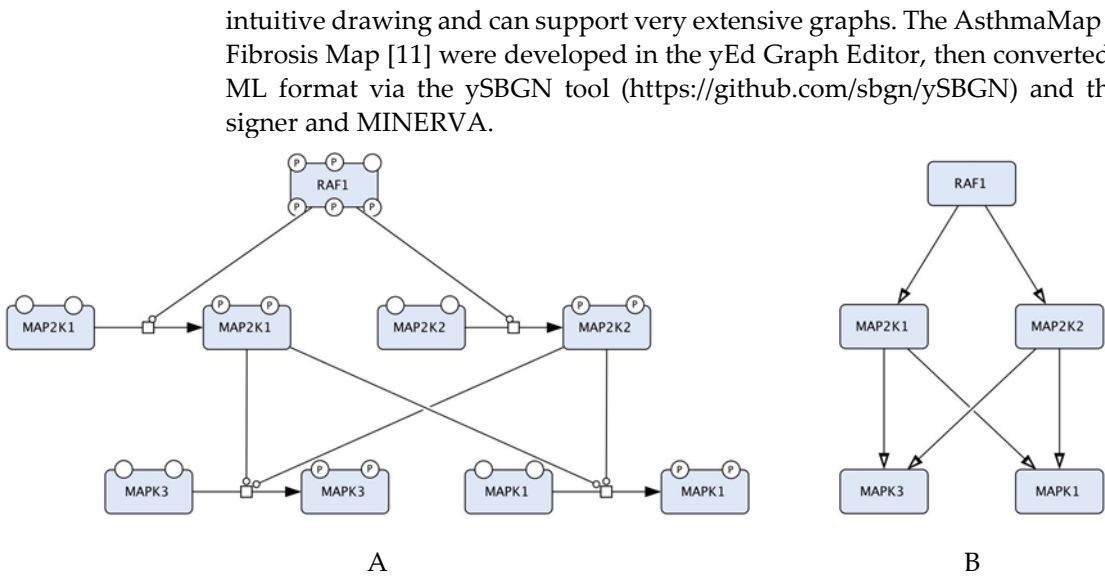
Outcomes of this step guide the curation efforts in steps 5 and 6 (Figure 1), and the coverage of the biological functions may change depending on available knowledge. In effect, this step is often revisited, so it is important to maintain versioning of the map planning document (<https://disease-maps.org/template>).

## 3. Choosing a graphical standard and an editor

Majority of disease maps are created in Systems Biology Graphical Notation (SBGN) [19] using the CellDesigner editor (<https://www.celldesigner.org>), using the SBGN Process Description (PD) and Activity Flow (AF) languages [19] (Figure 2). The Entity Relationship (ER) language [20] is not currently employed in developing disease maps.

Process Description (PD) [21] and Activity Flow (AF) [22] are the two most used languages of the Systems Biology Graphical Notation (SBGN) standard ([www.sbgn.org](http://www.sbgn.org)) [19] (Figure 2). More information about the SBGN standard are available in the "SBGN Learning" materials at <https://sbgn.github.io/learning> and the "Figure to SBGN" page at <https://sbgn.github.io/figuretosbgn>.

CellDesigner (<https://www.celldesigner.org>) is a diagram editor that is frequently used for developing disease maps. It supports extensive diagrams and follows SBGN logic and makes it possible to draw SBGN-compatible diagrams in Process Description and Activity Flow type of languages using the default CellDesigner's palette or its Reduced Notation palette. Alternative solutions include the open-source web-based Newt Editor [23], the open-source VANTED editor with its SBGN-ED add-on [24] and the freely available yEd Graph Editor from yWorks (<https://www.yworks.com/products/yed>). Each tool has its own advantages. Newt offers support for the latest 0.3 SBGN-ML [25], SBGN bricks [26], advanced automatic layout [27], semantic validation of maps, conversion to and from CellDesigner [28] and SBML formats, and an easy switch to alternative colour schemes. SBGN-ED allows map validation [24], conversion from KEGG and SBML formats and PD-to-AF conversion [29]. yEd has an SBGN palette and SBGN layout tool [30], offers



**Figure 2.** An example that compares two representations in CellDesigner that correspond to Activity Flow (Reduced Notation palette in CellDesigner) and Process Description (default palette in CellDesigner). The two diagrams represent the same biological events but in two conceptually different languages. **A.** The Process Description representation of the RAF-MEK-ERK signalling: the process of MEK1/2 phosphorylation is catalysed by RAF1 and the process of ERK1/2 phosphorylation is catalysed by the phosphorylated MEK1/2. **B.** The Activity Flow representation of the RAF-MEK-ERK signalling: the activity of RAF1 stimulates the activity of MEK1/2 (MAP2K1 and MAP2K2 in official HGNC names), and the activity of MEK1/2 stimulates the activity of ERK1/2 (MAPK3 and MAPK1 in official HGNC names).

The choice of an editor is also connected to the visualisation platform used (Figure 1, step 7). The MINERVA and NaviCell [31, 32] platforms by design are compatible to CellDesigner’s XML. The MINERVA Platform also accepts SBGN-ML files and SBML files with layout [33], and can therefore be used with Newt, SBGN-ED, as well as with yEd in combination with the ySBGN converter. Importantly, the platform also supports export to these formats, allowing to review, improve and reupload disease map diagrams.

4. Designing map architecture

Depending on the goals of the project, map coverage and the resources available, there are several possibilities for designing map architecture and for step-by-step progress including starting from PD or AF layer and working with a single diagram or a set of sub-pathways (Table 2).

Suggested starting design is a PD diagram of mechanisms with accompanying top-level view of disease mechanisms. This design focuses on the prioritised pathways, allows reuse of the PD layer from other projects, and helps communicating with domain experts via the top-level view (Figure 3). Please note that each map version (A, B, C and D) is a functional, complete resource. Potentially the AF layer can be created automatically from the PD layer using CaSQ [37] or a PD-to-AF converter [29]. More options for building a multi-layered structure and example architectures of the published maps are available at <https://disease-maps.org/architecture>.

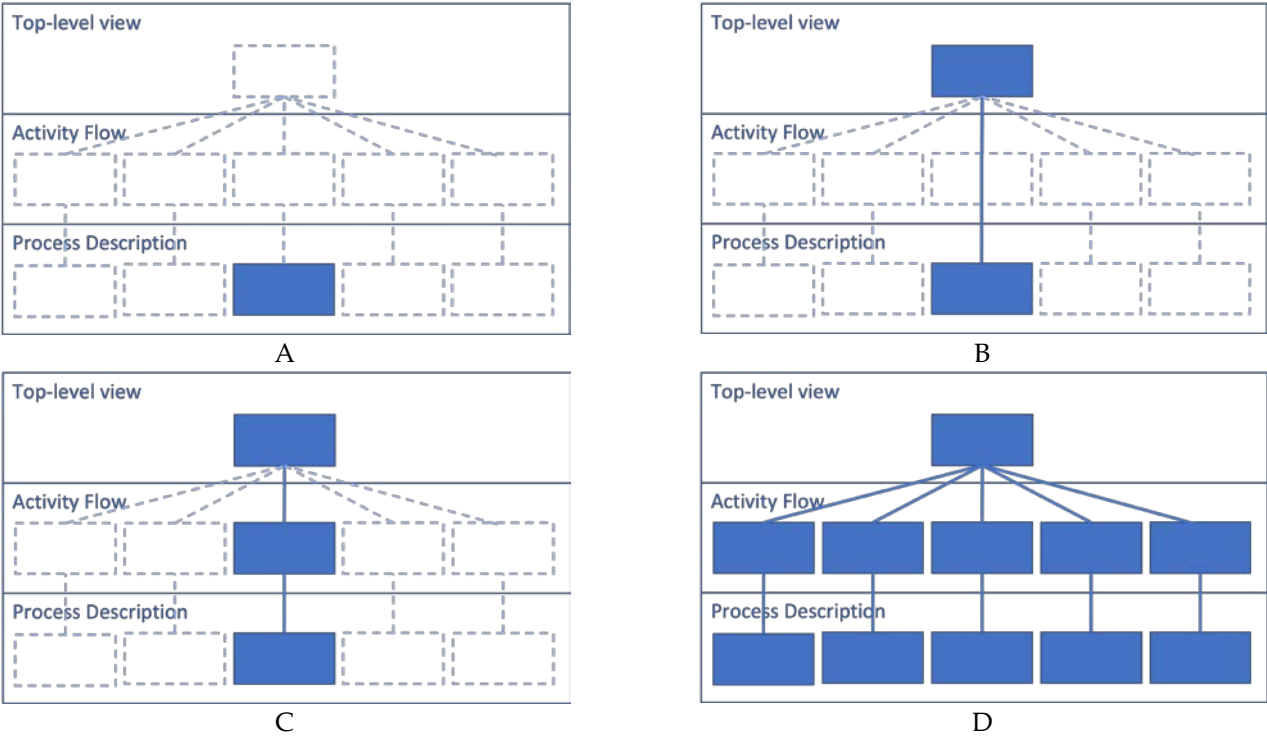
**Table 2.** Examples of map architecture from a single diagram to a hierarchically-organised multi-layered structure.

#	Disease Map	Architecture	Format	Top-level view
a	AlzPathway [34, 35], RA-Map [14]	Single diagram	Process Description	None



b	FluMap [36]	Single diagram	Process Description	None
c	Cystic Fibrosis Map [11]	Single diagram	Activity Flow	None
d	Parkinson’s Disease Map [13]; ACSN [12]	Hierarchical	Process Description, Activity Flow	Illustrative
e	ACSN [12]; RA-Atlas [10], COVID-19 Disease Map [5]	Hierarchical	Process Description	Integrating diagram*
f	AsthmaMap [3]	Hierarchical	Process Description, Activity Flow	Integrating diagram*

\*The integrating diagram has elements that represent lower-layer sub-pathways and this way unites them into a single virtual map.



**Figure 3.** A recommended way to gradually build a hierarchical structure starting from a single PD diagram and step by step making the map architecture more complex. Each map version works as a fully functional complete resource. **A.** Starting with a priority sub-pathway in PD. **B.** Building a top-level view with key mechanisms partly or entirely represented in the PD diagram. **C.** Automatically or manually creating an AF diagram that matches the content of the PD diagram. **D.** Extending the content to multiple diagrams with one integrating top-level view diagram that represents more detailed AF and PD layers.

Modularisation

Modular design of a disease map means developing diagrams that are manageable and reusable [38] for faster curation, reuse in other projects and downstream use in computational modelling [39–41]. A module is a diagram part that is self-contained and can be developed and maintained independently while being included in a larger network.

Modularisation can be applied at different scales including signalling subnetworks, e.g. the MAPK cascade, and on a larger scale, e.g. cell-specific pathways in mast cells, eosinophils and macrophages in asthma (AsthmaMap, <https://asthma-map.org/mr>).

### Examples

AlzPathway (<http://alzpathway.org>) [34, 35] catalogues signalling pathways in Alzheimer's disease as the first comprehensive map of a particular disease. It is a single PD diagram in CellDesigner with an overview image highlighting the included pathways.

The Parkinson's Disease Map [13] was originally designed as a single diagram integrating disease-relevant pathways, e.g. mitochondrial dysfunction, calcium homeostasis, synaptic pathobiology,  $\alpha$ -synuclein misfolding, failure of protein degradation systems, apoptosis and neuroinflammation. Later a top-level view was added in MINERVA for easier overview and navigation ("SHOW OVERVIEW" button), and a number of pathways were added as underlying PD diagrams (<https://pdmap.uni.lu/minerva>).

The FluMap [36] is a comprehensive map of the influenza A virus replication cycle. It is available as a single PD diagram with a simplified version in an AF-compatible format developed in the Reduced Notation in CellDesigner.

The Atlas of Cancer Signalling Networks (ACSN) [12] is a collection of large-scale PD diagrams that cover mechanisms of cancer and include cell cycle and DNA repair, cell survival, regulated cell death, telomere maintenance, epithelial-mesenchymal transition (EMT) and senescence, invasion and motility, and many more (<https://acsn.curie.fr/ACSN2/maps.html>). These modules are also available as part of a single extensive map (<https://acsn.curie.fr>).

The AsthmaMap includes three layers: the detailed Biochemical Mechanisms layer in PD (<https://asthma-map.org/bm>), the intermediate Molecular Relations layer in AF (<https://asthma-map.org/mr>) and the top-layer Cellular Interactions view in AF-compatible format (<https://asthma-map.org/ci>) [3]. The three layers are interconnected and presented in a hierarchically-organised map in MINERVA (<https://asthma.uni.lu>).

The Rheumatoid Arthritis Map (RA-Map) was initially designed as a single Process Description map [14] and then evolved into the Rheumatoid Arthritis Atlas (RA-Atlas), a collection of interconnected diagrams with an access point via a top-level view [10]. The RA-Atlas includes an updated version of the global RA-Map [14] covering relevant metabolic pathways and cell-specific molecular interaction maps for CD4<sup>+</sup> Th1 cells, fibroblasts, and M1 and M2 macrophages.

The COVID-19 Disease Map is a collection of interconnected Process Description diagrams developed by a community-driven effort in the answer to the SARS-CoV-2 pandemic (<https://covid.pages.uni.lu>). The map gathers diagrams developed by curators using CellDesigner and VANTED software, describing mechanisms of virus cycle in the host cells, its interactions with host molecular pathways, and the mechanisms of host response to the infection. These diagrams are organised into one hierarchical map via a top-level view diagram (<https://covid19map.elixir-luxembourg.org/minerva>). Notably, the COVID-19 Disease Map Community involves diagram curators of the Reactome and WikiPathways platforms, hosting their SARS-CoV-2-related diagrams.

### Choice and implementation of a map architecture

The choice of a map design should match the biological coverage of the map (Figure 1, step 2) and the overall progress of the project. The design should be adapted in case of a change in the coverage, timeline or available resources, as the stages introduced in Figure 3 are flexible. Considering the discussed examples, a multi-layered and modularised hierarchically-organised map should be considered as a default design (Figure 3D). Depending on a project, the focus can be only on one layer, PD or AF, and the top-level view is optional (Figure 3A and Figure 3B). Though, with adequate resources available, it is

recommended to work on a multi-layered hierarchical map as the state-of-the-art practice. The main advantages of this hierarchically-organised structure (Table 3):

- rich content and detailed representation supports different types of visualisation and different types of computational modelling;
- a possibility of reusing available resources;
- compatibility with similar disease map projects, map parts are easy to reuse in new projects;
- top-level view of disease hallmarks helps communicating with domain experts.

**Table 3.** Comparison of the top-level view vs. Activity Flow vs. Process Description layers in a hierarchically-organised map. Resources: KEGG: Kyoto Encyclopaedia of Genes and Genomes (KEGG) [42], OmniPath [43], SIGnaling Network Open Resource (SIGNOR) [44], Reactome [45], Recon and ReconMap [46, 47], PANTHER [48].

Features	Top layer: top-level view	Intermediate layer: AF	Detailed layer: PD
SBGN language	Activity Flow or free format**	Activity Flow	Process Description
Connectivity	Interactions	Interactions	Processes
Granularity	Least detailed: focus on key molecules and cell types	Intermediate level of details	Most detailed, the level of processes
Compatible databases	Not available*	KEGG signalling pathways, BioCarta, OmniPath, SIGNOR	Reactome, PANTHER, Recon X, KEGG metabolic pathways
Application/Use	Communication with domain experts for defining disease hallmarks	omics visualisation, network analysis, Boolean modelling, integration with machine learning inferred networks	omics visualisation, network analysis, ODE modelling, Boolean modelling (via CaSQ)

\*The closest similar visualisations are figures with an overview of disease mechanisms in review papers, representations with cell types, tissues and organs shown.

\*\*By free-format here we mean flexible representation with one necessary function: all lower-layer sub-pathways in a hierarchically-organised structure are represented by an element on the top-level view and this way integrate them into a single virtual map.

Finally, other design alternatives are possible, e.g. with more high-level layers or using other standard languages. For instance, the disease map approach supports the concept of adverse outcome pathways (AOPs) in the field of immunotoxicology (paper in preparation).

5. Formalising knowledge from literature and databases

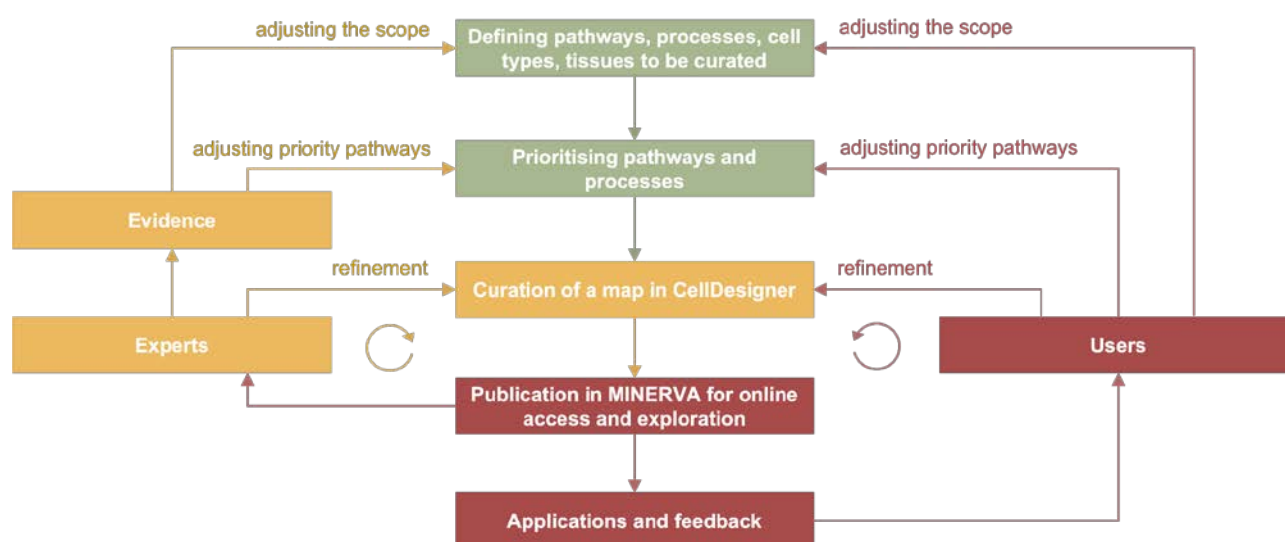
In the knowledge formalisation step, relevant causal interactions and processes are identified and extracted from selected sources, and encoded into a diagrammatic representation [4]. The sources include disease-relevant scientific papers and existing represen-



tations in pathway databases. Knowledge about molecular processes relevant in the context of a particular disease need to be represented in a standard way, for example, receptor activation, phosphorylation, complex formation, positive and negative regulation of transcription, transport, glycosylation, and more. Knowledge is transformed from textual descriptions and figures in scientific articles into a standardised diagrammatic representation. This representation is also machine-readable and can be further converted into executable computational models.

Knowledge about disease mechanisms is formalised in the SBGN standard within two conceptually different languages, AF and PD. AF is focusing on molecular activities and their sequential connectivity, and PD is focusing on the detailed description of the involved molecular processes (see Table 3, Figure 2). AF is compact and is suitable for overview diagrams and for highlighting key mechanisms. PD is detailed and is better for comprehensive representation of molecular events.

In the curation cycle (Figure 4) information is integrated from different sources such as review papers and discussions with domain experts, producing lists of relevant molecules, pathways and cell types. Prioritised pathways are curated in CellDesigner from literature or pathway databases, and visualised in MINERVA for use and review, leading to updates in CellDesigner.



**Figure 4.** Curation and refinement cycle in disease map development. In green are the stages from the planning phase, in yellow from the curation phase and in red from the publication and application stages. The cycle on the left includes feedback from domain experts involved in the development, and the cycle on the right includes feedback from users.

Disease map curation can be seen also as a graphical review of relevant disease mechanisms - a combination of scoping, rapid and systematic reviews [49]. A scoping review, corresponding to step 2 of the map development process (Figure 1), determines the coverage of biological functions, defines disease hallmarks, and identifies knowledge gaps. Rapid and systematic reviews, corresponding to steps 5 and 6 (Figure 1), narrow down the focus and collect evidence from literature and databases for drawing and annotating diagrams in an editor. These steps are usually performed by a dedicated curator with advice and feedback from a panel of domain experts.

This “graphical review” requires following diagram-specific knowledge representation formats. It demands consistency in using a standardised graphical language and describing connectivity between diagram elements that may span beyond the currently curated article. For example, if a paper states that IL6 activates epithelial cells, it is important to identify the relevant protein (IL6, UniProt:P05231), identify receptors involved (IL6RA, IL6ST, IL6RB), and describe the following signalling pathway, which may require looking for additional articles.

More details on knowledge formalisation for disease maps are available at <https://disease-maps.org/formalisation>: the webpage provides examples of search queries in PubMed and PubMed Central, examples of text-to-SBGN and figure-to-SBGN transformation.

6. Annotating map entities and interactions

Two objectives during the annotation step are identifying map components and providing evidence for interactions. Identification assigns stable identifiers or ontology terms to map components to uniquely define them. This ensures interoperability with systems biology tools and external databases. Providing evidence for interactions defines the context of a given interaction, e.g. a specific cell type, tissue, organism, or relevant pathology. Contrary to components, these annotations are not unique but interaction-specific, and require evidence from a relevant source, e.g. scientific publications, curated pathways, or interaction databases.

Annotating entities

Different types of map entities are shown in Table 4 together with recommended annotation.

**Table 4.** Recommended annotation of map entities. Resources: HUGO Gene Nomenclature Committee (HGNC, <https://www.genenames.org>), UniProt (<https://www.uniprot.org>), Complex Portal (<https://www.ebi.ac.uk/complexportal>) [50], ChEBI (<https://www.ebi.ac.uk/chebi>), PubChem (<https://pubchem.ncbi.nlm.nih.gov>), ChEMBL (<https://www.ebi.ac.uk/chembl>), DrugBank (<https://go.drugbank.com>), Cell Ontology (<https://www.ebi.ac.uk/ols/ontologies/cl>), BRENDA (<https://www.brenda-enzymes.org>), Cellosaurus (<https://www.cellosaurus.org>), GO (<http://geneontology.org>), MeSH (<https://www.ncbi.nlm.nih.gov/mesh>).

Biochemical entity	Naming*	Identifier
Protein	HGNC official symbol**	UniProt / HGNC
RNA	HGNC official symbol	HGNC
Gene	HGNC official symbol	HGNC
Metabolite	ChEBI / PubChem recommended name	ChEBI / PubChem
Drug	ChEBI / PubChem recommended name***	ChEBI / PubChem / ChEMBL / DrugBank
Complex	Specific name from literature or listing complex components: Element A:Element B	Not required. If available: Complex Portal
Compartment	Appropriate term from Cell Ontology, BRENDA, Cellosaurus, or a specific name from literature	Cell Ontology / BRENDA / Cellosaurus
Phenotype (biological processes)	Appropriate Gene Ontology (GO) Biological Process (BP) term if available****	GO Biological Process
Phenotype (symptom, disease state)	Appropriate MeSH term if available****	MeSH

\* Flexible if an identifier is provided.  
\*\* This way the protein will be annotated on upload to MINERVA but UniProt will be assigned only if there are no multiple proteins that correspond to that name.

\*\*\*\* Non-standard names are allowed as long as the annotation is provided: for example, “ERK1” instead of “MAPK3” (UniProt:P27361) or “PGE2” instead of “prostaglandin E2” (ChEBI:15551).

\*\*\*\* As in the literature or as suggested by the domain experts.

Proteins, RNAs and genes should be named according to the official symbols of the HUGO Gene Nomenclature Committee (HGNC, <https://www.genenames.org>). For example, while COX1 is a commonly used name, PTGS1 symbol is recommended. This provides consistency in naming and allows skipping manual addition of other stable identifiers in CellDesigner, because HGNC symbols are recognised by the MINERVA annotation tool on upload.

Annotation of complexes is optional because they can be identified by their components. However, it is recommended to name them by listing the molecules involved, for example “FCER1A:FCER1G:MS4A2”, or using specific names if available. When annotating, complex-specific annotations are recommended, e.g. GO Cellular Component (CC) or Complex Portal identifiers [50] (<https://www.ebi.ac.uk/complexportal>). For instance, GO CC term “Fc-epsilon receptor I complex” (GO:0032998) for the complex above.

Phenotypes or other elements representing higher order processes should be annotated with Gene Ontology (GO) Biological Process (BP) terms (<http://geneontology.org/docs/ontology-documentation>) [51] or MeSH terms (<https://www.ncbi.nlm.nih.gov/mesh>). GO BP is recommended for annotating components describing biological processes and MeSH terms are recommended for diseases-related terms.

Metabolites (“simple chemical” in SBN, or “simple molecule” in CellDesigner) are identified via the Chemical Entities of Biological Interest database (ChEBI, <https://www.ebi.ac.uk/chebi>) [52]. Alternatively, PubChem (<https://pubchem.ncbi.nlm.nih.gov>) [53], Human Metabolome Database (<https://hmdb.ca>) [54] or LIPID MAPS (<https://www.lipidmaps.org>) [55] can be used. Automatic MINERVA annotation works for standard ChEBI names, for example “arachidonic acid” (ChEBI:15843), however in other cases annotations need to be provided manually.

An extended description of annotation for disease maps at <https://disease-maps.org/annotation> offers suggested annotation of diagram information and specific details of annotation fields in CellDesigner.

Other useful sources on the topic of annotation are the recommendations of the logical modelling community [40], the Reactome data model and pathway development practices [8], the rules suggested for creating reusable pathway models [7], the documentation of the MI2CAST format (Minimum Information about a Molecular Interaction Causal Statement) [56] and the protocol for high-quality metabolic reconstruction [57].

### Annotating interactions

Scientific articles contain evidence that confirm the existence of molecular interactions in the context of a specific disease. Each interaction should be annotated with stable identifiers of relevant articles, a PubMed ID or a DOI. If pathway interactions are used, their identifiers can be used instead. Although some guidelines suggest a minimal number of references confirming an interaction [4], this should be adapted to the literature coverage of a given disorder.

Importantly, often the relevance of molecular processes in the context of a disease can only be confirmed with a combination of references. For example, one paper can confirm that a specific process is involved in a disease, and another paper will describe details of this process. Only together they offer strong evidence for the interaction.

Some molecular processes related to the disease of interest can be detected in non-human organisms, typically in well-established rodent models for that disease. For parts built based on evidence from cell or animal models, it is important to annotate the corresponding interactions. As proteins, RNA and genes are named using HGNC symbols (human gene symbols), it is recommended to annotate interactions with the NCBI taxon ID

of the non-human organism in which the molecular process was determined. For example, if the phosphorylation of STAT3 by JAK2 was determined in mice, the NCBI Taxonomy ID “10090” (NCBI:txid10090) annotation should be added to the interaction.

As most of disease-specific evidence comes from the literature, annotating map interactions requires maintaining a large repository of articles. Such a repository should be stored in a reference manager like Zotero (<https://www.zotero.org>) or Mendeley (<https://www.mendeley.com>). Articles should be archived while curating and annotating interactions, preferably in a structure mirroring the organisation of the map. Overall, it is important to check that all interactions in the diagram are annotated with evidence. Automated validation is strongly advised, for instance by using the validators available in the MINERVA Platform.

### 7. Visualising in a navigable tool

The MINERVA Platform [33, 58, 59] enables easy online access to and visual interactive exploration of disease maps. Examples are:

- Parkinson’s Disease Map (<https://pdmap.uni.lu>);
- AsthmaMap (<https://asthma.uni.lu>);
- Rheumatoid Arthritis Map (<https://ramap.elixir-luxembourg.org>);
- Cystic Fibrosis Map (<https://cysticfibrosismap.github.io>);
- COVID-19 Disease Map (<https://covid19map.elixir-luxembourg.org>).

Video tutorials are available for introducing key MINERVA functionalities with the AsthmaMap as an example: navigating the maps, adding comments, exploring drug targets and visualising data (<https://asthma-map.org/tutorials>).

Maps can be uploaded to MINERVA in several formats including CellDesigner SBML, SBML+layout+render, SBGN-ML and GPML [33]. The conversion module within the MINERVA Platform allows export to these formats as well. On upload users can choose from CellDesigner’s view and SBGN view.

Automatic annotation can be selected on upload. As discussed above, with HGNC names used for proteins, mRNAs and genes, the MINERVA Platform automatically adds links to relevant external databases.

Map browsing and zooming works in MINERVA in the Google Maps manner. For multi-layered maps, interconnected diagrams are organised into a single entity via the submaps function and are visible from the top-level diagram in the SUBMAPS tab (<https://asthma-map.org/tutorials/#navigating-the-asthmamap>). Search functionality queries all connected maps.

Data can be uploaded as a tab-separated file (<https://minerva.uni.lu/doc/examples>). This way, omics data (for example, differential expressed genes) can be converted into a gradient of colours on the map (<https://asthma-map.org/tutorials/#visualising-data>). Multiple datasets can be visualised simultaneously and compared, including a multi-omics combination (e.g. proteomics and metabolomics), or time series data with different time points visualised as separate datasets.

Drug target search connects via online queries to DrugBank (<https://www.drugbank.ca>) [60] and ChEMBL (<https://www.ebi.ac.uk/chembl>) [61, 62] databases and allows highlighting all entities affected by a drug (<https://asthma-map.org/tutorials/#exploring-drug-targets>).

Moreover, MINERVA API and plugins [59] allow the implementation of customised visual exploration and analytics pipelines.

An alternative web-based tool is NaviCell [31, 32] developed primarily to support the ACSN project. It offers functionalities for visualising omics data for performing functional analysis of the maps, and for computing aggregated values for sample groups and their visualisation on the maps (<https://navicell.vincent-noel.fr>). A video tutorial on ACSN in NaviCell is available at <https://acsn.curie.fr/documentation.html#video>.

### 8. Collecting feedback and updating the map

When the first complete version of a map is available online, it is advisable to request feedback from domain experts regarding the coverage of disease hallmarks, the depth and quality of pathways included, accuracy and adequacy of the disease mechanisms representation. This input can be collected in the form of 1) free-format notes, 2) suggested publications, or 3) direct comments on the map in MINERVA (<https://asthma-map.org/tutorials/#commenting>).

It is also suggested to have an automatic check of any inconsistencies in curation and annotation. This can be done using Neo4j-based tools (see the “Disease map management in the Neo4j graph database” section) or the automated verification in the MINERVA Platform ([https://minerva.pages.uni.lu/doc/admin\\_manual/v16.0/index/#configure-automatic-verification](https://minerva.pages.uni.lu/doc/admin_manual/v16.0/index/#configure-automatic-verification)).

### 9. Applying the resource

Map applications can play a role in the development process, as analysis and computational modelling may identify missing feedback loops or analytical endpoints, key pathways, or larger gaps in the network. Therefore, considering the use-cases of a map can help with its curation, revision and refinement [5].

One of map applications is omics data visualisation for its interpretation, hypothesis generation, and planning validation experiments. Another is exploration of drug targets in MINERVA using the corresponding functionality (<https://asthma-map.org/tutorials>). Finally, computational modelling allows to transform a static diagrammatic representation into a dynamic view of disease progression to be used for refining hypotheses and for predictions.

Demonstrated applications of disease maps include:

data visualisation:

- omics data visualisation [3, 12, 63];
- analysis of cell-specific mechanisms using single-cell expression data [5];
- RNA-Seq-based analysis of the activity of transcription factors [5];

network-based analysis:

- network analysis and representing map-based molecular signatures for sub-groups of patients in cancer [64, 65];
- structural network analysis together with omics data to rationalise the synergistic effects of drugs towards the design of complex disease stage-specific druggable interventions in preclinical studies [66];
- network-based analysis and prediction of epithelial-to-mesenchymal-like transition mechanisms followed by experimental validation with the use of an animal model, a transgenic mice [67];

computational modelling:

- creating computational models based on disease maps, for example for rheumatoid arthritis [16, 17] and atherosclerosis [15];
- creating causal-interaction networks based on a disease map [68];

in preclinical studies with follow-up validation experiments:

- in preclinical studies in cancer [65–67] with validation of the proposed hypothesis by follow-up experiments;

fusing disease maps with other solutions:

- integrating disease maps with machine learning inferred networks [17];



- identifying new crosstalks between pathways from combined analysis of interactions and text mining datasets [5].

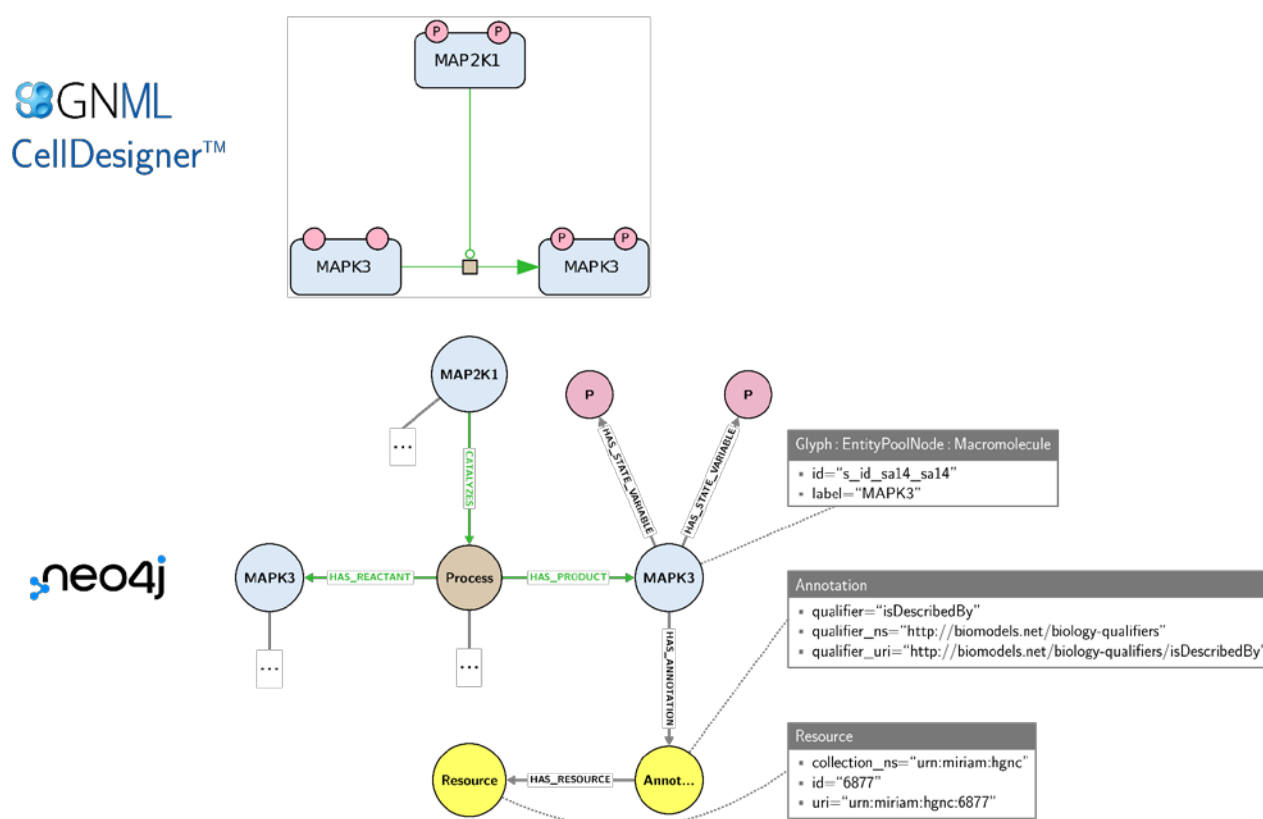
### Long-term sustainability of disease maps

Developing a disease map requires a significant amount of work thus it's important to ensure its long-term sustainability and use. First, the map should be deployed on a stable IT infrastructure for uninterrupted access to the content. This can happen via institutional resources or via a third party, such as the open-access map hosting service offered by ELIXIR Luxembourg (<https://elixir-luxembourg.org/services/catalog/minerva>). Second, a universal programmatic interface should be set up allowing reproducible queries to the disease map contents. The performance of the already available MINERVA API can be further improved by deployment of graph databases, supporting large and complex queries. Finally, a map should be set up as a FAIR resource (<https://www.go-fair.org/fair-principles>) [69], ensuring its long-term reuse. Following sections discuss the topics of graph database deployment and map FAIRification.

#### *Disease map management in the Neo4j graph database*

Maps offer visual representations primarily intended for human comprehension. For their exchange and edition, they are stored in machine-readable formats, e.g. CellDesigner SBML and SBGN-ML. While these formats are well suited for storing and exchanging graphical layout information, they do not offer a convenient solution for the network-based management (integration, querying, exploration) and analysis of collections of maps. The Neo4j graph database was proposed for an efficient management of large-scale highly-interconnected heterogeneous biological data [70], including biological pathway representations in systems biomedicine like Reactome [71], IntAct [72] and Recon2Neo4j [46, 73]. Collections of SBGN maps may also be stored and explored in a Neo4j database using the stonpy software (<https://github.com/adrienrougny/stonpy>), which builds a Neo4j graph from an SBGN-ML map or a CellDesigner map converted to SBGN-ML using cd2sbgnml [28]. Data model of stonpy is close to the visual representation of maps and that is easy to learn. In this model, SBGN nodes (e.g. proteins, states variables) are modelled using Neo4j nodes, while SBGN arcs (e.g. catalyses, production arcs) and relationships between concepts are modelled using Neo4j relationships (edges). Annotations are also stored and can be easily queried. Figure 5 shows an example of a map and an excerpt of its corresponding Neo4j graph as built by stonpy.

The stonpy workflow was used to perform a comparative analysis of integrated biological pathways in several major resources including the Atlas of Cancer Signalling Networks (ACSN), PANTHER, and Reactome [26, 38]. The cd2sbgnml [28] and stonpy tools were also used to develop the dedicated C19DM-Neo4j graph database component (<https://c19dm-neo4j.covid.pages.uni.lu>), which integrates the biological content of the COVID-19 Disease Map diagrams available in MINERVA.



**Figure 5.** An SBGN map and an excerpt of its corresponding Neo4j graph built using the stonpy software. SBGN nodes (e.g. proteins, states variables) are modelled using Neo4j nodes, while SBGN arcs (e.g. catalyses, production arcs) and relationships between concepts are modelled using Neo4j relationships (edges). Neo4j nodes are labelled (e.g. “Glyph”, “Macromolecule”) and may contain key-value pairs (e.g. the pair “label”/“MAPK3”). Additionally, annotations are stored in a structured form that can be easily queried. In the complete model (not shown), SBGN arcs are additionally modelled using Neo4j nodes, as they may contain SBGN nodes themselves.

#### *FAIR principles for disease maps available in MINERVA*

Disease Maps approaches evolve together with other systems biology communities such as COMBINE [74], CoLoMoTo [74], SysMod [75], SBGN [19], SBML [76] and FAIR [69] to enhance the accessibility and reusability of the resources [39].

The FAIR (Findable, Accessible, Interoperable and Reusable) principles (<https://www.go-fair.org/fair-principles>) establish criteria to assess and improve the reuse of scientific resources including data, metadata and infrastructure.

The underlying MINERVA Platform provides a set of functionalities that addresses several elements of FAIR:

- unified API calls to access the standardised content of all diagrams on the platform;
- functionality for translating between systems biology standards;
- authentication and authorisation procedures for user management;
- explicitly mentioning the reuse licence (for example, the COVID-19 Disease Map diagrams are released in MINERVA under the CC-BY 4.0 licence).

More specifically, MINERVA conforms to FAIR expectations in the following ways:

**Findability:** The content of the disease maps diagrams in MINERVA is annotated with a range of information elements from initial publications, connected to well-established biomedical resources (e.g. HGNC, UniProt, ChEBI, ChEMBL, PubMed, etc.) using standard identifiers. The maps are openly available online (e.g. at <https://covid19map.elixir-luxembourg.org>) and indexed in all major search engines.

**Accessibility:** The content of biological maps in MINERVA is accessible easily online via standard web protocols. Moreover, MINERVA provides uniform API calls to access the content of the diagrams and can provide required authentication (credentials: user ID and password) and authorisation procedures where necessary.

**Interoperability:** The biological content in MINERVA follows XML-based systems biology standards such as CellDesigner SBML; translation to other systems biology standard formats is also possible via MINERVA. Moreover, the content of maps is rich in metadata through annotations with information from initial publications connected to biomedical resources as recommended in Table 3, and there are ongoing efforts to connect the content to the Bricks Ontology [26].

**Reusability:** The biological content of disease map diagrams in MINERVA is released under a specific data usage licence, e.g. under the CC-BY 4.0 usage licence for COVID-19 Disease Map [5]. The biological diagrams are primarily edited using CellDesigner, thus complying with the incorporated rules for molecular mechanism representation, and are annotated with various types of information including references to publications (PMIDs) and standard identifiers for molecular entities (UniProt IDs for proteins, ChEBI IDs for metabolites, etc.).

By conforming to most aspects of the FAIR principles, the utility of disease maps available in MINERVA to the systems biomedicine community is greatly enhanced. While there is always room for improvement, and particularly in the domains of versioning and obsoleting policies, the application of FAIR principles throughout the development process of disease maps ensures the long-term relevance of the resource.

## Conclusion

Knowledge of disease mechanisms is dispersed across many publications and domains. Systematic reconstructions of disease mechanisms with the help from disease domain experts allow building reference resources where fragmented information is integrated into disease maps as conceptual disease models. These maps are applied for interpreting newly generated omics data and for computational modelling in order to study diseases as complex systems. For this reason, disease maps need to be constructed in a responsible way, based on a formal, structured and detailed protocol enriched with illustrative examples. This will ensure high-quality and consistency, as well as enable efficient coordination in large-scale collaborations and help young researchers to be faster introduced to best practices, tools and methods.

## Funding

This project has not received external funding.

## Competing interests

Authors declare no competing interests.

## Availability

All relevant data are provided within the paper.

## References

1. Mazein A, Ostaszewski M, Kuperstein I, Watterson S, Le Novère N, Lefaudeux D, et al. Systems medicine disease maps: community-driven comprehensive representation of disease mechanisms. *Npj Syst Biol Appl*. 2018;4.
2. Ostaszewski M, Gebel S, Kuperstein I, Mazein A, Zinovyev A, Dogrusoz U, et al. Community-driven roadmap for integrated disease maps. *Brief Bioinform*. 2018. <https://doi.org/10.1093/bib/bby024>.
3. Mazein A, Ivanova O, Balaur I, Ostaszewski M, Berzhitskaya V, Serebriyskaya T, et al. AsthmaMap: An interactive knowledge repository for mechanisms of asthma. *J Allergy Clin Immunol*. 2021;147:853–6.
4. Kondratova M, Sompairac N, Barillot E, Zinovyev A, Kuperstein I. Signalling maps in cancer research: construction and data analysis. *Database J Biol Databases Curation*. 2018;2018.

5. Ostaszewski M, Niarakis A, Mazein A, Kuperstein I, Phair R, Orta-Resendiz A, et al. COVID19 Disease Map, a computational knowledge repository of virus-host interaction mechanisms. *Mol Syst Biol.* 2021;17:e10387.
6. Tang YA, Pichler K, Füllgrabe A, Lomax J, Malone J, Munoz-Torres MC, et al. Ten quick tips for biocuration. *PLOS Comput Biol.* 2019;15:e1006906.
7. Hanspers K, Kutmon M, Coort SL, Digles D, Dupuis LJ, Ehrhart F, et al. Ten simple rules for creating reusable pathway models for computational analysis and visualization. *PLoS Comput Biol.* 2021;17:e1009226.
8. Varusai TM, Jupe S, Sevilla C, Matthews L, Gillespie M, Stein L, et al. Using Reactome to build an autophagy mechanism knowledgebase. *Autophagy.* 2020;0:1–12.
9. Ravel J-M, Monraz Gomez LC, Sompairac N, Calzone L, Zhivotovsky B, Kroemer G, et al. Comprehensive Map of the Regulated Cell Death Signaling Network: A Powerful Analytical Tool for Studying Diseases. *Cancers.* 2020;12.
10. Zerrouk N, Aghakhani S, Singh V, Augé F, Niarakis A. A Mechanistic Cellular Atlas of the Rheumatic Joint. *Front Syst Biol.* 2022;2.
11. Pereira C, Mazein A, Farinha CM, Gray MA, Kunzelmann K, Ostaszewski M, et al. CyFi-MAP: an interactive pathway-based resource for cystic fibrosis. *Sci Rep.* 2021;11:22223.
12. Kuperstein I, Bonnet E, Nguyen H-A, Cohen D, Viara E, Grieco L, et al. Atlas of Cancer Signalling Network: a systems biology resource for integrative analysis of cancer data with Google Maps. *Oncogenesis.* 2015;4:e160.
13. Fujita KA, Ostaszewski M, Matsuoka Y, Ghosh S, Glaab E, Trefois C, et al. Integrating Pathways of Parkinson's Disease in a Molecular Interaction Map. *Mol Neurobiol.* 2013;49:88–102.
14. Singh V, Kalliolias GD, Ostaszewski M, Veyssiere M, Pilalis E, Gawron P, et al. RA-map: building a state-of-the-art interactive knowledge base for rheumatoid arthritis. *Database J Biol Databases Curation.* 2020;2020.
15. Parton A, McGilligan V, Chemaly M, O'Kane M, Watterson S. New models of atherosclerosis and multi-drug therapeutic interventions. *Bioinforma Oxf Engl.* 2019;35:2449–57.
16. Aghakhani S, Soliman S, Niarakis A. Metabolic Reprogramming in Rheumatoid Arthritis Synovial Fibroblasts: a Hybrid Modeling Approach. 2022;:2022.07.20.500752.
17. Miagoux Q, Singh V, de Mézquita D, Chaudru V, Elati M, Petit-Teixeira E, et al. Inference of an Integrative, Executable Network for Rheumatoid Arthritis Combining Data-Driven Machine Learning Approaches and a State-of-the-Art Mechanistic Disease Map. *J Pers Med.* 2021;11:785.
18. Ochoa D, Hercules A, Carmona M, Suveges D, Gonzalez-Uriarte A, Malangone C, et al. Open Targets Platform: supporting systematic drug-target identification and prioritisation. *Nucleic Acids Res.* 2021;49:D1302–10.
19. Le Novère N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, et al. The Systems Biology Graphical Notation. *Nat Biotechnol.* 2009;27:735–41.
20. Le Novère N. Quantitative and logic modelling of molecular and gene networks. *Nat Rev Genet.* 2015;16:146–58.
21. Rougny A, Touré V, Moodie S, Balaur I, Czauderna T, Borlinghaus H, et al. Systems Biology Graphical Notation: Process Description language Level 1 Version 2.0. *J Integr Bioinforma.* 2019;16.
22. Mi H, Schreiber F, Moodie S, Czauderna T, Demir E, Haw R, et al. Systems Biology Graphical Notation: Activity Flow language Level 1 Version 1.2. *J Integr Bioinforma.* 2015;12:265.
23. Balci H, Siper MC, Saleh N, Safarli I, Roy L, Kilicarslan M, et al. Newt: a comprehensive web-based tool for viewing, constructing and analyzing biological maps. *Bioinformatics.* 2020. <https://doi.org/10.1093/bioinformatics/btaa850>.
24. Czauderna T, Klukas C, Schreiber F. Editing, validating and translating of SBGN maps. *Bioinforma Oxf Engl.* 2010;26:2340–1.
25. Bergmann FT, Czauderna T, Dogrusoz U, Rougny A, Dräger A, Touré V, et al. Systems biology graphical notation markup language (SBGNML) version 0.3. *J Integr Bioinforma.* 2020;17.
26. Rougny A, Touré V, Albanese J, Waltemath D, Shirshov D, Sorokin A, et al. SBGN Bricks Ontology as a tool to describe recurring concepts in molecular networks. *Brief Bioinform.* 2021. <https://doi.org/10.1093/bib/bbab049>.
27. Balci H, Dogrusoz U. fCoSE: A Fast Compound Graph Layout Algorithm with Constraint Support. *IEEE Trans Vis Comput Graph.* 2022;28:4582–93.
28. Balaur I, Roy L, Mazein A, Karaca SG, Dogrusoz U, Barillot E, et al. cd2sbgnml: bidirectional conversion between CellDesigner and SBGN formats. *Bioinforma Oxf Engl.* 2020;36:4975.
29. Vogt T, Czauderna T, Schreiber F. Translation of SBGN maps: Process Description to Activity Flow. *BMC Syst Biol.* 2013;7:115.
30. Siebenhaller M, Nielsen SS, McGee F, Balaur I, Auffray C, Mazein A. Human-like layout algorithms for signalling hypergraphs: outlining requirements. *Brief Bioinform.* 2018. <https://doi.org/10.1093/bib/bby099>.
31. Bonnet E, Viara E, Kuperstein I, Calzone L, Cohen DPA, Barillot E, et al. NaviCell Web Service for network-based data visualization. *Nucleic Acids Res.* 2015;:gkv450.
32. Kuperstein I, Cohen DPA, Pook S, Viara E, Calzone L, Barillot E, et al. NaviCell: a web-based environment for navigation, curation and maintenance of large molecular interaction maps. *BMC Syst Biol.* 2013;7:100.
33. Hoksza D, Gawron P, Ostaszewski M, Hausenauer J, Schneider R. Closing the gap between formats for storing layout information in systems biology. *Brief Bioinform.* 2019. <https://doi.org/10.1093/bib/bbz067>.

34. Mizuno S, Iijima R, Ogishima S, Kikuchi M, Matsuoka Y, Ghosh S, et al. AlzPathway: a comprehensive map of signaling pathways of Alzheimer's disease. *BMC Syst Biol.* 2012;6:52.
35. Ogishima S, Mizuno S, Kikuchi M, Miyashita A, Kuwano R, Tanaka H, et al. AlzPathway, an Updated Map of Curated Signaling Pathways: Towards Deciphering Alzheimer's Disease Pathogenesis. *Methods Mol Biol Clifton NJ.* 2016;1303:423–32.
36. Matsuoka Y, Matsumae H, Katoh M, Eisfeld AJ, Neumann G, Hase T, et al. A comprehensive map of the influenza A virus replication cycle. *BMC Syst Biol.* 2013;7:97.
37. Aghamiri SS, Singh V, Naldi A, Helikar T, Soliman S, Niarakis A. Automated inference of Boolean models from molecular interaction maps using CaSQ. *Bioinforma Oxf Engl.* 2020;36:4473–82.
38. Mazein A, Rougny A, Karr JR, Saez-Rodriguez J, Ostaszewski M, Schneider R. Reusability and composability in process description maps: RAS-RAF-MEK-ERK signalling. *Brief Bioinform.* 2021;22:bbab103.
39. Niarakis A, Waltemath D, Glazier J, Schreiber F, Keating SM, Nickerson D, et al. Addressing barriers in comprehensiveness, accessibility, reusability, interoperability and reproducibility of computational models in systems biology. *Brief Bioinform.* 2022;23:bbac212.
40. Niarakis A, Kuiper M, Ostaszewski M, Malik Sherif RS, Casals-Casas C, Thieffry D, et al. Setting the basis of best practices and standards for curation and annotation of logical models in biology-highlights of the [BC]2 2019 CoLoMoTo/SysMod Workshop. *Brief Bioinform.* 2021;22:1848–59.
41. Cooling MT, Nickerson DP, Nielsen PMF, Hunter PJ. Modular modelling with Physiome standards. *J Physiol.* 2016;594:6817–31.
42. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28:27–30.
43. Túrei D, Korcsmáros T, Saez-Rodriguez J. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat Methods.* 2016;13:966–7.
44. Licata L, Lo Surdo P, Iannuccelli M, Palma A, Micarelli E, Perfetto L, et al. SIGNOR 2.0, the SIGnaling Network Open Resource 2.0: 2019 update. *Nucleic Acids Res.* 2020;48:D504–10.
45. Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, et al. The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* 2022;50:D687–92.
46. Thiele I, Swainston N, Fleming RMT, Hoppe A, Sahoo S, Aurich MK, et al. A community-driven global reconstruction of human metabolism. *Nat Biotechnol.* 2013;31:419–25.
47. Noronha A, Daniélsdóttir AD, Gawron P, Jóhannsson F, Jónsdóttir S, Jarlsson S, et al. ReconMap: an interactive visualization of human metabolism. *Bioinforma Oxf Engl.* 2017;33:605–7.
48. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 2019;47:D419–26.
49. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Inf Libr J.* 2009;26:91–108.
50. Meldal BHM, Perfetto L, Combe C, Lubiana T, Ferreira Cavalcante JV, Bye-A-Jee H, et al. Complex Portal 2022: new curation frontiers. *Nucleic Acids Res.* 2022;50:D578–86.
51. Gene Ontology Consortium. The Gene Ontology resource: enriching a Gold mine. *Nucleic Acids Res.* 2021;49:D325–34.
52. Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, et al. ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res.* 2016;44:D1214–1219.
53. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* 2021;49:D1388–95.
54. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, et al. HMDB: the Human Metabolome Database. *Nucleic Acids Res.* 2007;35 Database issue:D521–526.
55. Sud M, Fahy E, Cotter D, Brown A, Dennis EA, Glass CK, et al. LMSD: LIPID MAPS structure database. *Nucleic Acids Res.* 2007;35 Database issue:D527–532.
56. Touré V, Vercruysse S, Acencio ML, Lovering RC, Orchard S, Bradley G, et al. The Minimum Information about a Molecular Interaction CAusal Statement (MI2CAST). *Bioinforma Oxf Engl.* 2021;36:5712–8.
57. Thiele I, Palsson BØ. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc.* 2010;5:93–121.
58. Gawron P, Ostaszewski M, Satagopam V, Gebel S, Mazein A, Kuzma M, et al. MINERVA-a platform for visualization and curation of molecular interaction networks. *NPJ Syst Biol Appl.* 2016;2:16020.
59. Hoksza D, Gawron P, Ostaszewski M, Smula E, Schneider R. MINERVA API and plugins: opening molecular network analysis and visualization to the community. *Bioinforma Oxf Engl.* 2019;35:4496–8.
60. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2018;46:D1074–82.
61. Bento AP, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M, et al. The ChEMBL bioactivity database: an update. *Nucleic Acids Res.* 2014;42 Database issue:D1083–1090.



- 
62. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 2012;40 Database issue:D1100-1107.
  63. Satagopam V, Gu W, Eifes S, Gawron P, Ostaszewski M, Gebel S, et al. Integration and Visualization of Translational Medicine Data for Better Understanding of Human Diseases. *Big Data.* 2016;4:97-108.
  64. Kuperstein I, Grieco L, Cohen DPA, Thieffry D, Zinovyev A, Barillot E. The shortest path is not the one you know: application of biological network resources in precision oncology research. *Mutagenesis.* 2015;30:191-204.
  65. Jdey W, Thierry S, Russo C, Devun F, Al Abo M, Noguez-Hellin P, et al. Drug-Driven Synthetic Lethality: Bypassing Tumor Cell Genetics with a Combination of AsiDNA and PARP Inhibitors. *Clin Cancer Res Off J Am Assoc Cancer Res.* 2016. <https://doi.org/10.1158/1078-0432.CCR-16-1193>.
  66. Monraz Gomez LC, Kondratova M, Ravel J-M, Barillot E, Zinovyev A, Kuperstein I. Application of Atlas of Cancer Signalling Network in preclinical studies. *Brief Bioinform.* 2019;20:701-16.
  67. Chanrion M, Kuperstein I, Barrière C, El Marjou F, Cohen D, Vignjevic D, et al. Concomitant Notch activation and p53 deletion trigger epithelial-to-mesenchymal transition and metastasis in mouse gut. *Nat Commun.* 2014;5:5005.
  68. Touré V, Flobak Å, Niarakis A, Vercruysse S, Kuiper M. The status of causality in biological databases: data resources and data retrieval possibilities to support logical modeling. *Brief Bioinform.* 2021;22:bbaa390.
  69. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data.* 2016;3:160018.
  70. Lysenko A, Roznovăț IA, Saqi M, Mazein A, Rawlings CJ, Auffray C. Representing and querying disease networks using graph databases. *BioData Min.* 2016;9:23.
  71. Fabregat A, Korninger F, Viteri G, Sidiropoulos K, Marin-Garcia P, Ping P, et al. Reactome graph database: Efficient access to complex pathway data. *PLoS Comput Biol.* 2018;14:e1005968.
  72. Del Toro N, Shrivastava A, Ragueneau E, Meldal B, Combe C, Barrera E, et al. The IntAct database: efficient access to fine-grained molecular interaction data. *Nucleic Acids Res.* 2022;50:D648-53.
  73. Balaur I, Mazein A, Saqi M, Lysenko A, Rawlings CJ, Auffray C. Recon2Neo4j: applying graph database technologies for managing comprehensive genome-scale networks. *Bioinforma Oxf Engl.* 2016. <https://doi.org/10.1093/bioinformatics/btw731>.
  74. Hucka M, Nickerson DP, Bader GD, Bergmann FT, Cooper J, Demir E, et al. Promoting Coordinated Development of Community-Based Information Standards for Modeling in Biology: The COMBINE Initiative. *Front Bioeng Biotechnol.* 2015;3:19.
  75. Dräger A, Helikar T, Barberis M, Birtwistle M, Calzone L, Chaouiya C, et al. SysMod: the ISCB community for data-driven computational modelling and multi-scale analysis of biological systems. *Bioinforma Oxf Engl.* 2021;:btb229.
  76. Keating SM, Waltemath D, König M, Zhang F, Dräger A, Chaouiya C, et al. SBML Level 3: an extensible format for the exchange and reuse of biological models. *Mol Syst Biol.* 2020;16:e9110.