

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO  
INSTITUTO DE COMPUTAÇÃO  
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

GABRIEL DE SOUZA BARBOSA

PEDRO PINTO MOYA

TRANSFORMAÇÃO DE DADOS DE PRONTUÁRIO PARA APOIO À PESQUISA  
CLÍNICA: UMA APLICAÇÃO NO CASO DA COVID-19 PARA O PROJETO VODAN BR

RIO DE JANEIRO

2023

GABRIEL DE SOUZA BARBOSA

PEDRO PINTO MOYA

TRANSFORMAÇÃO DE DADOS DE PRONTUÁRIO PARA APOIO À PESQUISA  
CLÍNICA: UMA APLICAÇÃO NO CASO DA COVID-19 PARA O PROJETO VODAN BR

Trabalho de conclusão de curso de graduação apresentado ao Instituto de Computação da Universidade Federal do Rio de Janeiro como parte dos requisitos para obtenção do grau de Bacharel em Ciência da Computação. (justificado sem recuo, início no meio da página).

Orientadora: Prof. Maria Luiza Machado Campos

Co-orientadora: Vânia Jesus de Araujo Soares Borges

RIO DE JANEIRO

2023

B238t

Barbosa, Gabriel de Souza

Transformação de dados de prontuário para apoio à pesquisa clínica: uma aplicação no caso da COVID-19 para o projeto VODAN BR/ Gabriel de Souza Barbosa e Pedro Pinto Moya. – 2023.

106 f.

Orientadora: Maria Luiza Machado Campos

Coorientadora: Vânia Jesus de Araujo Soares Borges.

Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação)  
- Universidade Federal do Rio de Janeiro, Instituto de Computação, Bacharel em Ciência da Computação, 2022.

1. VODAN. 2. ETL. 3. COVID-19. 4. Pentaho. 5. HUGG. I. Moya, Pedro Pinto. II. Campos, Maria Luiza Machado (Orient). III. Borges, Vânia Jesus de Araujo Soares (Coorient.). IV. Universidade Federal do Rio de Janeiro, Instituto de Computação. V. Título.

GABRIEL DE SOUZA BARBOSA

PEDRO PINTO MOYA

TRANSFORMAÇÃO DE DADOS DE PRONTUÁRIO PARA APOIO À PESQUISA  
CLÍNICA: UMA APLICAÇÃO NO CASO DA COVID-19 PARA O PROJETO VODAN BR

Trabalho de conclusão de curso de graduação  
apresentado ao Instituto de Computação da  
Universidade Federal do Rio de Janeiro como  
parte dos requisitos para obtenção do grau de  
Bacharel em Ciência da Computação.

Aprovado em 17 de ABRIL de 2023.

BANCA EXAMINADORA:



Documento assinado digitalmente  
MARIA LUIZA MACHADO CAMPOS  
Data: 29/06/2023 17:34:02-0300  
Verifique em <https://validar.iti.gov.br>

\_\_\_\_\_  
Profa. Maria Luiza Machado Campos. Ph. D.



Documento assinado digitalmente  
JOAO CARLOS PEREIRA DA SILVA  
Data: 03/07/2023 14:27:37-0300  
Verifique em <https://validar.iti.gov.br>

\_\_\_\_\_  
Prof. João Carlos Pereira da Silva. D.Sc.



Documento assinado digitalmente  
GISELI RABELLO LOPES  
Data: 04/07/2023 14:58:16-0300  
Verifique em <https://validar.iti.gov.br>

\_\_\_\_\_  
Prof. Giseli Rabello Lopes. D.Sc.



Documento assinado digitalmente  
VANIA JESUS DE ARAUJO SOARES BORGES  
Data: 04/07/2023 14:18:35-0300  
Verifique em <https://validar.iti.gov.br>

\_\_\_\_\_  
Vânia Jesus de Araujo Soares Borges, D.Sc.

Este trabalho é dedicado a Natalina  
Lopes e a toda a sua jornada, sem você, nada  
disso seria possível

## **AGRADECIMENTOS DE GABRIEL DE SOUZA BARBOSA**

Gostaria de primeiramente agradecer a Deus que me possibilitou viver toda essa jornada. A minha avó Natalina, grande sonhadora desta conquista e matriarca desta família, você foi e sempre será a inspiração para todos nós, onde quer que esteja.

Em seguida, quero agradecer aos meus pais, minha mãe que é meu porto seguro e que sempre esteve comigo sempre que eu precisava, mesmo antes que eu soubesse disso, ao meu pai que é meu grande exemplo e meu melhor amigo, o homem do qual me inspiro todos os dias a ser. As minhas irmãs Raphaela, que é minha joia mais preciosa e melhor amiga, e Renata que sempre foi minha companheira e esteve comigo nos momentos importantes pra me ensinar e aconselhar.

Aos meus primos Jacqueline e Leandro, que sempre foram minhas inspirações pessoais, obrigado por todos os momentos, ajuda, conselhos e por estarem sempre comigo. Aos meus tios, Helena, Geraldo e Zezé, que foram meus pilares por anos e que ainda são, vocês são minha família e agradeço a todos vocês por todo o amor que vocês me deram e ainda dão.

A minha esposa, amiga e amor da minha vida, Tainá, obrigado por estar comigo nos melhores e piores momentos e me ajudar a seguir em frente, você foi e é parte importante de toda essa jornada.

Aos meus amigos, Kath, Giovanni, Tardelli, Moya, Kah, Diego, Camila, Mario, Marcelo, Bianca, e a todos os outros que me acompanharam, motivaram e dividiram comigo as madrugadas viradas. Sem vocês tudo seria mais difícil.

Também gostaria de agradecer a Profa. Maria Luiza e a Vânia (Que sempre será chamada de professora por mim por todos os méritos a que tem direito), que foram verdadeiras guias para que esse trabalho fosse entregue com qualidade e apreço. Um agradecimento especial também a equipe do Graffe que foi uma das peças principais e primordiais para que esse trabalho fosse realizado.

## **AGRADECIMENTOS DE PEDRO PINTO MOYA**

Gostaria de primeiramente agradecer a minha família. Meus pais Miguel e Adriana, sem eles e seu amor não seria ninguém e nunca teria chegado aonde cheguei. A minha irmã Maria Clara, que consegue alegrar o mundo com seu sorriso. A meu irmão Ricardo, que teve paciência comigo durante muitas complicações e solavancos da vida, sempre contagiando com seu bom humor.

Gostaria também de agradecer a todos os grandes amigos que fiz durante esta jornada. André, Gabriel, Cíntia, João, Camila, Mario, Giovanni, e muitos outros que não conseguiria citar em apenas alguns parágrafos. Sem vocês eu não teria conseguido vir tão longe.

Por último, gostaria de agradecer às professoras Maria Luiza e Vânia Borges. Sem seus ensinamentos e instrução não teríamos conquistado esta vitória..

Em respeito aos 6.450.000 mortos pela  
COVID-19 até o momento da publicação deste  
trabalho, e suas famílias.



## RESUMO

Visando permitir que dados relativos a COVID-19 estejam disponíveis de forma global, com o objetivo de permitir o livre acesso à informação, foi criada a *Rede de Implementação Virus Outbreak Data Network* (VODAN). Este trabalho foi desenvolvido com o intuito de prover novos dados nesta rede de informação, contribuindo para o enfrentamento global do coronavírus. Para isso, os dados clínicos do Hospital Universitário Gaffrée Guinle (HUGG) são coletados, tratados e disponibilizados na forma de perguntas e respostas, de acordo com o formulário de registro de caso estabelecido pela Organização Mundial da saúde (OMS). Esse tratamento é parte essencial da Plataforma VODAN BR, parte da rede de implementação no Brasil, preparando os dados para o processo de FAIRificação de dados, que visa tornar mais fácil um dado de ser descoberto, acessado, integrado e reutilizado por outros pesquisadores e partes interessadas. Isso aumenta a visibilidade e o impacto dos dados, além de permitir sua validação e replicação por outros pesquisadores. A preparação desses dados agiliza o processo de FAIRificação, contribuindo para a disponibilização de dados e metadados na rede VODAN.

**Palavras-chave:** VODAN.;ETL; COVID-19; Pentaho; HUGG

## ABSTRACT

Aiming to make COVID-19-related data globally available with the objective of enabling free access to information, the Virus Outbreak Data Network Implementation Network (VODAN) was created. This work was developed to provide new data in this information network, contributing to the global response to the coronavirus. For this purpose, clinical data from the Gaffrée Guinle University Hospital (HUGG) is collected, processed, and made available in a question-and-answer format according to the case registration form established by the World Health Organization (WHO). This processing is an essential part of the VODAN BR Platform, part of the implementation network in Brazil, preparing the data for the process of data FAIRification, which aims to make data easier to discover, access, integrate, and reuse by other researchers and stakeholders. This increases the visibility and impact of the data, as well as allowing its validation and replication by other researchers. The preparation of this data speeds up the process of FAIRification, contributing to the availability of data and metadata in the VODAN network.

**Keywords:** VODAN; ETL; COVID-19; Pentaho; HUGG

## LISTA DE ILUSTRAÇÕES

<b>Figura 1:</b> Casos e mortes provenientes de COVID-19 durante o período de 31/12/2019 a 13/02/2020.....	14
<b>Figura 2:</b> Quantidade de mortes diárias entre o período de 18/03/2022 e 10/03/2022 .....	15
<b>Figura 3:</b> Mortes no Brasil durante a pandemia de COVID-19 entre Abril de 2020 e meados de Outubro de 2021 .....	16
<b>Figura 4:</b> Estrutura do projeto Data Together .....	20
<b>Figura 5:</b> Arquitetura do projeto VODAN.....	22
<b>Figura 6:</b> Exemplo de perguntas de parte do módulo 1 do FRC-OMS.....	24
<b>Figura 7:</b> Exemplo de perguntas de parte do módulo 2 do FRC-OMS.....	24
<b>Figura 8:</b> Exemplo de perguntas de parte do módulo 3 do FRC-OMS.....	25
<b>Figura 9:</b> Estrutura disponibilizada pelo HUGG em SGBD PostgreSQL .....	29
<b>Figura 10:</b> Estrutura de dados da Anamnese .....	30
<b>Figura 11:</b> Exemplo de informação descrita pela especialidade de nutrição .....	30
<b>Figura 12:</b> Exemplo de informação registrada na tabela de Anamnese .....	31
<b>Figura 13:</b> Exemplo da estrutura da tabela de controle.....	32
<b>Figura 14:</b> Exemplo da estrutura da tabela de dados demográficos.....	32
<b>Figura 15:</b> Exemplo da estrutura da tabela de evolução .....	33
<b>Figura 16:</b> Exemplo da estrutura da tabela de internação .....	34
<b>Figura 17:</b> Exemplo da estrutura da tabela de medicamentos.....	35
<b>Figura 18:</b> Exemplo das 30 opções de bancos como fontes para o Pentaho PDI. ....	37
<b>Figura 19:</b> Série de opções de input de informações do Pentaho.....	38
<b>Figura 20:</b> Exemplo do cálculo da idade a partir da data de nascimento no Pentaho. ....	38
<b>Figura 21:</b> Exemplo do cálculo da cotação em real de compras no exterior no Pentaho.....	39
<b>Figura 22:</b> Exemplo de sub-transformação. ....	40
<b>Figura 23:</b> Visão geral das tabelas envolvidas na operação do ETL.....	42
<b>Figura 24:</b> Arquitetura da primeira solução, com o funcionamento das transformações não detalhadas.....	43
<b>Figura 25:</b> Primeira parte da transformação central.....	44
<b>Figura 26:</b> Segunda parte da transformação central com etapas da busca por crformsID em evidência.....	45
<b>Figura 27:</b> Segunda parte da transformação central com etapas de sub- transformações em evidência.....	46
<b>Figura 28:</b> A sub-transformação createFormRecords. ....	47
<b>Figura 29:</b> A sub-transformação getProntuarios. ....	48
<b>Figura 30:</b> Lógica de geração de registros dos módulos 1 e 3 para a tabela tb_FormRecord. 48	
<b>Figura 31:</b> Exemplo de busca em BD que utiliza uma informação dinâmica representada por ' ? ' vinda dos registros presentes no fluxo. ....	49
<b>Figura 32:</b> Lógica de geração do registro dos módulos 2 para a tabela tb_FormRecord.....	50
<b>Figura 33:</b> Operação finais da sub-transformação createFormRecords .....	51
<b>Figura 34:</b> Arquitetura da sub-transformação getAssesstmentQuest.....	52
<b>Figura 35:</b> Terceira parte da transformação central .....	53
<b>Figura 36:</b> Query da operação tb_form.....	54
<b>Figura 37:</b> Arquitetura da chamada para as sub-transformações de geração de respostas.....	56
<b>Figura 38:</b> Desenho da arquitetura da transformação Q1 .....	57
<b>Figura 39:</b> Arquitetura da transformação Q1 no Pentaho .....	58
<b>Figura 40:</b> Exemplo de transformação com alta complexidade da primeira versão da solução. ....	59

<b>Figura 41:</b> Nova arquitetura.....	61
<b>Figura 42:</b> Estrutura da transformação de respostas para questões do formato sim, não e desconhecido.....	63
<b>Figura 43:</b> Fluxos de coleta de dados para a resolução de questões do formato sim, não e desconhecido.....	63
<b>Figura 44:</b> Fluxo de coleta de dados da tabela de anamnese.....	64
<b>Figura 45:</b> Fluxo de coleta das questões que serão respondidas na transformação de perguntas múltiplas escolhas e sua fonte de dados.....	65
<b>Figura 46:</b> Conjunto de filtros que direciona os registros para sua fonte de informação correta.....	66
<b>Figura 47.A:</b> Primeira parte da estrutura da busca de dados na fonte de evoluções.....	67
<b>Figura 47.B:</b> Segunda parte da estrutura da busca de dados na fonte de evoluções.....	67
<b>Figura 48:</b> Estrutura da busca de dados na fonte de medicamentos.....	68
<b>Figura 49:</b> Parte final da transformação model_yes_unk.....	68
<b>Figura 50:</b> Filtragem dos campos que serão utilizados na lógica de decisão da transformação de questões múltiplas escolhas.....	69
<b>Figura 51:</b> Operação de filtragem por regex.....	69
<b>Figura 52:</b> Regex utilizado nas buscas cadastrado no banco para cada questão individualmente.....	70
<b>Figura 53:</b> Relação de respostas possíveis para cada tipo de pergunta múltipla escolha.....	71
<b>Figura 54:</b> Lógica para extração dos dados das tabelas de informações do HUGG.....	72
<b>Figura 55:</b> Lógica da coleta de respostas de questões com extração de texto livre.....	73
<b>Figura 56:</b> Script de extração utilizado na transformação model_Extract.....	74
<b>Figura 57:</b> Parte 1 da Arquitetura da transformação UCI_GetSave.....	75
<b>Figura 58:</b> Parte 2 da arquitetura da transformação UCI_GetSave.....	76
<b>Figura 59:</b> Parte 3 da arquitetura da transformação UCI_GetSave.....	77
<b>Figura 60:</b> Parte 4 da arquitetura da transformação UCI_GetSave.....	78
<b>Figura 61:</b> Parte 5 da arquitetura da transformação UCI_GetSave.....	79
<b>Figura 62:</b> Arquitetura inicial da transformação Outcome.....	80
<b>Figura 63:</b> Lógica de resolução do estado de saúde do paciente no momento da alta/óbito.....	81
<b>Figura 64.A:</b> Arquitetura inicial da transformação get_old_questions.....	82
<b>Figura 65:</b> Lógica de resposta da transformação get_old_questions.....	83
<b>Figura 66.A:</b> Arquitetura inicial da transformação DurationDays.....	84
<b>Figura 67:</b> Lógica de cálculo da duração de dias em que o paciente mostrou determinado comportamento ou característica.....	85
<b>Figura 68:</b> Parte final da arquitetura da transformação DurationDays.....	86
<b>Figura 69:</b> Resultados gerados pelo ETL.....	87

## LISTA DE SIGLAS

CODATA - Committee on Data for Science and Technology

COVID – Corona Vírus Disease

ESPII - Emergência de Saúde Pública de Importância Internacional

ETL - Extract, Transform and Load

FIOCRUZ - Fundação Oswaldo Cruz

FRC-OMS - Formulário de Registro de Caso da OMS

OMS - Organização Mundial de Saúde

PDI - Pentaho Data Integration

RDA - Research Data Alliance

SARS-CoV-2 - Síndrome respiratória aguda grave coronavírus 2

UFRJ - Universidade Federal do Rio de Janeiro

UNIRIO - Universidade do Rio de Janeiro

VODAN - Virus Outbreak Data Network

VODAN BR - Virus Outbreak Data Network Brazil

WDS - World Data System

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>13</b>
1.1 MOTIVAÇÃO.....	16
1.2 OBJETIVO.....	18
1.3 ESTRUTURA .....	19
<b>2 A REDE VODAN E O PROJETO VODAN BR</b> .....	<b>20</b>
2.1 ARQUITETURA DO PROJETO .....	21
2.2 FRC-OMS .....	23
2.3 TRABALHOS RELACIONADOS .....	25
<b>3 SOLUÇÃO</b> .....	<b>27</b>
3.1 MODELO DE DADOS DO HUGG.....	28
3.2 ORIENTAÇÕES DA EQUIPE TÉCNICA DO HUGG.....	35
3.3 PENTAHO DATA INTEGRATION.....	36
3.4 VERSÃO 1.0.....	41
3.4.1 Transformação Central.....	43
3.4.2 Transformação Q1 .....	56
3.4.3 Problemas encontrados .....	58
3.5 VERSÃO 2.0.....	59
3.5.1 Arquitetura da solução.....	60
3.5.2 Resolução de questão de múltipla escolha ( <i>model_yes_unk</i> ).....	61
3.5.3 Resolução de questão de extração ( <i>model_extract</i> ).....	71
3.5.4 Duração e dias na UTI ( <i>UCI_getAndSave</i> ).....	74
3.5.5 Informações de saída ( <i>Outcome</i> ).....	79
3.5.6 Questões que referenciam outras questões ( <i>get_old_questions</i> ).....	81
3.5.7 Questões com duração de dias ( <i>DurationDays</i> ) .....	83
<b>4 CONCLUSÃO</b> .....	<b>87</b>
4.1 TRABALHOS FUTUROS .....	88
<b>REFERÊNCIAS</b> .....	<b>89</b>
<b>ANEXO A – EXTRATO DE ATAS DE REUNIÕES</b> .....	<b>93</b>
<b>ANEXO B – TABELA DE MEDICAMENTOS</b> .....	<b>98</b>
<b>ANEXO C – DICIONÁRIO DE DADOS DO HUGG</b> .....	<b>101</b>

## 1 INTRODUÇÃO

O coronavírus é uma família de vírus amplamente conhecida, responsável por causar doenças pulmonares em seres humanos. Até o início da pandemia de COVID-19 desencadeada no final de 2019, eram conhecidos 6 tipos diferentes de coronavírus: HCoV-229E, HCoV-OC43, HCoV-NL63, HCoV-HKU1 e o tipo mais mortal, o SARS-COV, responsável por causar a síndrome respiratória aguda grave (LAI et al., 2014).

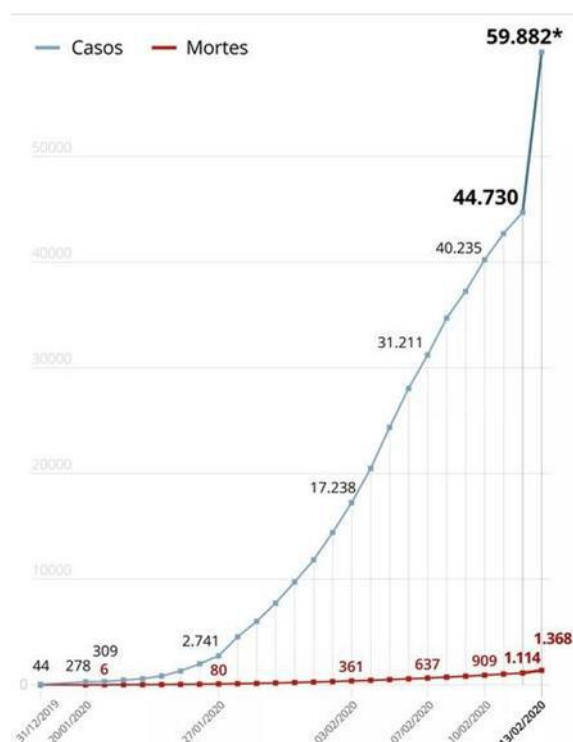
Em 31 de dezembro de 2019, a Organização Mundial de Saúde (OMS) recebeu uma série de relatos de casos de pneumonia assolando a cidade de Wuhan, na China. O crescente caso de doenças pulmonares semelhantes na região fez a organização monitorar os pacientes para descobrir a origem das enfermidades. Até aquele momento, ainda não havia ligação entre os relatos e uma nova doença (REUTERS, 2019).

Uma semana após a OMS ser notificada, em 7 de janeiro de 2020, as autoridades chinesas confirmaram a descoberta de um novo tipo de coronavírus, nomeado temporariamente de 2019-nCov. A primeira morte decorrente do novo vírus, ocorreu apenas 2 dias depois, em 9 de janeiro de 2020 (DE SÁ, 2021). Seria a primeira, das mais de 6 milhões de mortes que ocorreriam em mais de 2 anos de pandemia.

Após atingir 19 países, o surto foi declarado como uma Emergência de Saúde Pública de Importância Internacional (ESPII) pela OMS em 30 de janeiro de 2020, maior nível de emergência da Organização Mundial de Saúde.

Em 11 de fevereiro de 2020 o nome do novo coronavírus foi alterado de forma definitiva para SARS-CoV-2, responsável pela transmissão da doença COVID-19. No final do mesmo mês, em 26 de fevereiro, foi registrado o primeiro caso no Brasil (DE SÁ, 2021). Algumas semanas mais tarde, em 11 de março de 2020 a COVID-19 foi declarada pela OMS como uma pandemia, o vírus já havia se espalhado por diversos países do mundo, com focos de transmissão bem estruturados. De acordo com Moreira e Pinheiro (2021), na data havia mais de 118 mil casos com 4.291 mortes confirmadas. A progressão de casos e mortes cresceu de forma exponencial, atingindo o mundo em uma velocidade assustadora, como pode ser visto na Figura 1.

**Figura 1:** Casos e mortes provenientes de COVID-19 durante o período de 31/12/2019 a 13/02/2020.



Fonte G1: COM NOVA METODOLOGIA, NÚMERO DE CASOS DE CORONAVÍRUS NA CHINA VAI A 59,8 MIL; MORTES SOMAM 1,3 MIL, 2021.

A Europa foi considerada o centro ativo da COVID-19 pela OMS em 13 de março de 2020. Em 14 de março, a Espanha decretou lockdown. Neste período, os casos na Europa dobravam em períodos de 2 a 4 dias (DE SÁ, 2021). Neste momento o pânico e a desinformação assolaram diversos países, com corridas às lojas para busca de diversos itens de sobrevivência, com a população receosa de enfrentar um desabastecimento de produtos básicos. Gondolas de papel higiênico, comida e água desapareceram de diversos mercados. A proliferação de *fake news* e estudos inconclusivos bombardearam as redes sociais com informações conflitantes e prejudicaram de forma geral a resposta da população à pandemia.

Em 13 de setembro de 2020, a COVID-19 já havia vitimado mais de 132 mil pessoas no Brasil, se tornando a doença mais fatal em um período de um ano desde 1979, data da informação mais antiga encontrada no SIM (Sistema de Informação sobre Mortalidade) do Ministério da Saúde (MADEIRO, 2021). De acordo com Sanar (2021), no final do mesmo mês, em 30 de Setembro de 2020, o Brasil atinge a marca de mais de 1.000 óbitos em um único dia, precisamente 1.031 óbitos em 24 horas.

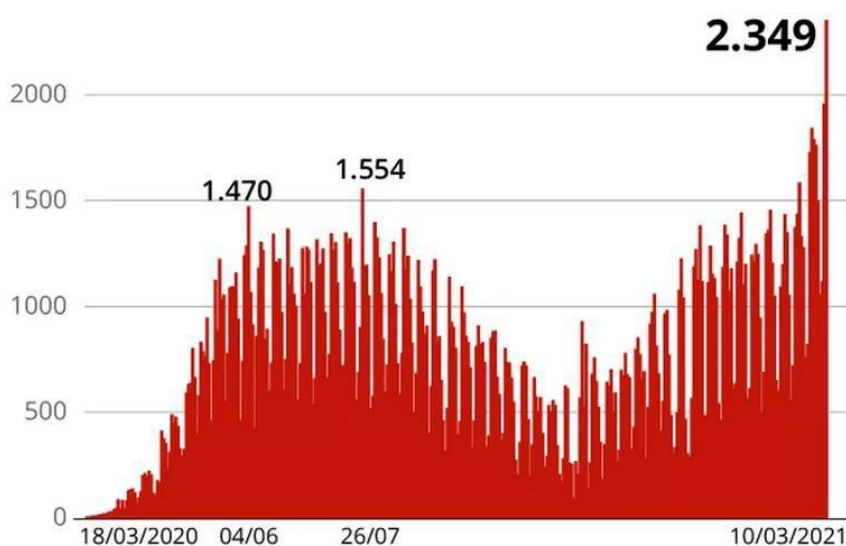


Segundo Holton (2021), finalmente em 8 de dezembro de 2020, após passar por todas as fases de segurança, a primeira vacina foi aplicada a Margaret Keenan, de 90 anos, no Reino Unido. Margaret se tornou o primeiro ser humano a ser vacinado contra COVID-19 na história.

Segundo Baddini e Fernandes (2021), no Brasil a vacinação começou em 17 de janeiro de 2021, com a enfermeira Mônica Calazans, de 54 anos, se tornando a primeira pessoa a ser vacinada no país. Vale frisar que em pandemias como a do COVID-19, a busca por vacinas se torna uma corrida em todos os países do mundo.

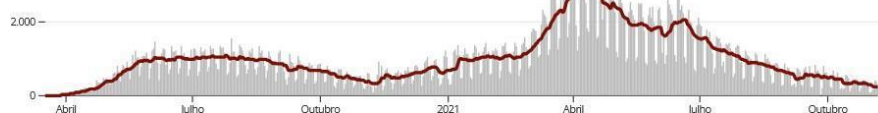
Por este motivo, durante os meses que se seguiram à vacinação, o Brasil enfrentou diversos casos de paralisação do programa de imunização por decorrência da falta de insumos para a produção e aplicação dos imunizantes. A Figura 2 destaca o número de mortos no período de tempo entre 18/03/2020 e 10/03/2021, cenário onde apesar da vacina existir e funcionar, não estava disponível para a população em massa, resultando em um contínuo aumento de casos (Aumento de casos em março em relação ao ano anterior). A Figura 3 destaca, em primeiro momento, que a quantidade de mortes diárias não sofreu nenhum impacto com a chegada do imunizante (Aumento de mortes em Abril, mesmo com o imunizante desenvolvido).

**Figura 2:** Quantidade de mortes diárias entre o período de 18/03/2020 e 10/03/2021



Fonte G1: BRASIL REGISTRA 2.349 MORTES EM 24 HORAS, 2021.

**Figura 3:** Mortes no Brasil durante a pandemia de COVID-19 entre Abril de 2020 e meados de Outubro de 2021



Fonte G1: Mortes e casos de coronavírus nos estados, 2021

A vacinação apresenta bons números com 55,89% da população totalmente imunizada e 72,97% com ao menos a 1ª dose do imunizante até 06 de novembro de 2021 (G1: Vacinação contra a covid: 55,89% da população está totalmente imunizada, 2021). De acordo com Alves (2021), apesar do otimismo quanto ao efeito da vacinação, vários países se encontram em novas crises de COVID-19, decorrentes principalmente de grupos da população aderentes a ideais anti-vacina. Neste momento de incerteza sobre o futuro, alguns efeitos colaterais da pandemia assolam diversos países, assim como o Brasil. Desemprego acentuado, aumento expressivo da inflação, desvio de recursos para tratamento da COVID-19, dentre outros. Nesse panorama, temos uma doença que não deve sumir nos próximos anos, mas que irá continuar presente na população mundial, assim como as outras variantes do coronavírus, unido a um cenário de recuperação econômica.

## 1.1 MOTIVAÇÃO

Neste contexto é importante seguirmos com projetos e políticas que estimulem o estudo da doença e nos permitam entender melhor seus impactos, efeitos e principalmente as formas de tratamento disponíveis. Ainda melhor, que esses projetos possam ser aproveitados como respostas a emergências e de forma proativa a novas e já conhecidas doenças que possam nos assolar no futuro.

Para efeito de comparação, de acordo com Morales (2021), a vacina mais rápida já feita pela humanidade, antes da pandemia de COVID-19 foi a vacina contra Caxumba, que levou 4 anos para ser finalizada. O imunizante contra COVID-19 foi desenvolvido no prazo de apenas 10 meses. Essa diferença considerável entre o tempo de pesquisa e desenvolvimento mostra como ainda há espaços de pesquisa sobre o vírus e seus efeitos.

Uma forma eficaz de estudar o impacto da doença é utilizar dados de pacientes de COVID-19, infectados durante a pandemia. Com milhões de casos ocorrendo no mundo desde o surgimento da pandemia, esses dados representam uma fonte importante para entender como a doença evolui contemplando uma grande variação de fatores. Por meio deles, é possível cruzar qual o impacto de determinado tratamento para pacientes com e sem comorbidades, assim como identificar aspectos que interferem na resolução e os principais fatores associados à alta e ao óbito, dentre outros.

Com o intuito de padronizar esses dados, de forma a disponibilizar as informações de forma global e facilitar o estudo e pesquisa, a OMS criou um Formulário de Registro de Caso (FRC), em inglês “*Case Record Form*”<sup>1</sup>, para o registro das informações dos pacientes. Esse formulário é referenciado neste trabalho como FRC-OMS. O FRC-OMS é dividido em 3 módulos. O primeiro módulo se refere a admissão do paciente, o segundo a evolução do paciente durante a sua internação na unidade de saúde, e o último contempla as informações sobre o desfecho do atendimento, como a alta ou o óbito do paciente.

No decorrer da pandemia, surgiram diversas iniciativas com o objetivo de auxiliar no combate ao vírus e na disseminação de informações corretas. De fato, apesar da informação ser um dos pilares no combate de epidemias virais, encontrar essa informação pode ser complicado, como ocorre hoje ao buscar informações sobre a Ebola, apesar de diversas epidemias terem ocorrido no passado. Além disso, de acordo com Henning et al (2019), transformar os dados para o modelo FAIR (*Findable, Accessible, Interoperable, Reusable*), como feito da iniciativa mundial GO FAIR<sup>2</sup>, é um caminho importante para democratizar o acesso à informação e salvar vidas.

Visando permitir que os dados relativos a COVID-19 estejam disponíveis de forma global, com o objetivo de permitir o livre acesso à informação, foi criada a *Rede Virus Outbreak Data Network*<sup>3</sup> (VODAN). Podemos definir a rede como uma infraestrutura de dados alinhada aos princípios FAIR e que apoia a troca de informações sobre pacientes infectados pelo COVID-19 (VEIGA *et al*, 2021), assim como futuras pandemias.

---

<sup>1</sup> [https://www.who.int/publications/i/item/WHO-2019-nCoV-Clinical\\_CRF-2020.4](https://www.who.int/publications/i/item/WHO-2019-nCoV-Clinical_CRF-2020.4)

<sup>2</sup> <https://www.go-fair.org/>

<sup>3</sup> <https://www.go-fair.org/implementation-networks/overview/vodan/>

Com a proposta de participar dessa rede internacional, disponibilizando dados coletados a partir de hospitais brasileiros, foi criada a *Virus Outbreak Data Network Brazil* (VODAN BR). Essa rede de implementação é coordenada pela Fundação Oswaldo Cruz (FIOCRUZ), com a participação da Universidade Federal do Rio de Janeiro (UFRJ), da Universidade Federal do Estado do Rio de Janeiro (UNIRIO) e da Universidade de Twente, na Holanda. Em sua fase inicial, atua com a coleta de dados do Hospital Universitário Gaffrée Guinle (HUGG), no Rio de Janeiro, do Hospital Municipal São José, especificamente voltado durante a pandemia para atendimento de pacientes graves da COVID em Duque de Caxias, e do hospital Israelita Albert Einstein em São Paulo (VEIGA *et al*, 2021).

## 1.2 OBJETIVO

O objetivo deste trabalho é o tratamento dos dados disponibilizados pelo HUGG, disponibilizados em um Sistema Gerenciador de Banco de Dados (SGBD) PostgreSQL, e a partir de uma ferramenta de extração, transformação e carga (do inglês *extraction, transformation e loading* – ETL), gerar os formulários correspondentes e os armazenar no sistema de gerenciamento de banco de dados (SGBD) MySQL. Essas informações servirão de insumo para a geração dos dados FAIR que serão disponibilizados na rede VODAN.

É importante notar portanto, que o foco deste trabalho será na dinâmica das atividades envolvidas no preenchimento do formulário com as informações corretas, através da leitura e tratamento de informações, em sua maioria, não estruturadas.

Como cada hospital possui seus próprios padrões para armazenar as informações dos pacientes e suas evoluções, é necessário padronizar os dados de forma a facilitar os processos de triplificação, que faz parte da transformação e visa tornar os dados mais facilmente processáveis por programas, e transformação dessas informações alinhados aos princípios FAIR. O modelo do FRC-OMS é uma excelente ferramenta, pois está disponível mundialmente e padroniza a organização da informação de pacientes com COVID-19.

### 1.3 ESTRUTURA

Este trabalho é dividido da seguinte forma; o próximo capítulo detalha o projeto VODAN BR, com foco inicial na arquitetura do projeto, analisando em seguida o FRC-OMS.

No terceiro capítulo apresentamos o modelo de dados disponibilizado pelo HUGG, seguido por uma aprofundada análise da solução proposta, com a primeira e segunda versões do projeto detalhadamente explicadas. Os pontos de melhoria entre as versões, arquitetura da solução e principais transformações também serão debatidos, assim como a estrutura de resposta para as questões de cada módulo do FRC-OMS.

Por fim, no capítulo 4 concluímos abordamos as dificuldades encontradas durante o trabalho e finalizamos com uma análise dos resultados obtidos e apresentamos os próximos passos para expandir e aperfeiçoar o produto gerado por este trabalho de conclusão de curso.

## 2 A REDE VODAN E O PROJETO VODAN BR

A Rede VODAN foi criada como uma das atividades realizadas pela colaboração DATA TOGETHER, uma união das quatro maiores organizações mundiais de dados, composta por Committee on Data for Science and Technology (CODATA<sup>4</sup>), Research Data Alliance (RDA<sup>5</sup>), World Data System (WDS<sup>6</sup>), e GO FAIR, como mostra a Figura 4. De acordo com GO FAIR (2020), a iniciativa tem por objetivo otimizar o ecossistema de dados abertos a pesquisa em nível internacional e identificar oportunidades e necessidades de criação de infraestruturas de suporte a dados científicos emergentes, como foi o caso da pandemia da COVID-19 com a criação da rede VODAN.

**Figura 4:** Estrutura do projeto *Data Together*



Fonte: GO FAIR, 2020

O foco da rede VODAN é fornecer uma infraestrutura de fácil acesso a dados de prontuário alinhado aos princípios *FAIR* de pacientes vítimas de vírus de alto contágio (VEIGA *et al*, 2021). Desta forma, a rede fica disponível não só para casos de COVID-19, mas também para outros tipos de contágios. O objetivo é criar uma comunidade através da construção de uma infraestrutura de dados distribuídos internacionalmente e interoperáveis, que ofereça suporte a respostas baseadas em evidências para os surtos virais

---

<sup>4</sup> <https://codata.org/>

<sup>5</sup> <https://rd-alliance.org/>

<sup>6</sup> <https://world-datasytem.org/>

(MONS, 2020).

Como mencionado anteriormente, de acordo com Veiga *et al* (2021), o projeto VODAN BR é o braço brasileiro da rede VODAN e foi concebido através do esforço conjunto da Fiocruz, UFRJ, UNIRIO e Universidade de Twente. Tem como parceiros o HUGG, o hospital municipal São José de Duque de Caxias, e o Hospital Israelita Albert Einstein. O objetivo do projeto é colaborar com a iniciativa VODAN BR, atuando principalmente em sua primeira fase, com a coleta, transformação e tratamento dos dados, em conformidade com o FRC-OMS.

## 2.1 ARQUITETURA DO PROJETO

O projeto VODAN BR visa a disponibilização de uma plataforma para FAIRificação de dados clínicos, inicialmente com foco em COVID-19, mas com possível uso para outras enfermidades.

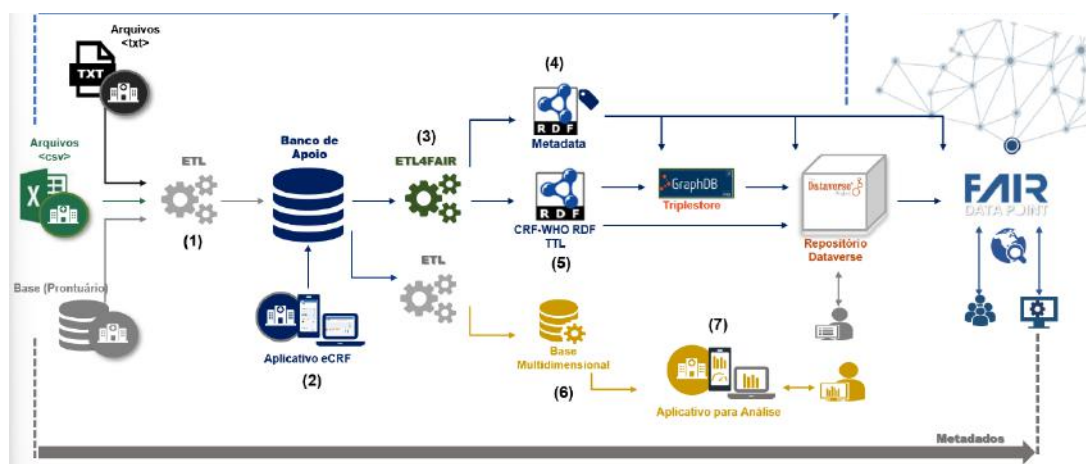
De forma resumida, a arquitetura da plataforma se inicia com um ETL que realiza a coleta dos dados de bancos, arquivos texto e CSV's, inserindo estes dados em um banco de dados de apoio, que também poderá ser alimentado pelo aplicativo eCRF. Os dados presentes neste banco são tratados pelo ETL4FAIR que realiza um processo de FAIRificação, tornando os dados e os metadados interligáveis. Por fim, os dados serão disponibilizados em triplestores e os metadados no FAIR DP. Importante definir que triplestores são bancos de dados que armazenam dados em formato de triplos RDF (Resource Description Framework), composto por sujeito, predicado e objeto. Eles permitem a armazenagem e consulta de grandes quantidades de dados estruturados

No decorrer desta subseção, iremos explicar em detalhes todos os passos descritos no resumo da arquitetura do projeto. Todas as menções às etapas da plataforma de FAIRificação estão referenciando a Figura 5, de arquitetura do projeto, onde foram inseridos números para cada passo, buscando simplificar o entendimento de cada etapa com o texto de explicação.

Para a coleta e tratamento desses dados clínicos, que serão integrados a rede VODAN, as três instituições de saúde disponibilizam seus dados de diversas maneiras. Como cada unidade de saúde tem autonomia para organizar suas informações, os dados podem ser fornecidos por XML, CSV, banco de dados ou por meio de outra tecnologia de armazenamento de informação, como arquivos textos.

Os dados precisam ser coletados e tratados para seu armazenamento no banco de apoio da Fiocruz. Este projeto seguirá o FRC-OMS. Para explicar a arquitetura do projeto vamos utilizar a Figura 5. Desta forma, as informações coletadas dos hospitais parceiros serão padronizadas em um banco de apoio, seguindo um modelo pré-determinado de organização da informação. Para esta coleta, será utilizado um ETL( número 1) customizado para este fim , foco deste trabalho.

**Figura 5:** Arquitetura do projeto VODAN



Fonte: CAMPOS *et al*, 2021

O Dados que já estejam no modelo da OMS podem ser inseridos diretamente no banco de apoio. O aplicativo eCRF VODANBR (número 2 da Figura 5) facilita a coleta dos dados e os organiza de acordo com o modelo. Esse aplicativo pode ser utilizado pelos hospitais para realizar o armazenamento das informações através do preenchimento do formulário de forma digital. (CAMPOS *et al*, 2021)

Os dados armazenados neste banco de apoio serão então tratados através do framework ETL4FAIR (número 3 da Figura 5), em um processo de FAIRificação. O



processo se inicia com uma análise dos dados e metadados que são tratados de acordo com modelo semânticos previamente definidos. Por fim, os dados e os metadados são interligáveis (números 4 e 5 respectivamente da Figura 5) (CAMPOS *et al*, 2021). De acordo com Jacobsen *et al* (2020), isto significa que os dados são transformados de forma a ser possível navegar pelos dados e metadados com objetivo a chegar em informações correlacionadas.

Além de disponibilizar os dados em tripletores e os metadados no FAIR DP, ponto de acesso para análise e visualização de dados tratados seguindo os princípios do FAIR Data, eles também serão disponibilizados em um banco multidimensional (número 6 da Figura 5), possibilitando um ponto de acesso para análises e visualizações dos dados tratados. Um aplicativo para análise dos dados será desenvolvido para visualização dos dados presentes na base multidimensional (número 7 da Figura 5).

Assim como descrito por Campos *et al* (2021), os dados são armazenados em um banco de apoio alinhado com as políticas FAIR e posteriormente exportados e tratados utilizando o modelo semântico para a representação do FRC-OMS. Desta forma, as informações coletadas e tratadas desde o fornecimento dos dados das instituições de saúde, passam a fazer parte da rede FAIR DP VODAN, o metadato passa a ser acessível de qualquer projeto e lugar do mundo, os dados estarão disponíveis aqueles que tiverem acesso.

## 2.2 FRC-OMS

Com a proliferação da COVID-19 em diversos países e a superlotação dos sistemas de saúde, a OMS percebeu que a estrutura dos dados dos pacientes internados por conta da doença era muito diferente ao redor do planeta. Com dados armazenados de forma heterogênea e com metodologias de acompanhamento diferentes, se tornou muito custoso analisar e interligar dados de diferentes fontes para aumentar o conhecimento sobre o vírus e sua doença.





que cercam estes tipos de dados. Há outros trabalhos de extração de informações clínicas que também se deparam com informações estruturadas e não estruturadas, porém, optaram por métodos de coleta diferentes.

O uso de machine learning adotado em FETTE *et al* (2012), mostra uma forma de coleta de dados não estruturados a partir de uma árvore terminologia, porém apresenta custos elevados para estruturar um modelo de aprendizado de máquina e pouco flexível em relação a novas informações que não estejam na árvore de terminologia gerada para o algoritmo do trabalho. Além disso, FETTE *et al* (2012) também tem como foco a coleta e inserção dos dados em um DW, não se preocupando em preencher um banco já existente, como neste trabalho, onde as informações são utilizadas para responder um questionário padronizado.

A extração de informações não estruturadas foi amplamente discutida em FORD *et al* (2016), assim como sua importância, onde novamente observamos as dificuldades encontradas em diversos métodos de extração de texto livre quando nos referimos a termos médicos. Em suma, até mesmo metodologias de NPL tradicionais podem apresentar novos erros decorrentes da diferença de escrita entre textos clínicos e convencionais.

É nítido, entretanto, que houveram diversos avanços nos mecanismos de análise de textos livres (não estruturados) como observado por MEYSTRE *et al* (2008), que faz um trabalho de análise dos métodos de análise de 1995 aos dias atuais, porém, evidencia ainda os problemas presentes atuais.

Muitas destas dificuldades foram encontradas neste trabalho, como a tradução de termos médicos, linguagem objetiva e escassa, dificultando a interpretação de termos de negação, além da grande variedade de situações e cenários. Algumas metodologias, inclusive combinadas, como visto em FORD *et al* (2016) podem ser implementadas neste trabalho, assim como discutido na seção 4.2 de trabalhos futuros.

### 3 SOLUÇÃO

Com base nas informações apresentadas no capítulo 2, observamos como requisitos relevantes do projeto: (i) a coleta de dados provenientes de prontuários eletrônicos; (ii) a limpeza e o tratamento desses dados transformando-os/adequando-os às questões de pesquisa clínica; (iii) a anonimização dos dados de pacientes, visando publicação futura com a devida proteção às informações de fórum pessoal; (iv) o tratamento mais assertivo possível às questões relacionadas a informações desestruturadas, como negação em texto livre.

Para atender os requisitos relevantes identificados, é necessário implementar uma solução que permita a coleta e tratamento dos dados brutos dos hospitais participantes do projeto, para o preenchimento dos FRC-OMS dos pacientes. Este trabalho tem por objetivo o tratamento dos dados de prontuário do HUGG, hospital universitário associado a UNIRIO e localizado no bairro da Tijuca, no Rio de Janeiro, Brasil. Para isso emprega uma amostragem de dados de pacientes disponibilizados pelo hospital. Cabe destacar a importância da participação dos especialistas do HUGG nas tomadas de decisão e no entendimento dos dados disponibilizados para o preenchimento do FRC-OMS.

Os dados do hospital foram disponibilizados através de um esquema específico no SGBD PostgreSQL que atende ao prontuário eletrônico do hospital. A separação dos esquemas de dados foi uma opção tomada por questões de segurança. Deste modo, os dados podem ser coletados sem interferir com as atividades do prontuário. Os dados tratados e transformados de acordo com o FRC-OMS são armazenados em um SGBD MySQL na FIOCRUZ. Com essa arquitetura, os registros gerados não ficam juntos ao banco de dados referente ao prontuário eletrônico dos pacientes.

Com essa estrutura estabelecida, precisamos de uma solução que permita a extração dos dados, provenientes do SGBD origem (PostgreSQL), sua transformação e inserção no SGBD destino (MySQL), ou seja, um ETL. Após pesquisa e a elaboração de alguns modelos, o Pentaho foi escolhido para ser a base da solução por ser uma ferramenta com facilidade de uso, flexibilidade, capacidade de integração com diversas fontes de dados (Know, 2022).

Este capítulo apresenta o modelo de dados disponibilizado pelo HUGG, onde examinamos as informações disponíveis para a ingestão do ETL e sua estrutura organizacional. Em seguida apresentamos uma análise sobre o Pentaho e sua importância dentro do projeto. Logo após, apresentamos e analisamos a primeira versão da solução, com as seções seguintes descrevendo

as principais transformações e por fim, os problemas encontrados nessa versão. Ao final, apresentamos a versão 2.0, iniciando pela sua arquitetura, seguido de suas principais transformações e funcionamento. Uma discussão sobre os resultados obtidos e os trabalhos futuros é apresentada no capítulo de conclusão.

### 3.1 MODELO DE DADOS DO HUGG

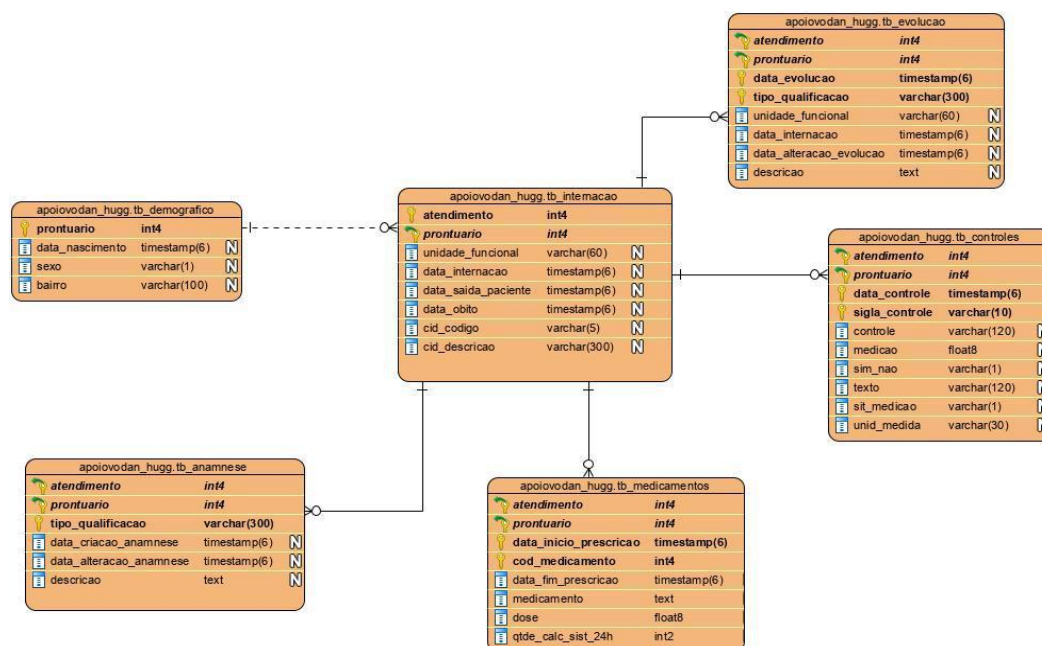
Para insumo de testes e suporte no desenvolvimento do ETL, foram extraídas do banco disponibilizado pelo HUGG as informações referentes à internação de 20 pacientes. Esta seção irá detalhar a estrutura dos dados disponibilizados, apresentando os dados disponibilizados para a elaboração das respostas do FRC-OMS.

O HUGG disponibilizou um esquema que contempla as seguintes tabelas: Anamnese (tb\_anamnese), Controles (tb\_controles), Demográfico (tb\_demografico), Evolução (tb\_evolucao), Internação (tb\_internacao) e Medicamentos (tb\_medicamentos). O modelo de dados com a estrutura das tabelas disponibilizadas, assim como seus relacionamentos, está representado na Figura 9.

Esse modelo tem por base a tabela Demográfica que apresenta as informações anonimizadas do paciente. A partir dessa tabela podemos acessar os atendimentos realizados que estão registrados na tabela internação, através do prontuário, que será utilizado como chave estrangeira entre as tabelas. Como a tabela de internação é central ao modelo de dados, associados a ela, com exceção da tabela Demográfica já comentada, temos : Evolução, responsável pelos registros de acompanhamento evolutivo do paciente; Controle, responsável pela medição de temperatura, batimentos cardíacos entre outros indicadores semelhantes; Medicamentos, que registra todos os medicamentos utilizados pelo paciente durante sua estadia e Anamnese, que possui as informações médicas da admissão do paciente. Um detalhamento de cada uma das tabelas é apresentado a seguir.

A tabela referente a Anamnese (tb\_anamnese) corresponde a uma entrevista inicial com o paciente, detalhando os sintomas declarados pelo paciente, buscando levantar informações que podem ser de grande importância na avaliação médica. Esta é a primeira etapa da qual o paciente passa, e por tanto, as informações presentes nesta tabela são fundamentais para o módulo 1 do questionário, que possui as questões referentes à admissão do paciente.

Figura 9: Estrutura disponibilizada pelo HUGG em SGBD PostgreSQL



A informação na tabela (tb\_anamnese) está dividida entre as colunas: atendimento, que identifica uma internação/atendimento do paciente, se repetindo para todos os dias nos quais o paciente ficou internado; prontuario, que é o identificador único de um paciente, e se mantém o mesmo caso haja novas visitas a unidade de saúde; data\_criacao\_anamnese e data\_alteracao\_anamnese que são referentes a data de entrada da informação no banco e a última vez que esta informação foi alterada, respectivamente; tipo\_qualificacao, que informa qual especialidade é responsável pelo registro, podendo ser Enfermagem, Medicina, Nutrição, Fisioterapia, Psicologia e Assistência Social; e por fim a descricao, que contém todas as informações referentes ao momento que o registro foi criado.

É importante frisar que as informações do campo descricao estão disponibilizadas de forma não estruturada, ficando o ETL responsável por obter as informações pertinentes para responder às perguntas do módulo 1 que dependem deste campo como fonte de dados. Os outros campos possuem informações estruturadas, como pode ser visto na Figura 10, que contém um exemplo dos registros encontrados na tabela.

**Figura 10:** Estrutura de dados da Anamnese

	Data Output	Explain	Messages	Notifications				
	atendimento integer		prontuario integer	data_criacao_anamnese timestamp without time zone	data_alteracao_anamnese timestamp without time zone	tipo_qualificacao character varying (300)	descricao text	
1	124000		78	2020-04-22 14:25:00.662	2020-04-22 14:27:54.309	NUTRIÇÃO	Paciente de 4...	
2	124456		78	2020-05-13 11:35:51.527	2020-05-13 11:51:22.646	ENFERMAGEM	# D 00 IH	
3	147299		78	2020-11-13 15:14:52.086	2020-11-13 15:29:00.125	ENFERMAGEM		
4	124112		78	2020-04-29 15:24:42.16	2020-04-29 15:27:04.823	SERVICO SOCIAL	Contato realiz...	
5	131100		78	2020-08-19 08:57:34.656	2020-08-19 08:57:54.813	PSICOLOGIA	@	
6	147299		78	2020-11-13 14:07:39.643	2020-11-13 14:29:14.763	FISIOTERAPIA	FISIOTERAPIA...	
7	124026		78	2020-04-22 10:29:51.177	2020-04-22 10:35:40.809	SERVICO SOCIAL	Contato telefô...	
8	124500		78	2020-05-14 10:16:52.428	2020-05-14 10:20:47.272	NUTRIÇÃO	Paciente caso	

Visto que o HUGG possui um processo de admissão no qual a anamnese é realizada através da entrevista do paciente por diferentes profissionais, há muitas especialidades inserindo dados no banco. Deste modo, os dados descritos no campo de descrição estão relacionados à área descrita no campo qualificação, pois os responsáveis pelo registro são os profissionais dessa especialização. Logo a depender da pergunta que estamos analisando para preencher o FRC-OMS, a resposta pode estar presente dentro da informação não estruturada de uma especialidade específica. Na Figura 11 por exemplo, podemos encontrar informações relacionadas à dieta do paciente, pois ela foi derivada da entrevista inicial feita pelo profissional de nutrição.

**Figura 11:** Exemplo de informação descrita pela especialidade de nutrição

```
Paciente de 44 anos, admitida no setor para tratamento (infecção por coronavírus,
com relato de hipertensão.
Conduta: paciente recebendo dieta via oral de consistência semilíquida,
hipossódica, fracionada em 5 refeições diárias.
Segue em acompanhamento.
```

É importante frisar que as colunas `atendimento` e `prontuario` funcionam como chave estrangeira para o relacionamento das tabelas. O `atendimento` por sua vez pode ser diferente para o mesmo `prontuario`, visto que a cada visita do paciente a unidade de saúde, o número do atendimento é alterado. Na Figura 12 podemos ver um exemplo de registro de Anamnese.



**Figura 12:** Exemplo de informação registrada na tabela de Anamnese

```
# D 00 IH
# Sem comorbidades
# Nega alergias
# COVID ?
# Swab na UPA copacabana dia 08/05
# Morse 20
# Braden 19
```

Paciente admitido neste setor procedente da regulação (UPA Copacabana). Relata início dos sintomas há 20 dias com tosse seca, febre noturna fazendo uso de dipirona, inapetencia, amargo na boca, olfato alterado (cheiro ruim). Internou na UPA dia 08/05, sendo regulado e transferido hj.

Ao exame: Paciente lúcido, orientado no tempo e no espaço, cooperativo, dispneico em uso de O2 sob óculos nasal, febril, hipocorado +/4+, acianótico, anictérico, tosse seca, refere baixa ingesta alimentar, mantendo acesso venoso periférico em MSD pervio e sem sinais flogísticos, Abdome flácido, indolor à palpação superficial e profunda, diurese espontanea em bom débito (sic), Funções intestinais presente nas ultimas 24h (ontem), MMII sem edema, panturrilhas livres.

Conduta: Monitorização multiparametrica nao invasiva, Swab nasal e retal, troca de acesso periférico.

A tabela de controle (`tb_controle`) possui medições do paciente ao longo do dia, permitindo acompanhar vários fatores que podem ser críticos em pacientes com COVID-19, como mostrado no exemplo da Figura 13. A estrutura da tabela possui informações sobre atendimento e prontuário, relativos ao atendimento e ao paciente, `datacontrole`, informando a data no qual o controle foi realizado, `medicao` e `unidmedida` que representam o valor obtido e a sua unidade de medida, `sim/nao` e `sitmedicao` que não serão utilizadas neste trabalho; `siglacontrole` e `controle`, que definem qual foi o fator medido naquele registro, podendo ser dos seguintes tipos: Alimentação por via oral; Amins via parenteral; Antibioticoterapia via parenteral; Avaliação de risco para quedas; Avaliação de risco para úlceras de pressão; Batimento cardíaco do feto; Curare; Drenagem sonda gástrica; Dreno abdominal esquerdo; Eliminação urinária; Emese; Escala de agitação e sedação de Richmond; Evacuação intestinal; Fração inspirada de o2; Frequência cardíaca; Frequência respiratória; Glicemia capilar; Hidratação venosa (eletrólitos se houver); Intensidade da dor; Medicação por via oral; Medicação via parenteral; Nutrição enteral; Nutrição parental; Outros drenos; Oximetria de pulso; Peso; Pressão arterial diastólica; Pressão arterial sistólica; Sedação via parenteral; Temperatura axilar.

Os registros desta tabela são gerados várias vezes ao dia para o mesmo paciente. Dessa forma teremos centenas de registros para uma dezena de pacientes, internados por alguns dias. Como muitas dessas informações não mudam em intervalos pequenos de tempo, teremos que tratar a grande quantidade de informações na elaboração da solução, para filtrar representações fiéis ao estado do paciente em um determinado período de tempo.

Seguindo as orientações da equipe médica do HUGG, foi escolhido utilizar apenas as informações do primeiro controle feito após as 12:00 horas.

**Figura 13:** Exemplo da estrutura da tabela de controle

	atendimento integer	prontuario integer	data_controle timestamp without time zone	sigla_controle character varying (10)	controle character varying (120)	medicao double precision	sim_nao character varying (1)
1	152453	71	2020-12-21 18:00:00	PAD	PRESSAO ARTERIAL DIASTOL...		73 [null]
2	153456	71	2021-01-12 09:58:00	SAO2	OXIMETRIA DE PULSO		64 [null]
3	150874	71	2020-12-11 12:00:00	HGT	GLICEMIA CAPILAR		117 [null]
4	124000	71	2020-04-21 17:05:00	FC	FREQUENCIA CARDIACA		81 [null]
5	151602	71	2020-12-15 01:54:00	FC	FREQUENCIA CARDIACA		80 [null]
6	130511	71	2020-08-14 16:00:00	PAD	PRESSAO ARTERIAL DIASTOL...		120 [null]
7	152453	71	2020-12-21 00:02:00	HGT	GLICEMIA CAPILAR		107 [null]
8	150874	71	2020-12-10 12:00:00	FR	FREQUENCIA RESPIRATORIA		30 [null]
9	150874	71	2020-12-09 22:01:00	ORAL	ALIMENTACAO POR VIA ORAL		300 [null]
10	124000	71	2020-04-26 14:00:00	PAD	PRESSAO ARTERIAL DIASTOL...		78 [null]
11	150874	71	2020-12-14 08:00:00	FR	FREQUENCIA RESPIRATORIA		20 [null]
12	124000	71	2020-05-12 18:00:00	SAO2	OXIMETRIA DE PULSO		98 [null]
13	150874	71	2020-12-05 12:00:00	Tax	TEMPERATURA AXILAR		36.2 [null]

A tabela demográfica (*tb\_demografico*) apresenta informações referentes aos dados demográficos do paciente, como *sexo*, *data\_nascimento* e *bairro* como apresentado na Figura 14. Além destes 3 campos, também se fazem presentes o *prontuário* e *atendimento*, identificando o paciente e individualmente seu atendimento, assim como feito nas outras tabelas do banco. Essas informações serão importantes, principalmente para o primeiro módulo do FRC-OMS, como será visto no desenvolvimento da solução.

**Figura 14:** Exemplo da estrutura da tabela de dados demográficos

	prontuario integer	data_nascimento timestamp without time zone	sexo character varying (1)	bairro character varying (100)
1	71	2000-10-23 00:00:00	F	PACIÊNCIA
2	71	1979-11-12 00:00:00	F	CAMPO GRANDE
3	71	1951-09-06 00:00:00	M	COPACABANA
4	71	1954-01-02 00:00:00	F	TIJUCA
5	71	1969-03-16 00:00:00	M	[null]
6	61	1960-01-31 00:00:00	F	VILA ISABEL
7	71	1984-06-05 00:00:00	F	PILARES
8	71	1951-02-23 00:00:00	M	BOTAFOGO
9	71	1961-09-27 00:00:00	M	SANTO CRISTO
10	71	1991-11-04 00:00:00	F	TAUÁ
11	71	1975-07-04 00:00:00	F	PORTUGUESA
12	71	1985-09-12 00:00:00	M	RICARDO DE ALBUQUERQUE

A tabela de evolução (*tb\_evolucao*) possui informações gerais de monitoramento

do paciente. A Figura 15 mostra um exemplo de como esses dados são disponibilizados. De modo similar a anamnese, diferentes profissionais registram as evoluções observadas no paciente. Deste modo, a tabela apresenta evolução de fisioterapia, de enfermagem, de médicos e etc, com um acompanhamento diário dos internados, criando uma descrição de seu estado e evolução ao longo dia. Os pacientes que estejam alocados em centros de tratamento intensivo (UTI) possuem um acompanhamento mais rígido, com vários registros sendo criados para cada especialidade em um único dia.

Os dados presentes nesta tabela são importantes para o módulo 2 do questionário, pois detém informações sobre toda a estadia do paciente. Além disso, esses dados estão organizados de forma não estruturada, ou seja, a informação se encontra presente em texto livre. Desta forma a tabela irá servir de fonte de informação para a maior parte das questões do módulo 2. Quais questões serão respondidas e os métodos utilizados, serão detalhados mais adiante no capítulo. Como cada profissional de saúde tem autonomia de escrever suas percepções do paciente de forma individual no campo descrição, a informação não segue uma estrutura pré-definida.

**Figura 15:** Exemplo da estrutura da tabela de evolução

	atendimento Integer	prontuario Integer	unidade_funcional character varying (60)	data_internacao timestamp without time zone	data_evolucao timestamp without time zone	data_alteracao_evolucao timestamp without time zone	tipo_qualificacao character varying (300)	descricao text
1	126756	130511	UNIDADE DE TERAPIA INTE...	2020-07-09 15:28:29.372	2020-07-16 13:23:07.339	2020-07-16 13:23:51.384	NUTRIÇÃO	Recebendo di...
2	130511		ENFERMARIA - COVID	2020-08-14 10:06:48.325	2020-08-20 11:15:00.568	2020-08-20 11:18:26.54	NUTRIÇÃO	Paciente rece...
3	165838		UNIDADE DE TERAPIA INTE...	2021-03-14 01:08:00	2021-03-15 11:40:05.94	2021-03-15 12:00:09.152	FISIOTERAPIA	Evolução Fisio...
4	124000		ENFERMARIA - COVID	2020-04-19 22:41:00	2020-05-22 22:44:19.935	2020-05-22 22:48:22.198	ENFERMAGEM	# IH: 19/04
5	139340		ENFERMARIA - COVID	2020-09-30 15:04:04.584	2020-10-04 10:23:13.753	2020-10-04 10:25:05.522	MEDICINA	EVOLUÇÃO M...
6	153456		ENFERMARIA - COVID	2021-01-04 08:12:21.651	2021-01-06 09:18:49.036	2021-01-06 09:24:48.696	ENFERMAGEM	# IH 04/01/20...
7	151602		ENFERMARIA - COVID	2020-12-09 18:37:00	2020-12-13 13:26:22.622	2020-12-13 13:28:01.575	NUTRIÇÃO	Paciente em tr...
8	124456		ENFERMARIA - COVID	2020-05-13 10:05:28.347	2020-06-01 09:35:36.177	2020-06-01 09:41:41.026	MEDICINA	# IH 08/05
9	150874		ENFERMARIA - COVID	2020-12-04 22:30:00	2020-12-14 23:52:50.768	2020-12-14 23:57:40.694	ENFERMAGEM	65 anos
10	146558		ENFERMARIA - COVID	2020-11-10 11:12:00	2020-11-12 10:52:02.037	2020-11-12 11:05:08.254	ENFERMAGEM	SDC, 12/11/2...
11	153456		ENFERMARIA - COVID	2021-01-04 08:12:21.651	2021-01-10 15:00:59.791	2021-01-10 15:10:11.926	NUTRIÇÃO	=> Paciente fe...
12	124000		ENFERMARIA - COVID	2020-04-19 22:41:00	2020-05-19 11:10:18.447	2020-05-19 11:11:10.599	MEDICINA	Paciente segu...
13	126756		UNIDADE DE TERAPIA INTE...	2020-07-09 15:28:29.372	2020-07-14 20:38:34.026	2020-07-14 20:50:57.969	FISIOTERAPIA	# PLANTÃO FL...
14	124456		ENFERMARIA - COVID	2020-05-13 10:05:28.347	2020-05-21 12:42:12.061	2020-05-21 12:46:12.407	MEDICINA	# Plantão Qu...
15	124500		UNIDADE DE TERAPIA INTE...	2020-05-14 17:35:13.384	2020-05-19 10:08:44.438	2020-05-19 14:02:37.637	MEDICINA	EVOLUÇÃO TE...

A tabela de evolução (tb\_evolucao) é constituída por 8 colunas, sendo as duas primeiras de atendimento e prontuário, para identificar cada registro; A unidade\_funcional identifica se o paciente se encontra em um centro de tratamento intensivo ou na enfermaria. A data\_evolucao e data\_alteracao\_evolucao são para organização temporal dos registros; O tipo\_qualificacao identifica a especialidade responsável pelo acompanhamento e a descricao, possui informações não estruturadas relativas à evolução e situação do paciente naquele momento.

A tabela de internação (`tb_internacao`) possui os seguintes campos: `data_internacao` e `data_saida_paciente` (ou `data_obito` caso o paciente venha a falecer) que representam respectivamente a entrada e saída do paciente na instituição de saúde; a `unidade_funcional` que informa de qual unidade funcional o paciente foi admitido/liberado ao final da internação. A Figura 16 mostra um exemplo de alguns destes registros. Cada registro também é identificado pela dupla de `atendimento` e `prontuario`. Esses dados são importantes para o último módulo do FRC-OMS.

**Figura 16:** Exemplo da estrutura da tabela de internação

	<code>atendimento</code> integer	<code>prontuario</code> integer	<code>unidade_funcional</code> character varying (60)	<code>data_internacao</code> timestamp without time zone	<code>data_saida_paciente</code> timestamp without time zone	<code>data_obito</code> timestamp without time zone
1	131100	700036	ENFERMARIA - COVID	2020-08-18 15:53:00	2020-08-19 10:00:00	[null]
2	130511	700009	ENFERMARIA - COVID	2020-08-14 10:06:48.325	2020-08-20 09:54:00	[null]
3	153456	700007	ENFERMARIA - COVID	2021-01-04 08:12:21.651	2021-01-12 20:21:00	2021-01-12 20:00:00
4	124242	700035	ENFERMARIA - COVID	2020-05-02 17:09:06.253	2020-05-04 12:30:00	[null]
5	125666	700038	ENFERMARIA - COVID	2020-06-26 15:41:38.992	2020-06-30 15:00:00	[null]
6	124456	700007	ENFERMARIA - COVID	2020-05-13 10:05:28.347	2020-06-10 17:00:00	[null]
7	139340	800000	ENFERMARIA - COVID	2020-09-30 15:04:04.584	2020-10-06 15:30:00	[null]
8	150874	700006	ENFERMARIA - COVID	2020-12-04 22:30:00	2020-12-15 16:26:00	[null]
9	124676	700035	ENFERMARIA - COVID	2020-05-23 14:33:00	2020-06-02 13:00:00	[null]
10	156607	700039	ENFERMARIA - COVID	2021-01-25 16:31:59.133	2021-01-28 11:45:00	[null]
11	124112	700035	ENFERMARIA - COVID	2020-04-26 11:00:00	2020-05-01 15:05:00	[null]
12	124000	700000	ENFERMARIA - COVID	2020-04-19 22:41:00	2020-05-24 12:16:00	[null]
13	165838	700038	UNIDADE DE TERAPIA INTE...	2021-03-14 01:08:00	[null]	[null]
14	152453	700004	ENFERMARIA - COVID	2020-12-16 10:29:16.694	2020-12-24 12:10:00	[null]

A tabela de medicamentos (`tb_medicamentos`) apresenta informações sobre remédios receitados aos pacientes durante o período de internação, como exemplificado na Figura 17. Ela possui as seguintes colunas: dados que identificam o tipo da medicação utilizada (`medicamento`), sua dosagem (`dose`), início (`data_inicio_prescicao`) e fim da prescrição (`data_fim_prescicao`), além da unidade de medida (`unid_medida`) e sigla do medicamento (`sigla_med`). Para entender os dados desta tabela, os médicos e profissionais parceiros do HUGG disponibilizaram um mapeamento dos medicamentos adotados, que pode ser visto no material no anexo B. Essas informações ajudam a entender juntamente com as outras tabelas, quais foram as ações escolhidas pelos profissionais de saúde durante o tratamento e seus efeitos. Além disso, também foi possível identificar o tipo de medicamento utilizado, se foi um antibiótico, um anti viral ou outro tipo de medicação mapeada.

**Figura 17:** Exemplo da estrutura da tabela de medicamentos

atendimento integer	prontuario integer	data_inicio_prescricao timestamp without time zone	data_fim_prescricao timestamp without time zone	cod_medicamento integer	medicamento text	dose double precision
126756	760006	2020-07-19 14:02:59.303	2020-07-20 14:00:00	281190	HidroCLOROTiazid...	25
124112	760005	2020-04-28 15:36:04.506	2020-04-29 11:00:00	14826	DIPIRONA AMP 2 ...	1
124112	760005	2020-04-26 18:11:25.97	2020-04-27 11:00:00	271851	OSELTAMIVIR CAP	75
124500	760002	2020-05-18 14:00:00	2020-05-19 14:00:00	400903	INSULINA HUMAN...	1
125666	760008	2020-06-28 20:58:56.691	2020-06-29 14:00:00	14826	DIPIRONA AMP 2 ...	1
147299	760009	[null]	[null]	[null]	[null]	[null]
124242	760005	2020-05-03 17:09:09.298	2020-05-04 14:00:00	223190	CLOROQUINA	450
139340	510000	2020-10-02 22:24:13.71	2020-10-03 14:00:00	148261	LACTULOSE XPE - ...	15
125666	760008	2020-06-27 14:00:00	2020-06-28 14:00:00	242853	ENOXAPARINA SO...	40
124000	760009	2020-05-04 14:00:00	2020-05-05 13:15:36.007	218480	MEROPENEM PO L...	2000
126756	760006	2020-07-23 14:00:00	2020-07-24 14:00:00	17680	PrednISONA COMP	20
125666	760008	2020-06-26 23:13:34.001	2020-06-27 14:00:00	281190	HidroCLOROTiazid...	25
151602	760005	2020-12-11 14:00:00	2020-12-12 14:00:00	264202	INSULINA HUMAN...	12

### 3.2 ORIENTAÇÕES DA EQUIPE TÉCNICA DO HUGG

No decorrer deste trabalho, esbarramos com diversas questões técnicas para a interpretação dos dados disponibilizados pela equipe do HUGG. Para sanar essas questões, uma equipe de profissionais do HUGG envolvidos nos processos de tratamento e internação se mostrou disponível e esclareceu como manipular e tratar os dados de forma a representar da maneira mais precisa possível o que foi descrito pelos diferentes profissionais de plantão.

As orientações e direcionamentos passados por eles foram a base para que as transformações e a arquitetura do projeto fosse formada. Por este motivo, antes de entrarmos no uso da ferramenta Pentaho Data Integration (PDI) vamos revisar as premissas estabelecidas para interpretar os dados descritos na subseção anterior referente ao modelo de dados do HUGG.

A primeira premissa estabelecida foi a divisão das qualificações em Medicina, Enfermagem, Fisioterapia, Serviço Social, Nutrição, Fonoaudiologia, Terapia Ocupacional e Psicologia. Os tipos referentes à Medicina, Enfermagem e Fisioterapia foram definidos como prioritários, pois possuem as informações relevantes para responder as perguntas do FRC-OMS.

É de suma importância esclarecer que, no decorrer deste trabalho, não foi realizada nenhuma filtragem em relação ao tipo de especialidade, a fim de garantir a maior amplitude

de informações possíveis. Dessa forma, o ETL (Extract, Transform, Load) construído buscou de maneira abrangente, nas fontes de informação disponíveis (Evolução, Anamnese, Medicamentos, entre outras), os dados conforme o cadastrado na ferramenta utilizada. Essa abordagem foi adotada tendo em vista que não é possível assegurar que as informações de interesse estejam necessariamente em um único tipo de especialidade. Logo, a opção de não limitar as especialidades analisadas no ETL se justifica pela busca pela obtenção de um conjunto amplo de informações relevantes para a pesquisa.

As decisões para cada um dos módulos do FRC-OMS são apresentadas a seguir. Para o módulo de admissão haverá um registro por atendimento, sendo que: (i) os resultados de exames serão referentes a primeira coleta após a admissão; (ii) os dados de medicação serão os ministrados no primeiro dia; e (iii) o primeiro controle será registrado como sintomas da admissão. Para o módulo de desfecho haverá apenas um registro de alta/óbito/transferência por atendimento, contemplando as informações da conclusão do atendimento. Para o módulo de acompanhamento haverá um registro por dia de internação referente ao atendimento, contemplando (i) as informações de sinais vitais e controles obtidas às 12:00 ou a próxima após este horário (caso não haja registro às 12:00) e (ii) as informações de medicamentos administrados durante o dia.

Para facilitar a obtenção de informações disponíveis nos questionários, houve diversas reuniões do projeto VODAN, onde participaram os membros da equipe do HUGG e da área de TI do projeto. O foco destas reuniões foi discutir melhores abordagens para o tratamento dos dados provenientes dos prontuários e convertê-los para os formatos de questões/respostas estabelecido pelo FRC-OMS. A partir das atas das reuniões foi elaborado um documento com as orientações de obtenção de dados para cada conjunto de questões. Essas orientações estão alinhadas com o modelo de dados apresentado na última seção e podem ser encontrados no anexo A deste trabalho. Importante ressaltar que o anexo C foi utilizado para compreender os dados levantados.

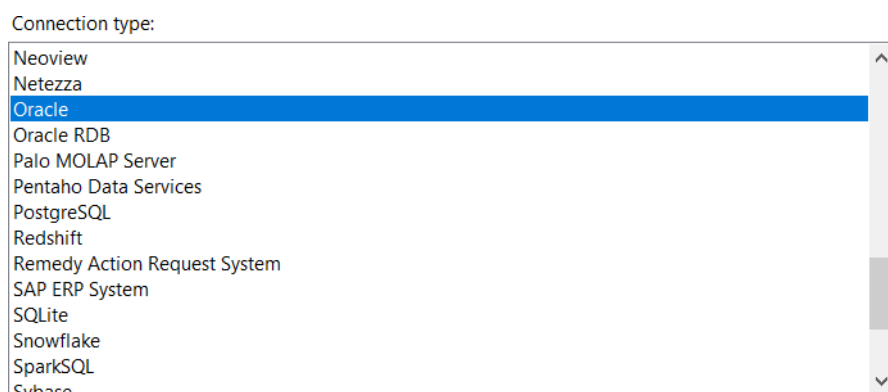
### 3.3 PENTAHO DATA INTEGRATION

O Pentaho se define como uma ferramenta de integração e análise de dados a nível

empresarial. (Target Trust, 2021) Em um cenário onde precisamos coletar e transformar uma grande quantidade de dados em pouco tempo, como o escopo deste trabalho, se faz necessário utilizar uma das muitas ferramentas disponíveis no mercado para agilizar o desenvolvimento e focar na implementação da solução.

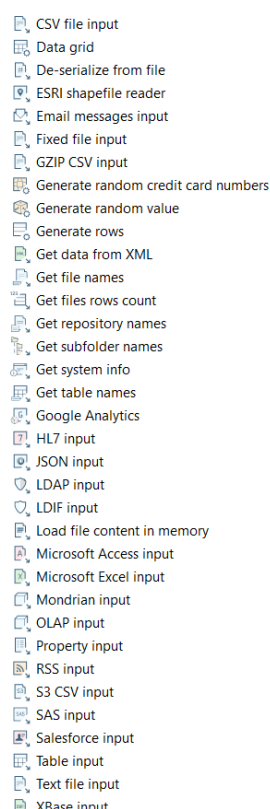
A ferramenta funciona a partir de transformações, que podem ser definidas como um conjunto de operações, interligadas em um fluxo, que são constituídos desde sua extração, tratamento e gravação. Há uma série de opções destas operações para serem escolhidas e interligadas. Geralmente a extração de dados é o primeiro passo das transformações construídas. O Pentaho nos dá uma série de opções para esse fim, como a extração a partir de bancos relacionais e não relacionais, e de fontes de informações em arquivo, como o xls, como ilustrado na Figura 18.

**Figura 18:** Exemplo das 30 opções de bancos como fontes para o Pentaho PDI.



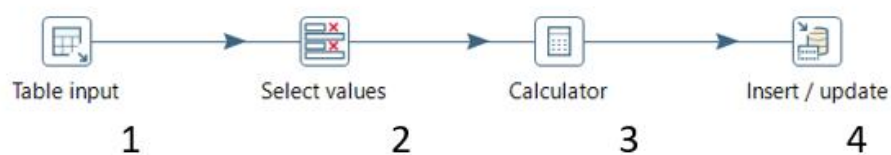
A partir da extração dos dados, realizamos uma série de transformações. Com as informações geradas efetuamos a gravação dos resultados em um destino propício. O Pentaho possui uma série de diferentes operações possíveis, como mostra a Figura 19. Para entender melhor, podemos partir de um exemplo prático. Em uma situação hipotética onde precisamos extrair uma informação de um banco de dados, efetuar um cálculo, como por exemplo a idade a partir da data de nascimento e armazenar o resultado em outra tabela, teremos 4 operações, conforme apresentado na Figura 20.

**Figura 19:** Série de opções de input de informações do Pentaho



Na Figura 20, a primeira operação será a de extração do dado (operação 1). Em seguida, filtramos as colunas que vamos utilizar e quais não serão necessárias. Nesse caso, apenas a data de nascimento seria mantida, com as outras colunas excluídas do fluxo. Posteriormente o cálculo da idade é realizado no Pentaho utilizando o calculator (operação 3). Assim temos acesso à data do sistema e podemos fazer a subtração das datas. Por fim, precisamos gravar o resultado (operação 4) de *insert/update*, garantindo que não iremos gravar dados repetidos no banco destino.

**Figura 20:** Exemplo do cálculo da idade a partir da data de nascimento no Pentaho.

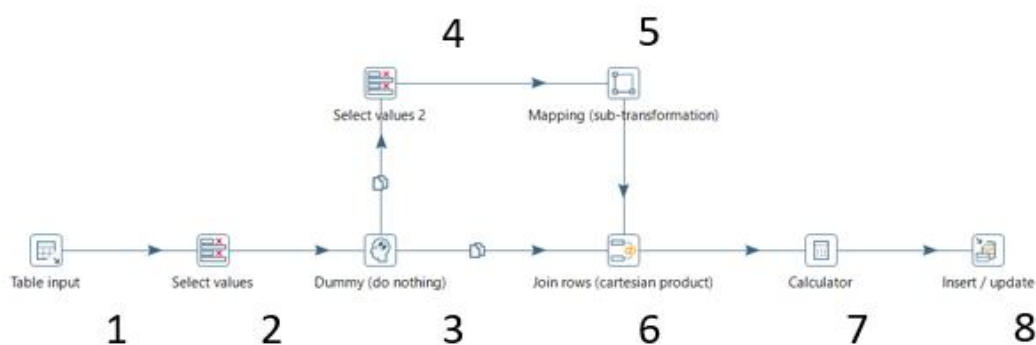




Conforme o número de operações aumenta, a complexidade da estrutura também. Em um exemplo um pouco mais sofisticado, podemos pensar em uma situação hipotética onde é necessário calcular o custo de várias compras, feitas em vários países, em datas diferentes, com moedas e cotações diferentes e obter o resultado em reais de cada uma delas.

Neste exemplo, que pode ser visto na Figura 21, precisamos primeiro fazer a extração das compras que vamos fazer o cálculo. Começamos com a extração dos dados pelo *table input* (operação 1), que opera de forma semelhante ao exemplo anterior. Seguimos pelo filtro dos dados (operação 2) que serão utilizados, pois nesse caso apenas precisamos das informações referentes as compras feitas. Em seguida passaremos por uma operação nova, o *dummy* (operação 3). O *Dummy* serve para ser uma operação nula, ela não executa nenhuma transformação nos dados, mas permite que eles sejam passados para uma ou mais operações. Neste caso, precisamos que os dados trafeguem em duas direções, para servir de input para a coleta dos câmbios nos dias das transações, e posteriormente para unir essa informação ao valor da compra e efetuar o cálculo.

**Figura 21:** Exemplo do cálculo da cotação em real de compras no exterior no Pentaho.



Para a coleta dos dados, podemos utilizar uma sub-transformação. Essa sub-transformação será importante pois teremos outras operações envolvendo a ação de coletar o câmbio, como o tratamento dos dados, consultas a outras tabelas, união de informações e seleção das informações pertinentes ao nosso objetivo. Assim, para não poluir a transformação principal com uma série de operações e fluxos que são pertinentes apenas a

coleta dessa informação, podemos criar essa sub-transformação. Ela é constituída de uma entrada esperada, que no caso será os países e datas das transações, seguido por uma série de operações e uma saída, para devolução dos dados a transformação de origem.

Na Figura 22 podemos visualizar um exemplo desta sub-transformação, com um input (operação 1), onde há o recebimento de parâmetros, seguida de uma busca em uma tabela no banco de dados (operação 2), um filtro (operação 3) e por fim, o retorno do processamento da sub-transformação (operação 4). A sub-transformação pode realizar uma série de operações entre seu início e seu fim.

**Figura 22:** Exemplo de sub-transformação.



Para alimentar essa sub-transformação, precisamos filtrar os dados (operação 4), que posteriormente serão enviados para ela. Seguiremos na Figura 21 para entender o restante do fluxo. Em seguida, teremos que unir o resultado da sub-transformação (operação 5) com os dados que restam no fluxo. Podemos fazer essa união a partir de uma operação de *join* (operação 6), que permite escolher quais os campos entre as duas fontes de informação deverão ser comparados para unir os registros. Desta forma podemos juntar os dados do câmbio com a compra específica feita naquele país e naquela data.

Por fim, com cada compra com seu respectivo câmbio e valor, podemos utilizar a operação de *calculator* (operação 7) para obter o valor da compra em reais. Essa informação é então gravada no banco com uma operação de *insert/update* (operação 8).

Os conceitos do Pentaho utilizados nesta seção, serão a base para as transformações utilizadas no desenvolvimento desta solução e que serão descritos no decorrer deste capítulo. Não abordaremos neste trabalho de conclusão de curso conceitos aprofundados da ferramenta Pentaho. Entretanto, a lógica das transformações e como elas funcionam e

se interligam para resultar nos FRC-OMS preenchidos a partir dos dados brutos disponibilizados pelo HUGG será explorada.

### 3.4 VERSÃO 1.0

A primeira versão desta solução foi planejada observando a arquitetura de tabelas e dados presentes na estrutura do banco de dados destino, hospedado na FIOCRUZ. É importante lembrar que o objetivo deste trabalho é extrair dados brutos provenientes do HUGG, tratá-los e criar FRC-OMS respondidos para cada paciente internado, considerando sua evolução, tratamento recebido e, se disponível, o desdobramento do tratamento em alta ou falecimento.

Assim como debatido no capítulo 2, apenas as tabelas em laranja da Figura 23 a seguir serão trabalhadas na solução. Desta forma, a primeira versão da solução se divide em 4 principais partes. A primeira parte, que iremos chamar de transformação central e que será detalhada na próxima seção, é responsável por preencher as tabelas relativas a informações sobre os questionários, excluindo as perguntas e respostas em si, ou seja, os metadados. A tabela `tb_QuestionGroupFormRecord`, que armazena as informações de resposta das perguntas do questionário, é respondida por outras três partes que são divididas nas transformações Q1, Q2 e Q3, respondendo respectivamente os módulos 1, 2 e 3 do questionário.

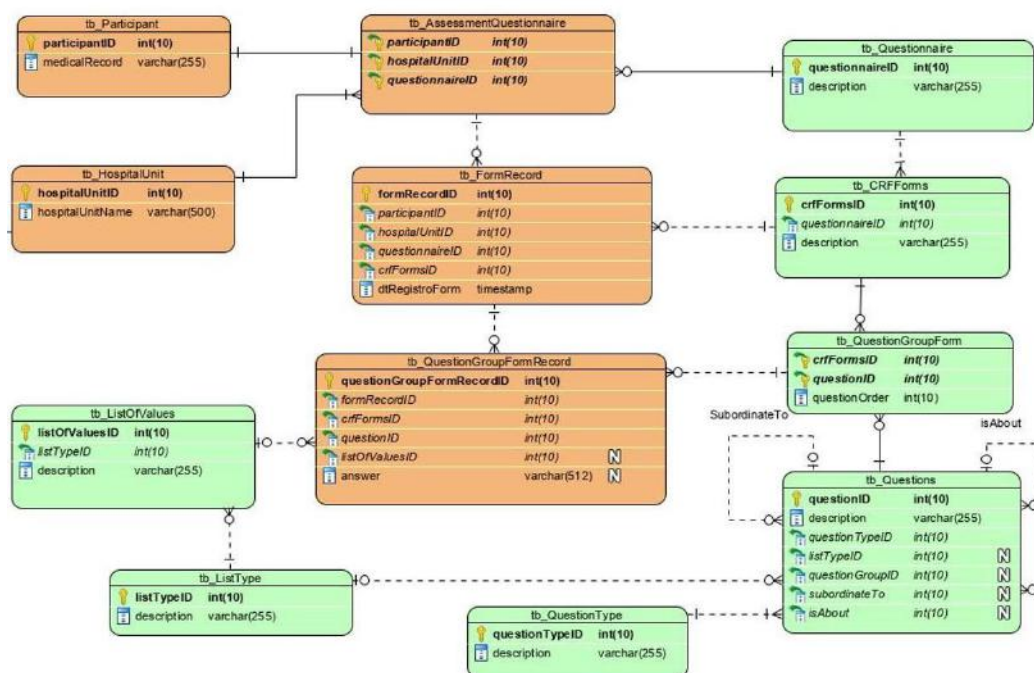
No decorrer do capítulo iremos detalhar a solução que preenche e interage com as tabelas da Figura 23. Por este motivo é importante resumir o funcionamento de cada uma de forma breve, para que as operações desenvolvidas estejam de acordo com o esperado no resultado final gravado no banco de apoio.

A tabela `tb_QuestionGroupFormRecord` explicada anteriormente é relacionada as tabelas `tb_FormRecord`, que armazena a informação de cada módulo do questionário para um determinado paciente e a `tb_questionGroupForm`, que interliga as informações da questão a resposta do formulário, sendo que neste caso vale frisar que as tabelas `tb_ListType`, `tb_QuestionType` e `tb_ListOfValues` são tabelas de

suporte, que detalham informações de questões múltiplas escolhas e a tipagem da questão, além disto a `tb_ListOfValues` será utilizada mais a frente para responder questões multipla escolhas, pois possuem a resposta padronizadas para cada tipo de questão multipla escolha.

A tabela `tb_CRFForms` detalha informações de módulo para um determinado questionário, neste caso, o FRC-OMS possui 3 módulos sempre. Essa tabela é interligada a `tb_Questionnaire` que define um questionário, por sua vez ligada a `tb_AssessmentQuestionnaire` que é o registro de um FRC-OMS para um determinado paciente. Além das tabelas que registram o questionário, o modelo apresenta `tb_Participant` que possui informações sobre o paciente e `tb_HospitalUnit` que concentra as informações sobre o hospital.

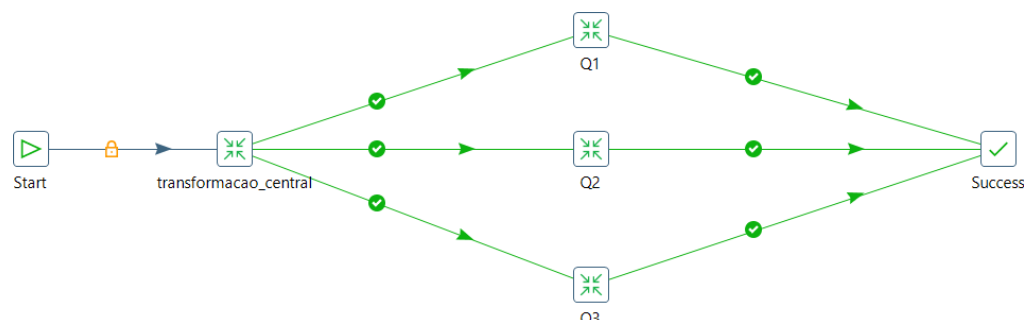
**Figura 23:** Visão geral das tabelas envolvidas na operação do ETL.



Dessa forma, temos uma primeira solução que pode ser representada pela arquitetura da Figura 24. Uma transformação central, responsável pela coleta dos pacientes e preenchimento das informações que descrevem os módulos e as respostas que serão computadas. E três outras transformações, executadas de forma paralela e que são responsáveis pela extração, transformação e resolução de cada resposta do módulo

correspondente. Assim a transformação Q1 irá responder às perguntas do primeiro módulo, Q2 do segundo módulo e Q3 do terceiro e último módulo.

**Figura 24:** Arquitetura da primeira solução, com o funcionamento das transformações não detalhadas



Cada uma das transformações da arquitetura possui uma lógica interna, eventualmente composta por sub-transformações. Cada uma delas será detalhada nas próximas seções, com explicações de seu funcionamento, sub-transformações e objetivo.

### 3.4.1 Transformação Central

A transformação central é o núcleo do funcionamento desta solução. Ela é responsável por inicializar as outras transformações que irão gerar as respostas para os módulos do FRC-OMS. Como ela é bastante extensa, possuindo mais de 40 operações, vamos dividir a sua explicação em três partes.

A primeira parte da transformação central, que pode ser vista na Figura 25, é composta de uma operação de coleta de dados de internação (operação 1), que coleta as informações dos pacientes que serão tratados durante a execução do ETL. Nesse momento, dados como o prontuário e a data de entrada e saída são coletados. Em seguida há duas operações, uma de *sort rows* (operação 2), que faz a ordenação dos dados, e outra de *unique rows* (operação 3), para eliminar dados duplicados.

Vale frisar que em vários pontos da solução essas operações de ordenação e *unique* estarão juntas pois o "*unique rows*" do Pentaho apenas consegue eliminar registros duplicados quando há uma ordenação pela chave de filtro anteriormente. Ou seja, se vamos manter um campo específico para cada registro, eliminando os registros com o campo

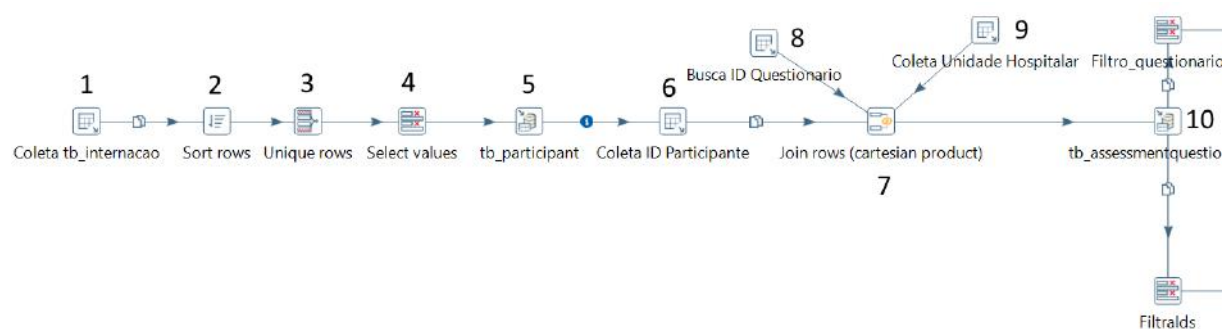
duplicado, precisamos ordenar os registros pelo campo em questão.

Uma operação de *select values* (operação 4) seleciona apenas o prontuário, eliminando os outros dados que não serão utilizados. Com os prontuários filtrados, os dados são inseridos na tabela de *tb\_Participant* (operação 5), sendo que o prontuário é inserido na coluna *MedicalRecord*. Em outros pontos da solução, quando precisarmos relacionar uma informação de outra tabela do banco MySQL da FIOCRUZ com o paciente, iremos usar esta tabela.

A próxima operação de coleta do Id do participante (operação 6), pois apesar de possuímos os valores dos prontuários de cada paciente, não temos o id que foi gerado automaticamente pelo banco de dados da FIOCRUZ. Esse dado será utilizado para correlacionar os dados com outras tabelas do BD. Cada id de paciente é mergeado com o retorno de duas novas buscas, a Busca ID Questionário (operação 8) e a Coleta Unidade Hospitalar (operação 9). A primeira busca qual o id de questionário está atribuído ao FRC-OMS, enquanto a segunda colhe o id atribuído ao HUGG na tabela de hospitais.

A operação de merge (operação 7) une a informação das três fontes, ou seja, cada paciente terá um registro com id do paciente, id do FRC-OMS e id do hospital HUGG. Esses dados são passados para uma operação de *insert/update* (operação 10) que insere os dados no banco na tabela *tb\_assessmentquestionnaire*.

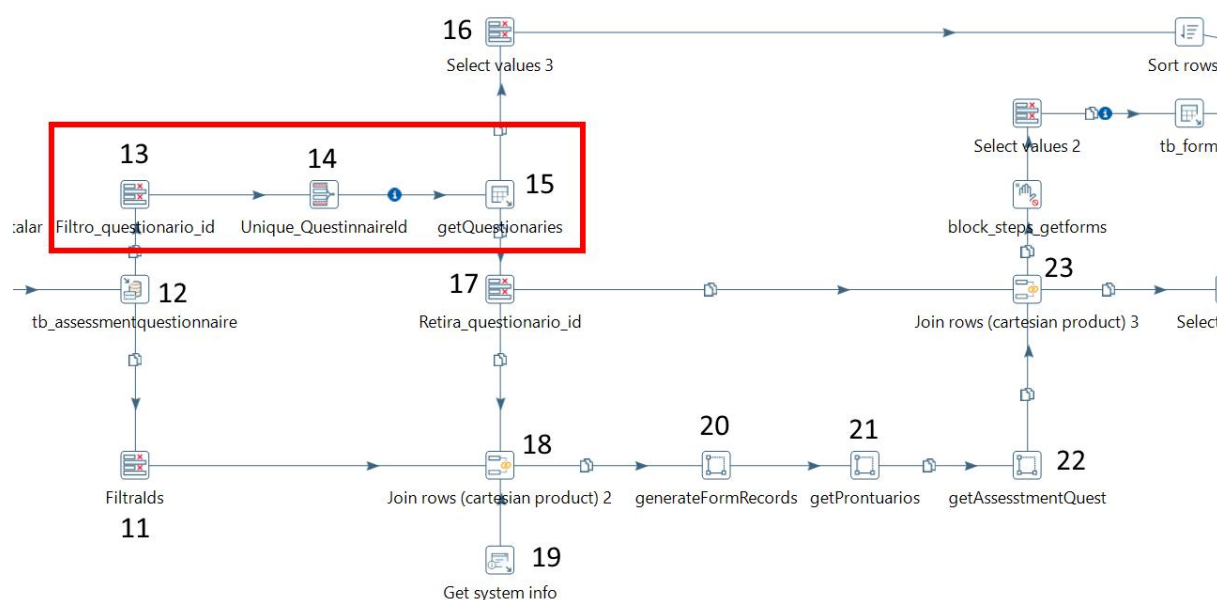
**Figura 25:** Primeira parte da transformação central.



A segunda parte da transformação central, visto na Figura 26, se inicia a partir do *insert/update* (operação 12) feito na tabela *tb\_assessmentquestionnaire*. As informações desta operação são subdivididas em dois fluxos, ambos para operações de

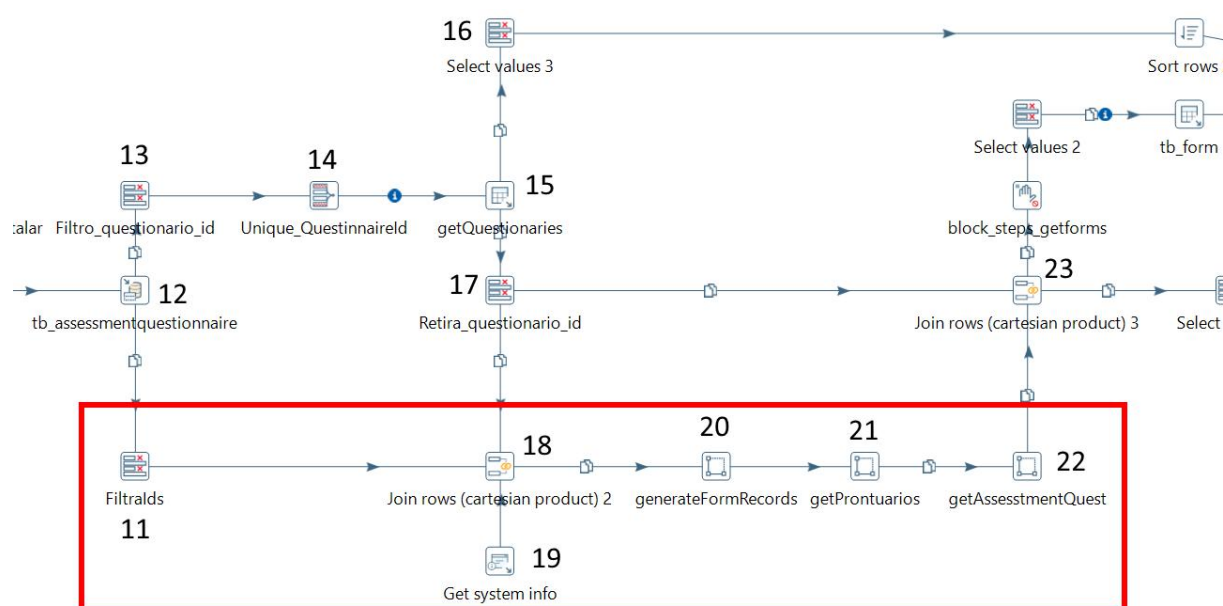
seleção de campos. A primeira (operação 13), seleciona o id do questionário, que é filtrado, para reduzirmos o número de registros a 1 (operação 14), e buscamos os dados de *crfFormsID* da tabela *tb\_crfforms* através deste id de questionário (operação 15). O *crfFormsID* é uma chave estrangeira que interliga as questões ao questionário, através da tabela *tb\_QuestionGroupForm*. Essa informação será necessária mais a frente, para coletarmos as questões que serão respondidas e para separarmos as questões em módulos 1, 2 e 3.

**Figura 26:** Segunda parte da transformação central com etapas da busca por *crfFormsID* em evidência.



Seguindo para a Figura 27, os dados do *insert/update* feito na tabela *tb\_assessmentquestionnaire* também seguem para uma seleção de campos onde apenas as informações de ids de hospital (operação 11), questionário e pacientes são mantidos. Essas informações são unidas com o *crfFormsID* obtido no fluxo superior e a data atual do sistema (operação 18). Esses dados são então passados para um conjunto de 3 sub-transformações (operações 20, 21 e 22), sendo que as duas últimas são alimentadas com as informações da sub-transformação anterior, seguindo um fluxo de execução. Cada uma destas sub-transformações será explicada em detalhes em seguida.

**Figura 27:** Segunda parte da transformação central com etapas de sub- transformações em evidência.



Ainda na Figura 27, a sub-transformação *generateFormRecords* (operação 20) irá criar registros para os módulos 1,2 e 3. É importante lembrar que há um registro de módulo 2 para cada dia de internação do paciente, e um registro apenas para os módulos 1 e 3, representando sua admissão e sua alta/óbito respectivamente.

A primeira parte da transformação consiste em buscar os prontuários correspondentes aos ids de pacientes recebidos no input da transformação. Como precisamos manter as informações recebidas, adicionando mais um elemento a cada registro trabalhado, vamos efetuar uma busca seguida de uma união.

Vale frisar que o Pentaho trabalha com fluxos de registros, logo temos um conjunto de registros que são transportados através das operações. Para acrescentarmos mais uma informação ao registro, precisamos dividir o fluxo em dois. Um fluxo irá ser utilizado como input na query, retornando um conjunto de prontuário e paciente, enquanto outro fluxo manterá as informações recebidas na requisição.

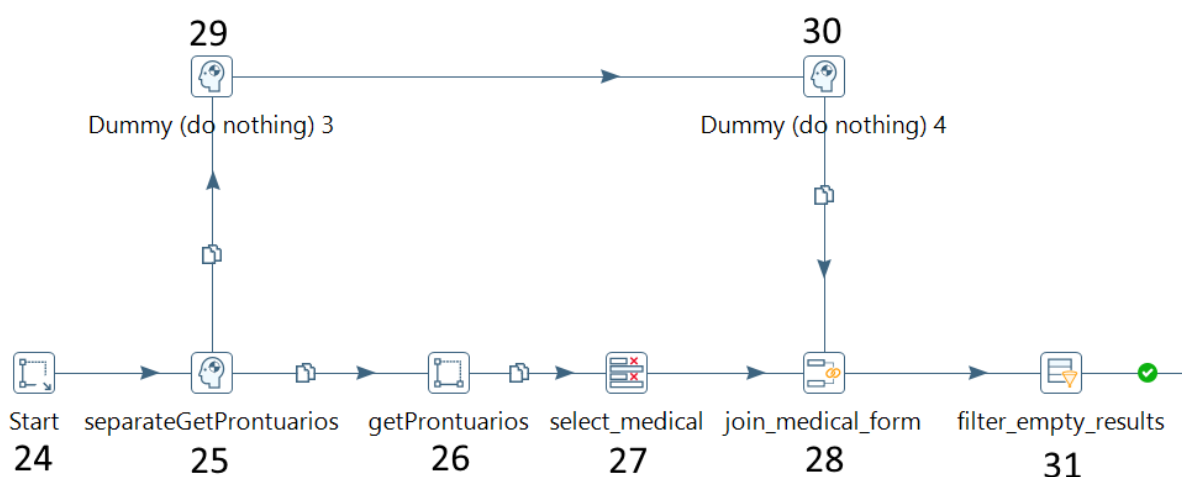
Com o retorno da *query*, podemos unir os fluxos com um *join* e garantir através de uma condição de união que as informações de prontuário só serão agregadas ao registro quando o id de paciente retornado na *query* for igual ao id de paciente a ser mergeado.



A Figura 28 mostra os passos descritos anteriormente. A sub-transformação se inicia na operação *START* (operação 24), seguido pelo *dummy separateGetProntuarios* (operação 25) que separa o fluxo de registros em dois. O primeiro fluxo segue para a sub-transformação *getProntuários* (operação 26), que faz a coleta dos prontuários dos ids de pacientes recebidos, conforme explicitado na Figura 29. O retorno da *getProntuarios* traz consigo a dupla prontuário e id de paciente que são combinados no *join\_medical\_form* (operação 28) com o segundo fluxo, que não sofreu alterações em seus registros pelas operações 6 e 7.

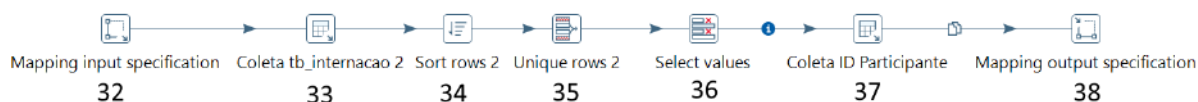
É importante ressaltar que a união só ocorre para os casos em que o id de paciente recebido da *getProntuários* (operação 26) for igual ao id de paciente do fluxo que contém o restante das informações. Desta forma ao chegarmos na operação *filter\_empty\_results* (operação 31) com os dados do início da sub-transformação, unidos ao prontuário do respectivo paciente. Nessa operação de filtro, eliminamos os registros que porventura possam ter vindo sem os dados corretos e seguimos para a segunda parte da sub-transformação.

**Figura 28:** A sub-transformação *createFormRecords*.



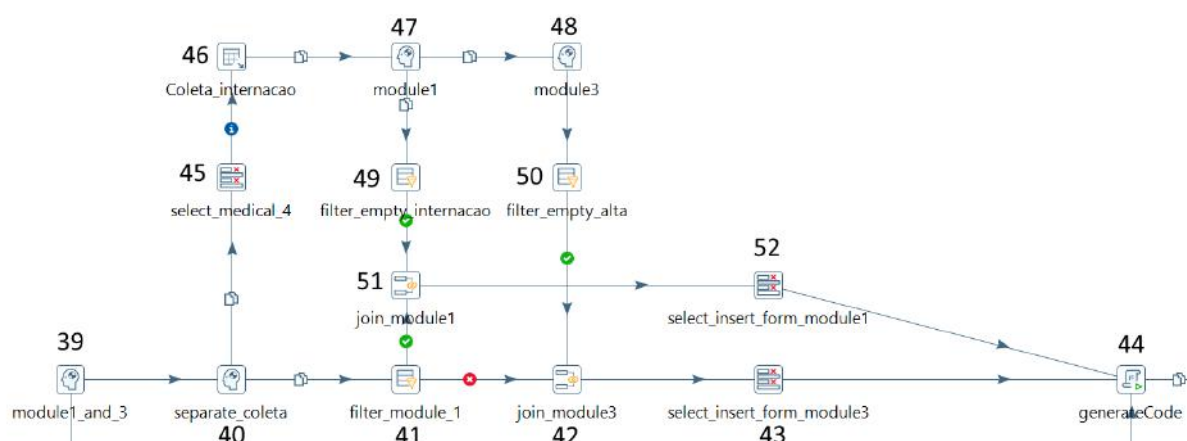
Apesar das operações e da lógica da *getProntuarios* serem simples, são muito utilizadas durante toda a solução. Por esse motivo, inserir os passos em uma sub-transformação facilita a replicação da funcionalidade de buscar prontuários para qualquer ponto da arquitetura que for necessário. A Figura 29 mostra as operações presentes nesta sub-transformação.

**Figura 29:** A sub-transformação *getProntuarios*.



A segunda parte do *createFormRecords* se divide em dois fluxos, conforme pode ser visto na Figura 30. O primeiro trata de gerar os registros correspondentes aos módulos 1 e 3, que possuem apenas uma entrada cada por paciente. Enquanto um segundo fluxo irá gerar os registros relativos ao módulo 2, onde haverá o cálculo da quantidade de dias de internação de cada paciente e a geração de um registro por dia internado.

**Figura 30:** Lógica de geração de registros dos módulos 1 e 3 para a tabela *tb\_FormRecord*.

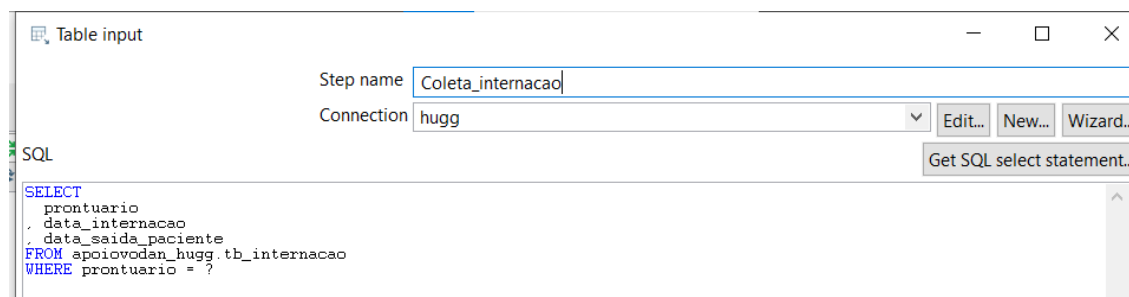


Para os módulos 1 e 3, após a coleta dos prontuários, precisamos das informações da tabela *tb\_internacao* (operação 46) presente no banco de dados do HUGG. Também vamos precisar dividir os registros em duas linhas de trabalho, uma para os registros de entrada (módulo 1) e outra para os registros de saída (módulo 3).

A operação *separate\_coleta* (operação 40) divide os fluxos em busca das informações do banco, através do *Coleta\_internacao* (operação 46), e em *filter\_module\_1* (operação 41) que separa os registros de cada módulo. Vale ressaltar que antes do *coleta\_internacao* há um filtro de campos, denominado *select\_medical\_4* (operação 45), isso ocorre, pois, operações de busca em BD que possuem uma cláusula de busca que use um valor do registro, como retornar as informações apenas referentes a um paciente em seu respectivo registro, só podem ser chamadas quando o fluxo possui apenas os campos que serão

utilizados, como visto na Figura 31.

**Figura 31:** Exemplo de busca em BD que utiliza uma informação dinâmica representada por '?' vinda dos registros presentes no fluxo.



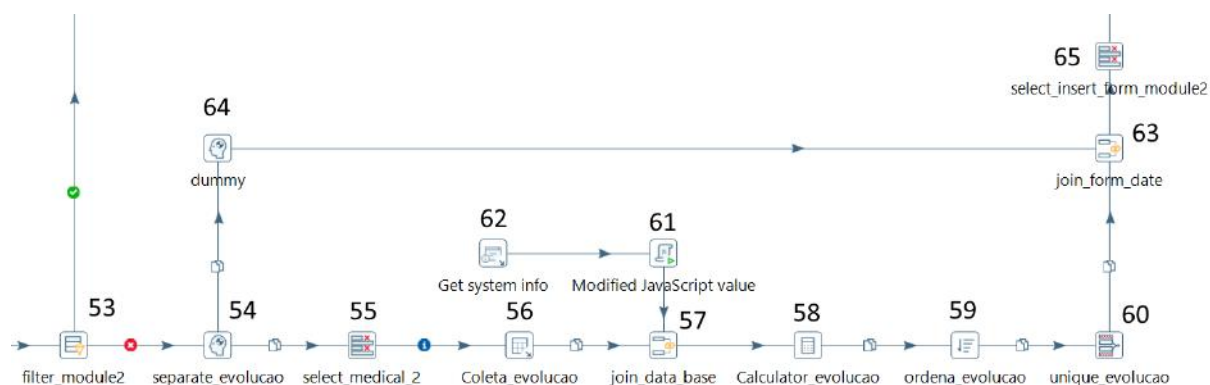
Os dados obtidos na consulta da tabela `tb_internacao` (operação 46) são enviados para as operações `module1` e `module3` (operações 47 e 48 respectivamente), que irão passar os dados para os fluxos dos módulos 1 e 3. Cada um deles irá então para as transformações `filter_empty_internacao` (operação 49) e `filter_empty_alta` (operação 50), para eliminar casos em que o paciente possa não ter dados na tabela ou ainda esteja internado. Então, os fluxos convergem para o `join_module1` (operação 51) e `join_module3` (operação 42), onde a informação de cada paciente é agregada ao seu respectivo registro, através da igualdade entre o prontuário do registro e o prontuário do dado obtido da consulta ao BD.

Com os registros corretamente unidos, terminamos esse setor da lógica com um filtro dos campos nas operações `select_insert_form_module1` (operação 52) para o módulo 1 e `select_insert_form_module3` (operação 43) para o módulo 3, retirando as informações que não serão pertinentes. Os fluxos seguem para a última parte da sub-transformação, que será explicada após debatermos a lógica de geração de registros do módulo 2. Desta forma, teremos uma visão completa do ciclo de vida da sub-transformação e poderemos compreender os dados que são enviados para a transformação central ao fim de sua execução.

A lógica para a geração dos registros do módulo 2, que pode ser analisada na Figura 32, se inicia com o `dummy_separate_evolucao` (operação 54), que divide o fluxo em 2. Isso ocorre pois como o módulo 2 terá um registro por dia de internação, precisamos multiplicar cada registro de paciente pelo número de dias de internação. Um Fluxo será enviado para o `join_form_date` (operação 63) diretamente e será mergeado (operação 63) com  $n$  registros

encontrados na lógica do segundo fluxo, que calcula os n dias de internação e gera um registro para cada um.

**Figura 32:** Lógica de geração do registro dos módulos 2 para a tabela `tb_FormRecord`.



O segundo fluxo, com o cálculo de dias, irá primeiramente para a operação `select_medical_2` (operação 55) para deixar apenas o campo `medicalRecord` que é utilizado na busca de informações na tabela de evolução, na operação seguinte denominada `Coleta_evolucao` (operação 56). Como há diversos registros de evolução de cada paciente, precisamos definir para qual dia cada registro representa.

O cálculo para obter o dia referente ao registro de internação se depara com algumas dificuldades, como o fato de alguns pacientes passarem mais de um mês ou até um ano internados. É preciso calcular a data a partir de uma data base. Nesse caso, as transformações `Get system info` e `Modified JavaScript value` (operações 62 e 61 respectivamente) geram uma data válida para cálculo de 01 de janeiro de 1980. Com essa data podemos calcular a quantidade de dias que se passou e achar um identificador único para cada dia internado.

Cada registro do fluxo recebe a data gerada, através da operação `join_data_base` (operação 57). As informações são enviadas para `Calculator_evolucao` (operação 58) que subtrai a data do registro da data base obtida de 1980 e calcula a quantidade de dias que se passaram. As operações `ordena_evolucao` e `unique_evolucao` (operações 59 e 60 respectivamente) filtram os registros deixando apenas um registro por dia de internação, através dos dias de diferença calculados anteriormente.

Por fim, os dados já filtrados são combinados aos dados crus (operação 63), de forma

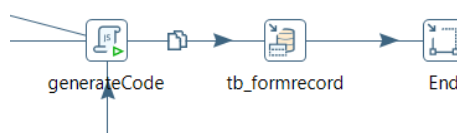
que teremos um registro por dia de internação. Com o término da lógica da geração de registros do módulo 2, o fluxo segue para o *select\_insert\_form\_module2* (operação 65) trocando o campo *data\_evolucao*, advinda do banco de dados, para *dtRegistroForm*, garantindo que a data de obtenção da evolução é a mesma do registro do módulo no banco.

Ambas as transformações, dos módulos 1,2 e do módulo 3 seguem então para as operações finais, como pode ser observado na Figura 33. Os fluxos convergem na operação *generateCode* que gera um campo em *javascript* concatenado do id do hospital, questionário, participante, *form* e data do registro.

O código se faz necessário por uma limitação da ferramenta Pentaho, que não consegue identificar registros duplicados quando a inserção é diferenciada por todas as colunas da tabela. Com um campo contendo a concatenação dessas informações, podemos garantir que apenas teremos registros únicos na tabela verificando apenas uma das colunas disponíveis.

Com o código gerado, o fluxo finalmente chega em *tb\_formrecord* armazenando as informações na tabela *tb\_formrecord*. Os dados armazenados são enviados de volta para a transformação pai, a partir de uma transformação de END.

**Figura 33:** Operação finais da sub-transformação *createFormRecords*



Com o fim da sub-transformação *createFormRecords*, podemos retornar a transformação central. Seguindo a arquitetura da Figura 27, a próxima sub-transformação é a *getProntuarios* (operação 21), que já foi descrita anteriormente e que coleta a relação id do paciente com seu prontuário. Os dados colhidos desta sub-transformação são enviados para a *getAssesstmentQuest* (operação 22).

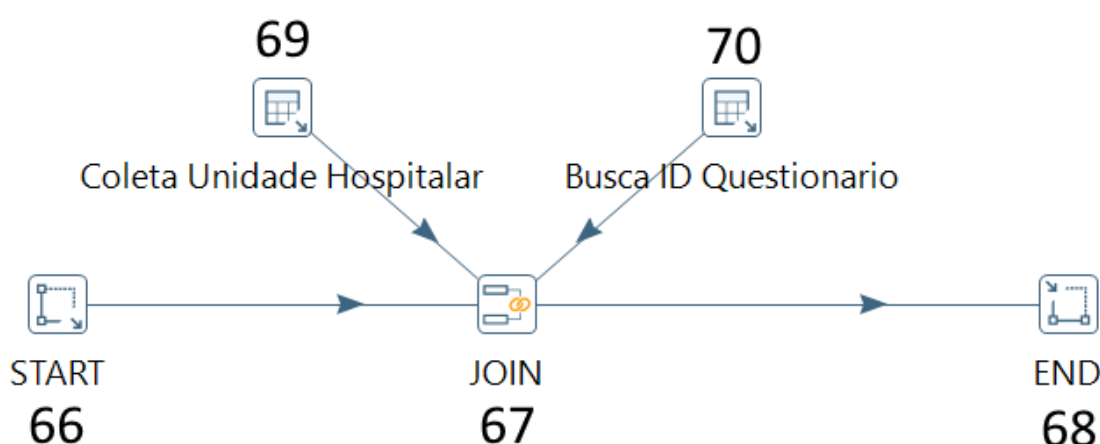
Na Figura 34, é possível ver a arquitetura da sub-transformação *getAssesstmentQuest*, que já foi mostrada na figura 27 como operação 22. O objetivo desta transformação é acrescentar a cada registro paciente/prontuário, recebido da *getProntuarios* (operação 21 da Figura 27), os dados de *hospitalUnitID* e *questionnaireID*. Esses dados serão

importantes para posteriormente coletar as perguntas de cada módulo na tabela `tb_question` como pode ser observado na Figura 23.

A sub-transformação começa com a operação START (operação 66), que nesse caso recebe registros com as informações de id do paciente e prontuário. Seguindo para um JOIN (operação 67), que une o id do hospital e do questionário, proveniente cada um de suas respectivas operações de busca em tabela (operações 69 e 70 respectivamente), que são únicos para todos os pacientes. Os registros, que agora contam com as 4 informações, seguem para a operação de saída (operação 68), retornando ao fluxo da transformação central

Retornando à transformação central, temos agora um conjunto de operações da Figura 35 que tem por objetivo transformar as informações provenientes da `getAssesstmentQuest` (operação 22 da Figura 27) em um conjunto de dados no qual cada registro simboliza um paciente, para uma questão e um módulo. Desta forma, para as perguntas do módulo 1, haverá um registro desta pergunta para cada paciente, assim como as questões referentes ao módulo 3. As perguntas do módulo 2 terão um registro por dia de internação do paciente.

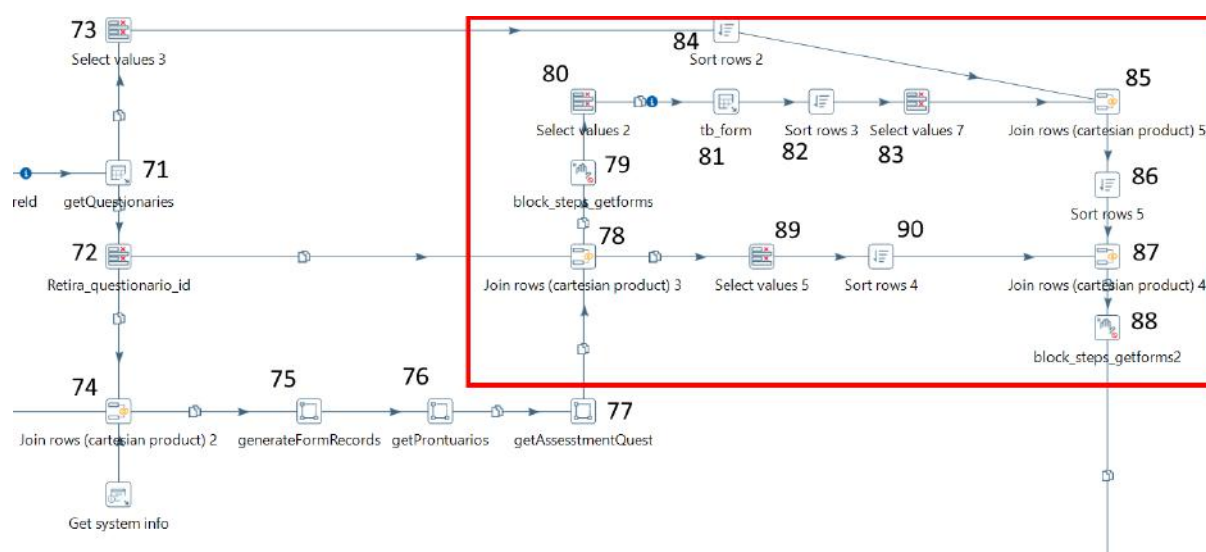
**Figura 34:** Arquitetura da sub-transformação `getAssesstmentQuest`



Como já gravamos os módulos do FRC-OMS efetuando os cálculos de dias de internação, entrada na unidade de saúde e alta/óbito, basta ligarmos os registros que temos no fluxo após o `getAssesstmentQuest` (operação 77 da Figura 35) , que representa um registro por paciente, aos módulos de cada paciente. Obtendo assim um registro para cada pergunta que deve ser respondida para cada paciente.

Para elucidar o objetivo da próxima parte da transformação central, vamos pensar em um exemplo prático. Sejam três questões: A,B e C. Relativas aos módulos 1,2 e 3 respectivamente. Digamos que nossa base de dados possui informações de apenas 1 paciente, internado por 5 dias antes de sua alta. Ao sairmos da sub-transformação *getAssesmentQuest* teremos 5 registros no fluxo, cada um com 4 informações. O id do paciente, seu prontuário e os dados de hospital e questionário.

**Figura 35:** Terceira parte da transformação central



O conjunto de operações da Figura 35 irá transformar esses registros de modo que teremos 1 registro para a questão A, de módulo 1, referente a entrada do paciente na unidade de saúde, 5 registros da questão B, referentes a cada dia de internação e 1 registro da questão B referente ao módulo 3, de alta. Totalizando 7 registros, um para cada vez que as perguntas devem ser respondidas.

Conforme aumentamos a quantidade de pacientes na base de dados, multiplicamos a quantidade de respostas que devem ser geradas. Para efeito de comparação, em um banco com apenas 20 pacientes, temos mais de 10.000 respostas geradas, que envolvem extração, análise de texto, dentre outros tipos de transformação.

Seguindo a arquitetura da Figura 35, os registros seguem para *Join rows (cartesian product) 3* (operação 78), onde recebem o *crfFormsID*, único para todas as perguntas. Deste ponto em diante o fluxo é dividido em dois. O primeiro é levado até a operação *block\_steps\_getforms* (operação 79).

Esse tipo de operação tem por objetivo garantir que uma execução só será feita após outra ser finalizada. Como temos diversas divisões e uniões, esse bloqueio impede que o fluxo seja executado até que todos os registros tenham passado pelo *generateFormRecords* (operação 75) .

Seguindo o fluxo, os registros passam para *Select values 2* (operação 80), onde o *crfFormId* é retirado para permitir utilizarmos os dados como parâmetros na operação *tb\_form* (operação 81), que busca na tabela *tb\_formrecord* os FRC-OMS dos

**Figura 36:** Query da operação *tb\_form*

pacientes. Neste momento estamos coletando as informações gravadas na transformação *createFormRecords*. A *query* utilizada é representada na Figura 36 abaixo:

```
SELECT
  formRecordID
  , participantID
  , hospitalUnitID
  , questionnaireID
  , crfFormsID
FROM vodan_br_bd.tb_formrecord
WHERE hospitalUnitID = ? AND participantID = ? AND questionnaireID = ? AND crfFormsID = ?
```

Ainda na Figura 35, as duas operações seguintes, *Sort rows 3* (operação 82) e *Select values 7* (operação 83), tratam os dados e trocam o nome da coluna *participantID* para *participantID\_2*. O objetivo desta troca é permitir que possamos usar o campo *participantID\_2* como critério de união na operação *Join rows (cartesian product) 4* (operação 87), unindo registros que tenham o mesmo id de paciente, representado no campo *participantID* do segundo fluxo advindo de *Join rows (cartesian product) 3* (operação 78).

A operação *Join rows (cartesian product) 5* (operação 85) une cada registro com a descrição do módulo, sendo eles *Admission form*, *Follow-up* e *Discharge/death form*, relativos aos módulos 1,2 e 3 respectivamente. Essas descrições serão usadas para diferenciarmos de qual módulo cada questão faz parte. É importante ressaltar que o id dessas descrições pode mudar dependendo do banco de dados onde a transformação rode. Por conta disso, usaremos a descrição para saber qual o módulo está sendo tratado.

Por fim, terminamos esta parte da transformação central com a operação *Join rows (cartesian product) 4* (operação 87), que une o segundo fluxo advindo de *Join rows (cartesian product) 3* (operação 78) aos registros que geramos do primeiro fluxo, do qual



acompanhamos. Os dados deste fluxo secundário possuem o prontuário de cada paciente. Informação que foi perdida após a consulta feita na operação *tb\_form*, que retém apenas os dados retornados na *query*. Desta forma, unimos novamente os prontuários aos seus respectivos pacientes.

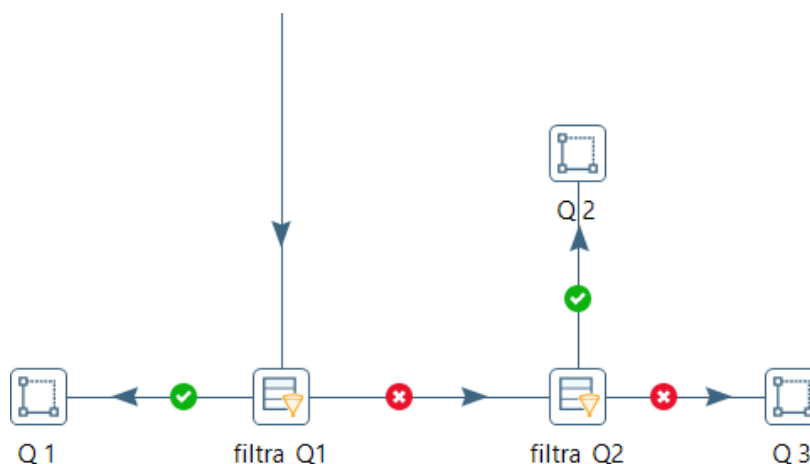
Novamente a operação *block\_steps\_getforms2* (operação 88) é um bloqueio inserido apenas para garantir que as operações de geração de respostas, localizadas na próxima parte da transformação central, só sejam executadas quando os fluxos anteriores tiverem sido encerrados.

A última parte da transformação central é composta por três sub-transformações principais, sendo elas Q1, Q2 e Q3, cada uma gerando respostas para questões dos módulos 1, 2 e 3 respectivamente, representadas na figura 37. Desta forma podemos paralelizar a execução desta parte da solução, que por possuir diversos tipos de análise diferentes acaba sendo a mais custosa em termos de processamento.

Seguindo a arquitetura presente na Figura 37, os registros chegam à operação *filtra\_Q1*. É importante frisar que temos um registro para cada pergunta a ser respondida por paciente e módulo. Assim, todo o registro que chega às sub-transformações seguintes, deve ser respondido segundo a origem do seu dado, seja anamnese, medicamentos, controle ou qualquer outra tabela do banco de dados do HUGG que devemos obter a informação pertinente à questão.

As operações *filtra\_Q1* e *filtra\_Q2* funcionam para separar os registros referentes ao módulo 1, 2 ou 3, enviando para as sub-transformações Q1, Q2 e Q3 respectivamente. Na próxima seção iremos analisar o funcionamento da transformação Q1.

**Figura 37:** Arquitetura da chamada para as sub-transformações de geração de respostas.



### 3.4.2 Transformação Q1

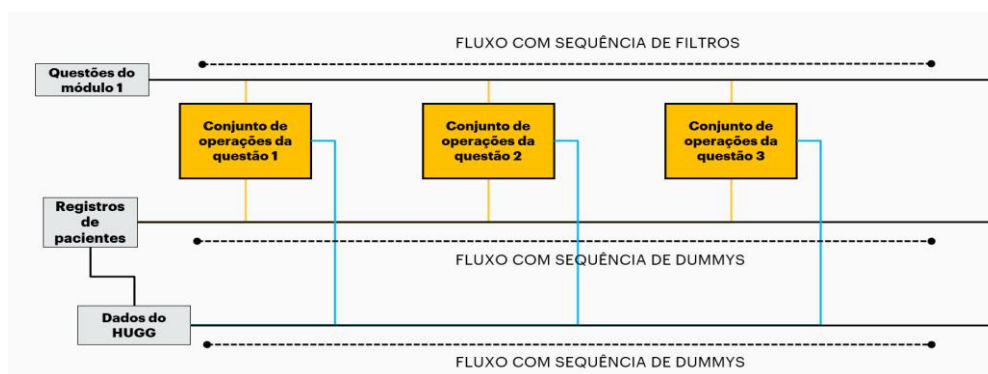
A sub-transformação Q1 tem por objetivo gerar as respostas das questões pertinentes ao módulo 1 do questionário de COVID da OMS. Ela é alimentada pelos registros advindos do fluxo da transformação central, que contém um registro por paciente. Desta forma, a arquitetura da transformação foi pensada para utilizar esse fluxo como base, retirando os registros de cada questão e os enviando para um conjunto de operações que irá gerar a resposta da questão em particular.

Podemos descrever esta arquitetura como um fluxo principal, que contém uma série de *dummys* que duplicam os registros de pacientes e os enviam para um conjunto de operações. Cada conjunto é responsável por gerar a resposta de uma pergunta. Um fluxo superior detém todas as perguntas daquele módulo e segue por um conjunto de filtros, que direcionam as informações da questão para o conjunto de operações referentes a ela.

Como as questões necessitam de dados para serem respondidas, que advém das tabelas do banco do HUGG, como a evolução, controle, medicamentos, entre outras, temos um conjunto de fluxos na parte inferior, cada um relativo a uma origem de dados que seguem por um conjunto de *dummys*. Como os *dummies* não fazem nada, cada dado que é recebido por um *dummy*, passa por ele sem nenhum tratamento e pode ser enviado para um ou mais fluxos, logo o *dummy* neste caso funciona como um cruzamento de fluxos, ele recebe as informações do fluxo e envia as mesmas informações para os fluxos de saída, garantindo

que ambos possuem as mesmas informações do fluxo original. Desta forma cada *dummy* duplica os registros daquela base para o conjunto de operações que necessitam desses dados para gerar a resposta da questão. A Figura 38 demonstra a arquitetura pensada para solucionar o problema, com divisão do fluxo utilizando dummies.

**Figura 38:** Desenho da arquitetura da transformação Q1



Conforme podemos examinar na Figura 39, a parte superior da transformação, cercada pelo tracejado roxo, representa o fluxo com as questões do módulo 1. Um conjunto de filtros definidos como Filtro\_1, Filtro\_2 e etc... dividem o fluxo, coletando os registros referentes a cada questão para enviá-los aos blocos abaixo, envoltos pelo tracejado amarelo. Os registros que não fazem parte da questão do filtro, são enviados ao filtro seguinte.

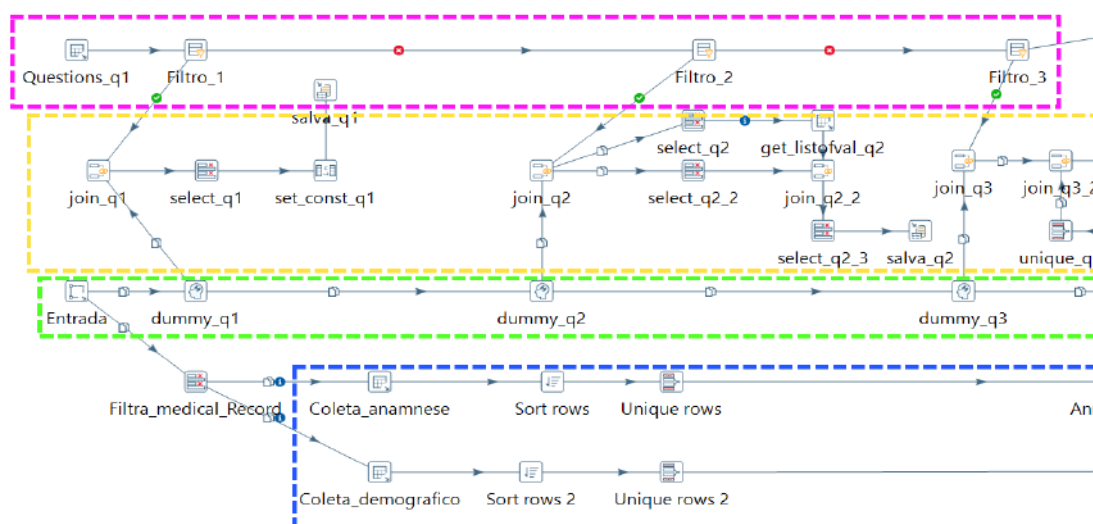
Cada conjunto de operações dentro do tracejado dourado representa a resolução de uma questão. A primeira questão, cuja transformações apresentam `_q1` no nome, são referentes a responder qual o hospital de origem do FRC-OMS, no caso sempre será o HUGG. As questões seguintes possuem uma lógica própria para cada tipo de resolução.

O bloco tracejado em verde representa os registros dos pacientes, que são recebidos da transformação central. Cada paciente é representado por um registro neste fluxo, e para cada questão é enviado a quantidade de pacientes. Desta forma, sempre haverá uma resposta para cada paciente da transformação.

Por último, na parte inferior da figura, em tracejado azul, podemos ver os fluxos de coleta de dados. O fluxo superior, com início na operação `Coleta_anamnese` traz as informações da tabela de anamnese do banco de dados do HUGG, enquanto o fluxo inferior

colhe os dados da tabela de dados demográficos. Sempre que uma pergunta necessita desses dados, um *dummy* é inserido e os dados são replicados para o conjunto de operações desta questão.

**Figura 39:** Arquitetura da transformação Q1 no Pentaho



As questões do módulo 1 foram desenvolvidas utilizando a arquitetura da transformação Q1 até a questão 48, quando a resolução de problemas que serão descritos na próxima seção resultou em uma segunda versão desta solução. É importante ressaltar que este trabalho tem por objetivo a resolução de 313 questões, contabilizando os três módulos do questionário.

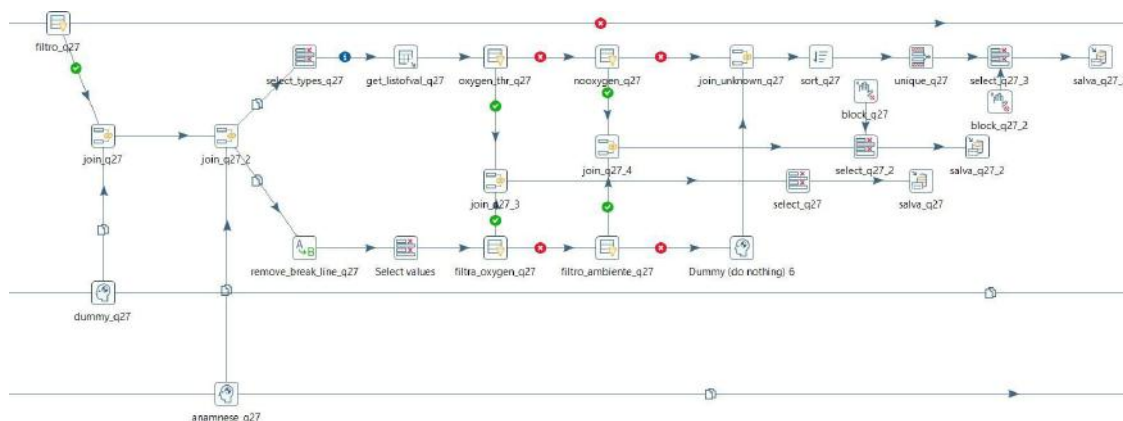
### 3.4.3 Problemas encontrados

A arquitetura da transformação Q1 se mostrou funcional, porém não prática. Conforme o nível de detalhes das questões aumenta, a complexidade das soluções aumenta de forma exponencial. Como não há uma estrutura de reutilização da lógica construída na questão anterior para a próxima, a cada nova questão é necessário repensar a lógica do zero, aumentando de forma considerável o trabalho envolvido na resolução de uma única pergunta, como exemplificado na Figura 40.

Esse problema ocorre por conta da independência dos conjuntos de operações de cada questão. Conforme visto na seção anterior, cada conjunto recebe um bloco de registros

independentes de pacientes, questões e dados advindos do banco do HUGG. Desta forma não há compartilhamento de informações entre os blocos ou da lógica da solução.

**Figura 40:** Exemplo de transformação com alta complexidade da primeira versão da solução.



Com um tempo elevado de construção de cada bloco de operações por questões, o modelo Q1 se tornou insustentável, e por consequência os modelos Q2 e Q3 também, por seguirem a mesma lógica de arquitetura.

Outro ponto que contribuiu para uma nova versão da solução, foi a manutenibilidade das operações. Com a independência dos blocos, um ajuste na lógica da geração de resposta de uma questão, além de custoso, pode necessitar de ajuste em outros blocos, com funções semelhantes. Aumentando também a complexidade dos documentos e guias da solução, que necessitam explicar a lógica de cada bloco individualmente.

### 3.5 VERSÃO 2.0

Muito usada no desenvolvimento de softwares, sistemas complexos e conceituada na Engenharia de Software, a modularidade é conhecida pela filosofia popular “dividir para conquistar” (OLIVEIRA, 2017). Seguindo esta vertente, a segunda versão desta solução busca utilizar as vantagens provenientes da modularidade para agilizar, simplificar e otimizar os processos de desenvolvimento e o resultado final da extração, que é o produto final deste trabalho.

Esta seção se inicia com a descrição da arquitetura da solução, seguindo através de três subseções seguintes explicando como essas transformações conseguem responder

múltiplas questões utilizando as mesmas transformações, com parametrizações pelo banco de dados. Por fim, iremos cobrir o conjunto de transformações auxiliares que são responsáveis pela resolução das questões subsequentes que não puderam ser simplificadas em transformações genéricas, como as questões relacionadas a alta/óbito.

### 3.5.1 Arquitetura da solução

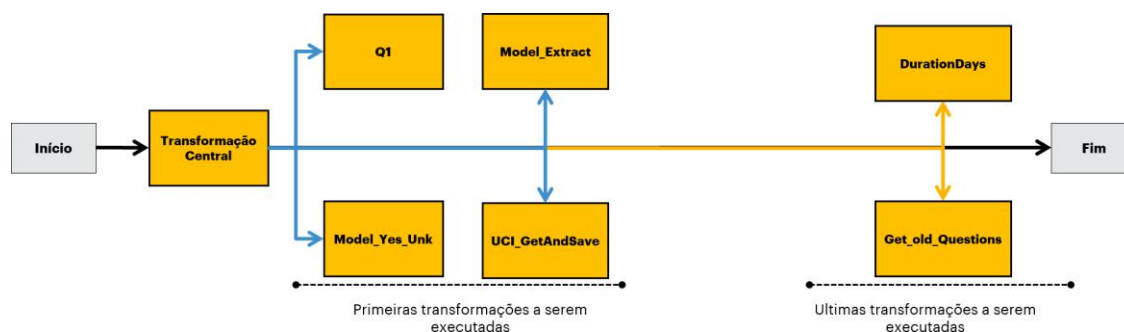
Através das transformações Q1, Q2 e Q3, foi possível obter alguns padrões de resposta, como as múltiplas escolhas, as de extração, que necessitam extrair um dado de um texto e aquelas que dependem de outras questões, como algumas perguntas do módulo 3.

Desta forma, foram elaboradas sete transformações principais, conforme mostra a Figura 41 sendo elas *:model\_yes\_unk*, *model\_extract*, *Q1*, *UCI\_getAndSave*, *Outcome*, *DurationDays* e *get\_old\_questions*. A primeira, *model\_yes\_unk*, é responsável pelas questões múltiplas escolhas de sim, não ou desconhecido. Seu funcionamento será explicitado na próxima subseção. A *model\_extract* é responsável pela resolução de questões que necessitem de extração da resposta, como a temperatura do paciente por exemplo. A transformação Q1 foi explicada nas subseções anteriores. *UCI\_getAndSave* é responsável por descobrir se o paciente esteve na UTI durante sua internação e se sim, quantos dias ele ficou e sua data de saída, se houver. *Outcome* trata algumas questões específicas de alta/óbito. *DurationDays* é responsável pela resolução de questões que envolvam duração de dias durante a evolução do paciente. Por último, *get\_old\_questions* é uma sub-transformação que é responsável por responder questões do módulo 3 que dependam de questão de outros módulos, como se o paciente utilizou determinada medicação durante o tratamento.

As transformações funcionam a partir da leitura de dados advindos do banco de dados do MySQL, da Fiocruz. Essas informações presentes no banco detalham quais são as questões respondidas por cada transformação e quais os parâmetros serão utilizados. Grande parte das transformações funcionam a partir de regex como será detalhado nas próximas subseções. Logo é possível, alterando o regex e alguns parâmetros, utilizar o mesmo conjunto de operações, para responder a questões diferentes.

A lógica explicada anteriormente para gerar os dados necessários para iniciar a resolução das questões foi mantida, com alterações na arquitetura apenas nas partes referentes às operações que geram as respostas às questões do FRC-OMS. Na Figura 41 é possível ver uma visão geral da arquitetura da nova solução.

**Figura 41:** Nova arquitetura.



É importante ressaltar que as questões presentes na sub-transformação Q1 são as mesmas elaboradas durante a primeira versão deste trabalho. A maioria destas possui a possibilidade de serem transferidas para as transformações genéricas. Esse esforço, porém, não faz parte do escopo deste trabalho, que visa garantir construir uma ferramenta modular e de fácil manutenção, além de disponibilizar os resultados deste ETL. Por este motivo, não haverá um novo capítulo re-apresentando esta sub-transformação.

### 3.5.2 Resolução de questão de múltipla escolha (*model\_yes\_unk*)

As questões de múltipla escolha podem ser definidas como aquelas em que, através de um regex, se verifica se a informação existe para o paciente em questão, no dia ou em toda a sua estadia. Ela é respondida simplesmente por sim, no caso de a informação estar presente nos registros procurados, ou desconhecido, caso a informação não esteja disponível para aquele paciente.

Um exemplo disso é o uso de respirador. Tanto o módulo 2 quanto o módulo 3 apresentam questões acerca desse assunto. Caso o ETL identifique que houve uso de ventilação artificial, através do match do texto com um regex cadastrado previamente no banco de dados destino, da Fiocruz, é marcada a opção "Sim" para a questão. Caso não seja encontrada informações sobre o uso dessa ventilação, o ETL marca a opção

desconhecida.

Nos casos em que a informação é negada, essa métrica é contabilizada de forma simples na versão atual deste trabalho, não havendo um robusto mecanismo de verificação de negação. Desta forma, "Houve uso de [ ] e ventilação" e "Não houve uso de ventilação" podem não apresentar distinção de significado para o atual conjunto de operações se a informação não for cadastrada nas configurações de forma manual. Logo, em ambos os casos o ETL poderia entender que houve uso de ventilação mecânica. A implementação desta funcionalidade de forma robusta e inteligente se encontra nas melhorias sugeridas de trabalhos futuros e pode fazer uso de IA e técnicas de aprendizado para aumentar a assertividade das informações analisadas.

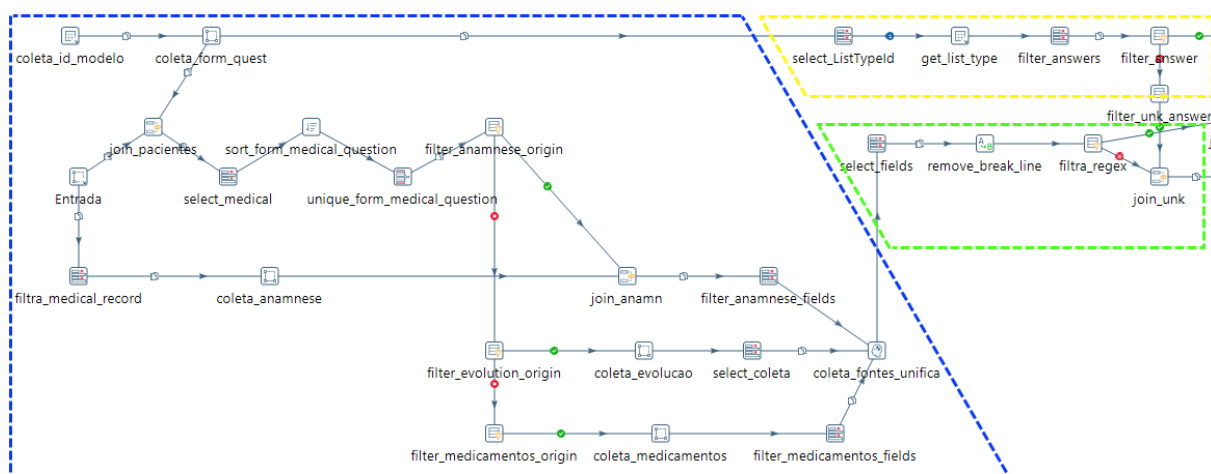
Esta sub-transformação pode ser dividida em 3 partes principais. a obtenção e filtragem das informações dos pacientes. a coleta das respostas que para aquela questão em específico simbolizam "Sim" ou "Desconhecido", visto que as questões podem ter respostas diferentes, como "ok", "não ok", "presente" e etc., que apresentam por sua vez, id's diferentes no banco, logo precisamos recolher as respostas corretas para cada questão. a última parte é responsável pela escolha da resposta através do regex cadastrado no banco. As informações do paciente que foram coletadas na parte 1 da transformação são filtradas através do regex do banco e caso seja encontrado correspondência, a questão é respondida com um "Sim", caso contrário, caso nenhuma das informações daquela questão possua o regex, a questão é respondida como "Desconhecido".

Os dados iniciais recebidos pela sub-transformação de entrada são os mesmos recebidos pela sub-transformação Q1 explicada anteriormente, ou seja, informações do id do paciente, hospital, questionário e FRC-OMS. Essas informações vão ter o objetivo principal de passarem por 3 subfluxos, cada um coletando de uma fonte de informação diferente, sendo elas anamnese, evolução e medicamentos.

A Figura 42 ilustra a divisão da primeira parte desta transformação. Em azul, parte da transformação responsável pela coleta das informações. Em verde e amarelo, início das partes responsáveis pela decisão da resposta e coleta da opção de "Sim" ou "desconhecido" de cada questão, respectivamente.

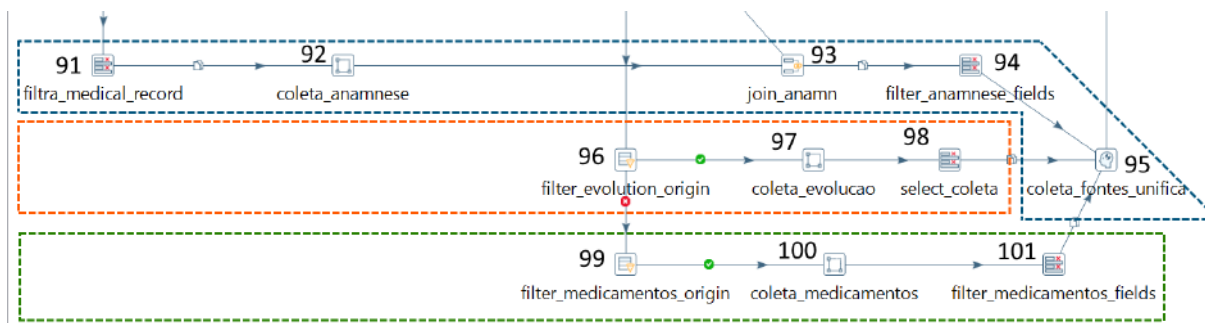


**Figura 42:** Estrutura da transformação de respostas para questões do formato sim, não e desconhecido.



Na Figura 43, temos um zoom na parte inferior da Figura 42, onde podemos ver que em azul há o fluxo de coleta de dados de anamnese, e em laranja e verde, os fluxos de coleta de dados de evolução e medicamentos, respectivamente.

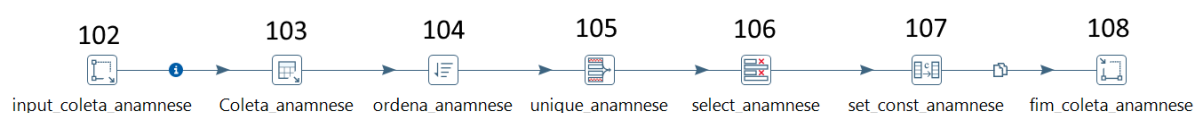
**Figura 43:** Fluxos de coleta de dados para a resolução de questões do formato sim, não e desconhecido.



No fluxo de anamnese, visto na Figura 43, primeiro filtramos os dados (operação 91) para manter apenas as informações do registro médico do paciente. Isso ocorre pois como cada paciente possui apenas um conjunto único de anamnese por especialidade médica, não precisamos de outras informações para coletar dados dessa fonte. Esses registros médicos são passados para a sub-transformação de anamnese (operação 92) que faz a coleta no banco de dados. Em seguida os dados são combinados com as questões que serão respondidas com esta fonte de dados (operação 93) e são filtradas na operação *Coleta\_Fontes\_Unifica* (operação 95), para excluirmos dados que não serão usados posteriormente.

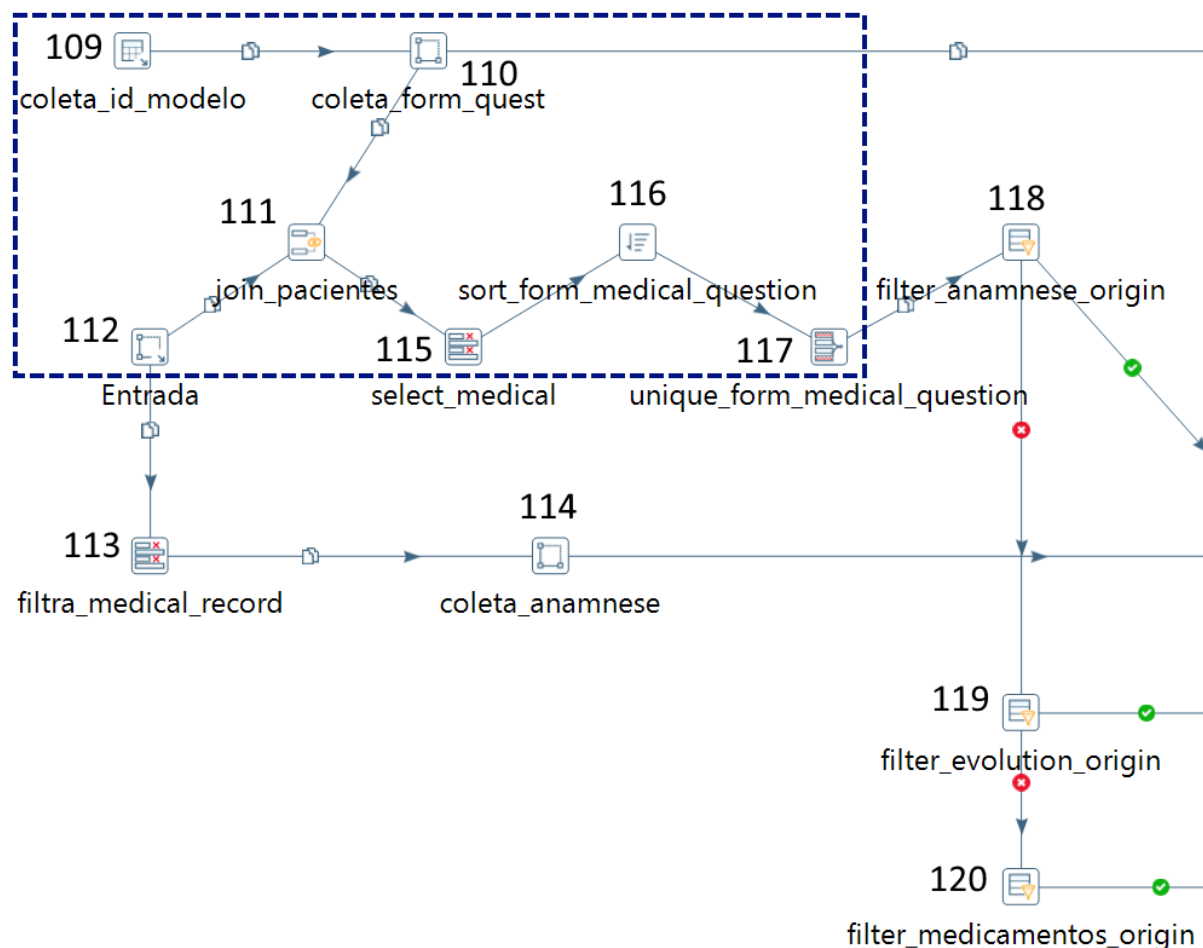
O fluxo de coleta de anamnese é simples, como pode ser visto na Figura 44. Começando com a operação de input (operação 102), com os dados do id do paciente, é feita uma consulta SQL através da operação coleta\_anamnese (operação 103) para procurar registros de anamnese para este Id. Em seguida, eles são ordenados e unificados (operações 104 e 105 respectivamente), para excluir qualquer redundância que possa ter vindo do banco de dados. Os dados são então tratados, para apresentar a origem com o valor 1 (operações 106 e 107), lógica que será explicada posteriormente durante o setor da transformação que identifica se a resposta deve ser "Sim" ou "Desconhecido". Os dados são então devolvidos para o fluxo padrão (operação 108).

**Figura 44:** Fluxo de coleta de dados da tabela de anamnese



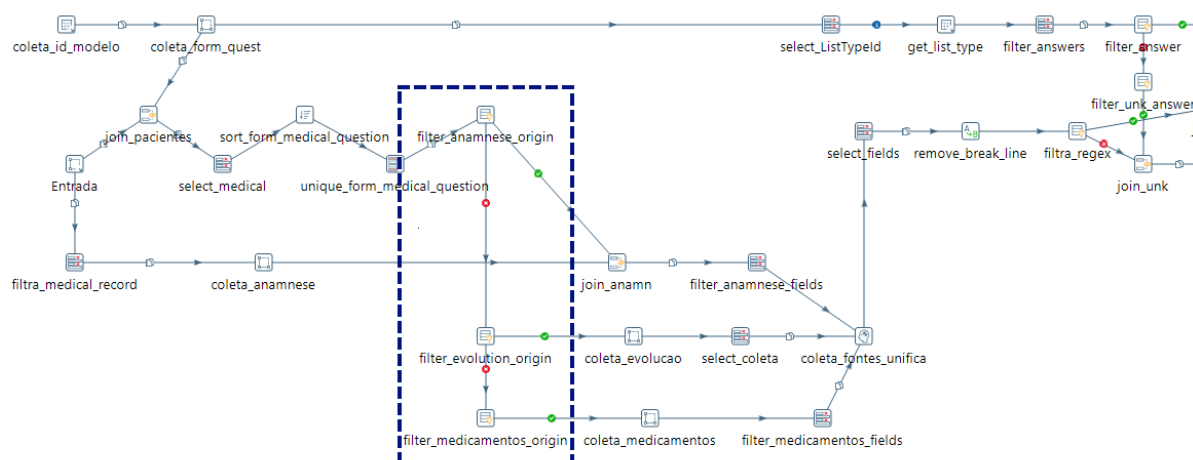
Os fluxos de coleta de evolução e medicamentos não recebem apenas o id do paciente, mas sim um conjunto de informações que são advindas de um conjunto de operações detalhadas na figura abaixo. Essas operações têm por objetivo unir as questões de origem de cada fonte, as informações de cada paciente. Desta forma, teremos para cada registro no fluxo o dado da questão que está sendo respondida e o paciente que irá ser analisado para responder aquela questão. Teremos por tanto  $n$  (sendo  $n$  o número de questão múltipla escolha) multiplicado por  $m$  (sendo  $m$  o número de pacientes) registros no fluxo.

**Figura 45:** Fluxo de coleta das questões que serão respondidas na transformação de perguntas múltiplas escolhas e sua fonte de dados.



Na Figura 45 podemos analisar que a operação de *Coleta\_Id\_Modelo* (operação 109) é responsável por pegar as questões que serão respondidas na transformação de múltipla escolha. Após isso, é recolhida a fonte de dados dessas questões (operação 110). Os registros encontrados são combinados com os registros de pacientes que serão tratados nesta execução do ETL (operação 111). É feita uma operação de *select\_medical* (operação 115) para filtrar os dados que não serão utilizados e uma ordenação e unificação (operações 116 e 117 respectivamente) para evitar que redundâncias do banco porventura possam estar sendo analisadas. O conjunto de filtros, representado na Figura 46, que vemos na sequência, são os responsáveis por direcionar cada questão a seu fluxo de origem (operações 118, 119 e 120 da Figura 45).

**Figura 46:** Conjunto de filtros que direciona os registros para sua fonte de informação correta.



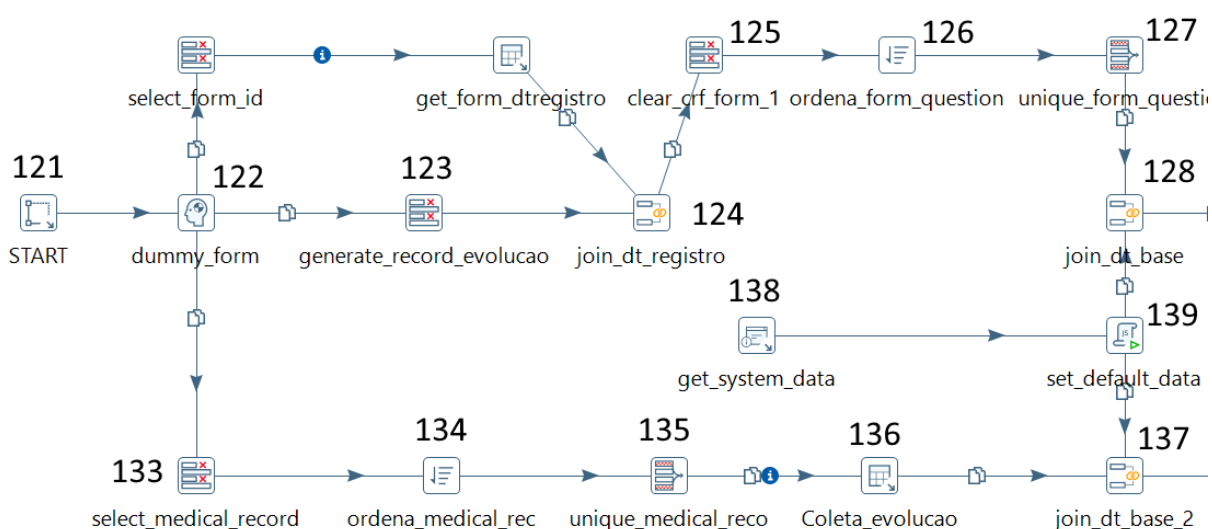
As sub-transformações de coleta de evolução e medicamentos seguem premissas semelhantes. Ambas coletam dados de diversos dias e condensam as informações para cada paciente. A coleta evolução possui uma estrutura para coletar um dado por dia. A parte superior da sub-transformação faz a coleta dos módulos do paciente. Como cada paciente pode possuir vários registros de módulos 2 e apenas 1 módulo 1 e 3, nos casos de questões envolvendo o módulo 2, é preciso que as questões e os dados coletados estejam de acordo com o registro do módulo da data em questão. Para questões do módulo 3, é levada em consideração todos os dias internação do paciente.

A parte inferior da transformação de coleta de evolução, como pode ser visto nas Figura 47.A e Figura 47.B, é responsável por buscar, para cada id de paciente, suas informações de evolução. Duas transformações de busca de data do sistema são utilizadas para pegar o dia atual, que será utilizado para calcular os dias que serão filtrados por paciente (operações 138 e 139). Neste ponto, indicado na Figura 47 pelas operações de união *join\_dt\_base* e *Join rows* (cartesian product) (operações 128 e 137), tanto os módulos quanto os dados apresentam uma data de base. Em seguida nas operações de *Calculator\_evolucao\_coleta* e *Calculator\_evolucao\_coleta\_2* (operações 140 e 129 respectivamente) é calculado o dia de cada informação. No caso dos módulos, é calculado e mantido um registro por dia, nos dados de evolução a lógica é a mesma.

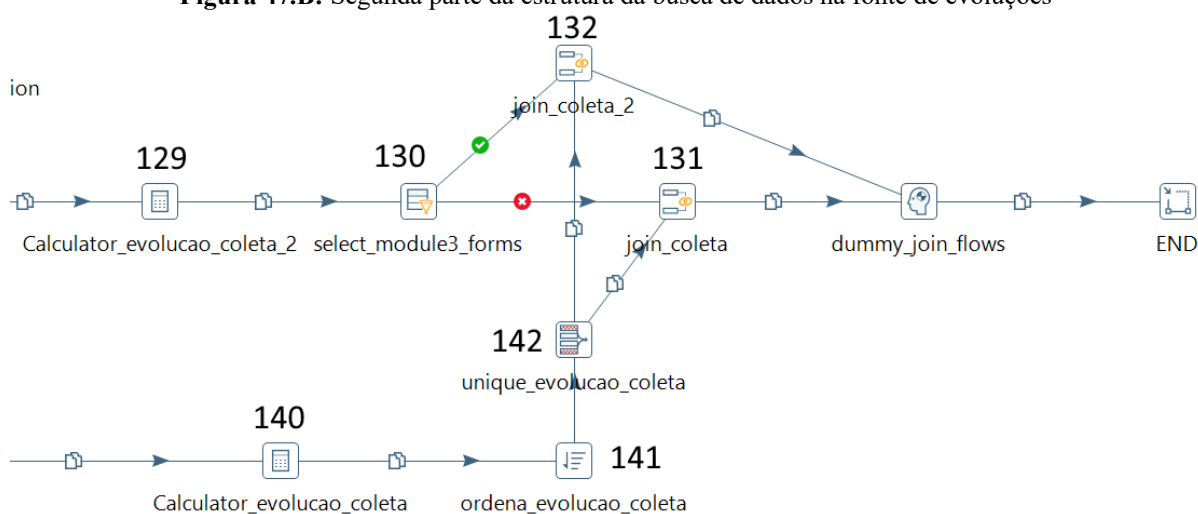
Por fim, dados e módulos são combinados nas duas transformações de merge finais (operações 131 e 132). É importante ressaltar que as questões de módulo 3 são unidas em outra operação (operação 132) para que seja garantido que durante a análise dos dados,

todos os dias sejam contados, ou seja. Para módulos 2 criamos um registro com o módulo do dia e uma evolução deste dia. Para registros do módulo 3, criamos uma redundância com vários registros para esse mesmo módulo, mas cada um com os dados de evolução de um dia. Isso garante que as questões do terceiro módulo serão respondidas com “Sim” se ao menos uma evolução tiver a característica procurada.

**Figura 47.A:** Primeira parte da estrutura da busca de dados na fonte de evoluções.



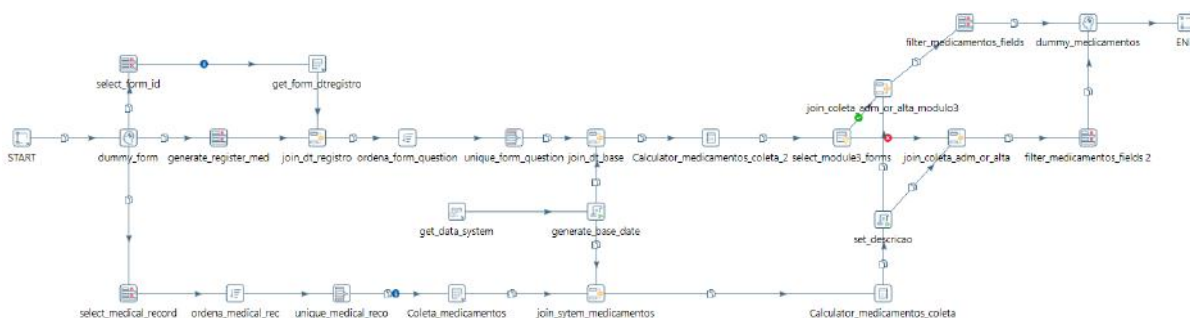
**Figura 47.B:** Segunda parte da estrutura da busca de dados na fonte de evoluções



A coleta de medicamentos segue exatamente a mesma lógica da coleta de evolução, como mostrado na Figura 48, a única mudança significativa, além dos nomes das operações, é que a fonte dos dados passa de ser a tabela de evolução para ser a tabela de medicamentos.

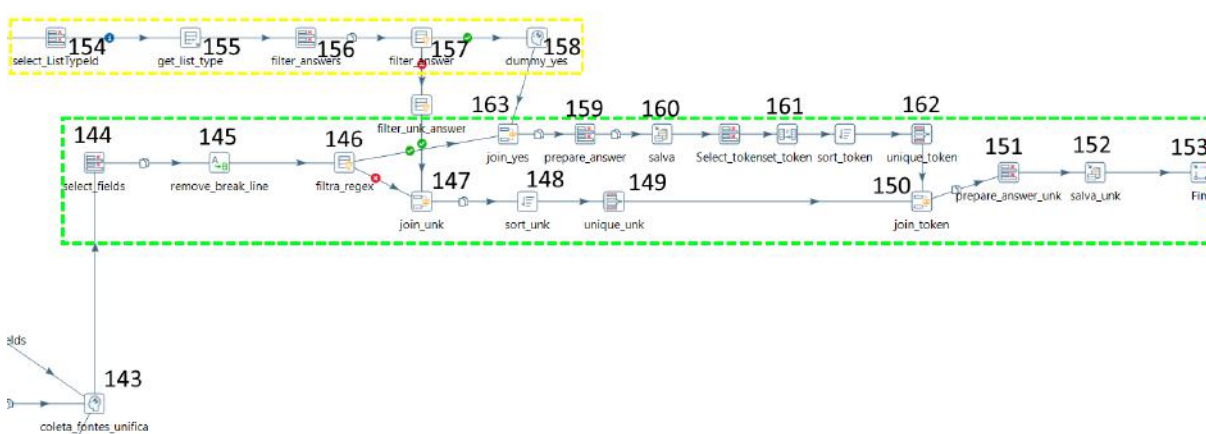
Apenas um dado será analisado por registro do módulo 2, enquanto todos os dados, sendo um por dia, dos pacientes internados serão analisados para responder questões do módulo 3.

**Figura 48:** Estrutura da busca de dados na fonte de medicamentos.



Em seguida, o fluxo de dados oriundo das três fontes de informação, anamnese, evolução e medicamentos, já unida a informação do paciente e da questão a ser respondida, é enviado para a série de operações em verde na Figura 49, responsável pela lógica de decidir se a resposta deve ser "Sim" ou "Desconhecido" para cada questão de cada paciente.

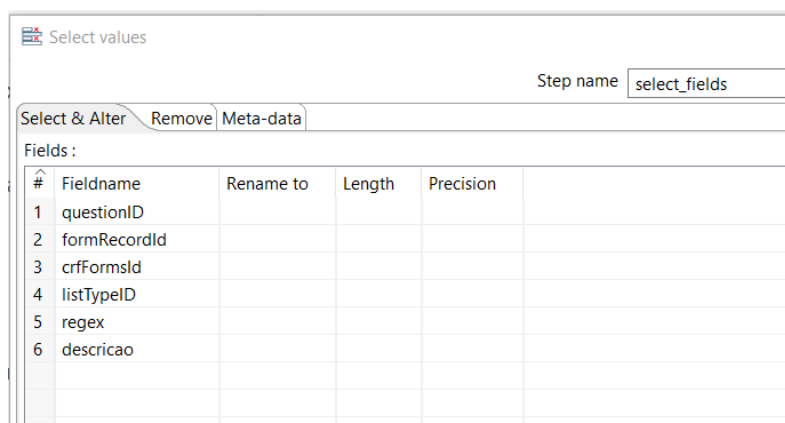
**Figura 49:** Parte final da transformação model\_yes\_unk



As próximas operações serão referentes à Figura 49. A primeira operação é a de *select\_fields* (operação 144), que faz a limpeza dos campos obtidos para filtrar aqueles que serão utilizados, conforme vista na Figura 50. Alguns campos acabam sendo duplicados em operações de merge durante o funcionamento do ETL, por esse motivo é importante

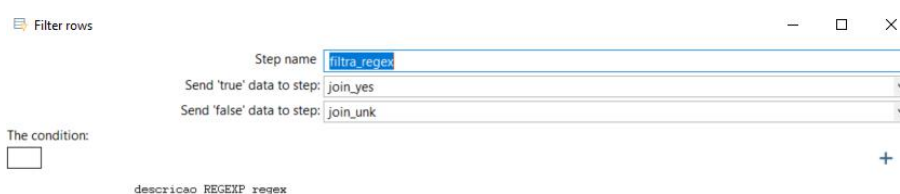
efetuar essa seleção para a limpeza dos dados e otimização da execução.

**Figura 50:** Filtragem dos campos que serão utilizados na lógica de decisão da transformação de questões múltiplas escolhas.



Em seguida o fluxo é tratado na operação *remove\_break\_line* (operação 145), responsável pela retirada das quebras de linha no texto oriundo das fontes de informação. Vale lembrar que os dados analisados são, em sua maioria, provenientes de campos livres escritos por profissionais de saúde. Por limitações da ferramenta, campos de texto com quebra de linha podem apresentar resultados incorretos após análises com regex. O texto, já tratado deste problema, segue para uma operação de filtro regex.

**Figura 51:** Operação de filtragem por regex.



Conforme pode ser visto na Figura 51, a operação de filtragem por regex nada mais é do que uma busca no campo texto, procurando uma expressão que se encaixe com o regex cadastrado no banco de dados para a questão a ser respondida. Caso o regex seja encontrado, o registro segue para a operação *join\_yes* (operação 163) caso contrário, segue para a operação *join\_unk* (operação 147). Na Figura 52 é possível observar que para cada questão cadastrada no banco, há um registro de regex, utilizado para efetuar a busca pelo texto livre.

**Figura 52:** Regex utilizado nas buscas cadastrado no banco para cada questão individualmente

questionID	description								regex	regex_delimitador
84	Falci-parum malaria	6	11	7	32	NULL	1	4	.*(malaria falci-parum)+.*	[NULL]
85	HIV	6	11	7	32	NULL	1	4	.*(HIV)+.*?	[NULL]
88	Non-Falci-parum malaria	6	11	7	32	NULL	1	4	.*(malaria não falci-parum)+.*	[NULL]

A tabela escolhida para o regex foi uma cópia da `tb_questions`, já mencionada no início deste capítulo, com a inserção de 3 campos novos, o campo `regex` utilizado para filtragem nesta transformação, o `regex_delimitador` e o `regex_limpeza` que serão explicados na transformação `model_extract` mais adiante neste capítulo.

As duas operações seguintes iniciam os fluxos de respostas positiva, "Sim", da questão ou inconclusiva, "Desconhecido" (operações 159 e 160). Entretanto ambas necessitam do Id da resposta correspondente. Cada questão múltipla escolha possui um id chamada `listTypeId`, que é uma chave estrangeira para uma tabela de respostas possíveis. Antes de gravarmos o resultado do filtro de regex anterior, precisamos saber o id da resposta positiva e inconclusiva para cada questão. Esse conjunto de operações responsável por buscar essa informação pode ser visto na parte em amarelo da Figura 49.

A lógica de busca pelo id da resposta positiva e inconclusiva começa pela operação `select_ListTypeId` (operação 154) que irá isolar o `ListTypeId` procurado. Em seguida ele será utilizado como parâmetro de busca na tabela de `listValues` que contém as opções de resposta múltipla escolha para cada tipo de questão como pode ser visto na Figura 53. O resultado da `query` é filtrada pelas operações `filter_answer` e `filter_rows` (operação 157), que garante que a resposta positiva vai para o fluxo de resposta positiva e que a inconclusiva vai para o fluxo inferior de respostas "Desconhecido".

Seguindo a arquitetura da Figura 49, podemos dividir o restante do fluxo em duas partes, o superior relacionado a resposta positiva e o inferior relacionado a resposta inconclusiva. Para respostas positivas a primeira operação é a de `join_yes` (operação 163) que une os dados oriundos das fontes de informação com a resposta positiva encontrada anteriormente, após isso uma operação de `prepare_answer` (operação 159) trata o dado semelhante ao que foi efetuado na transformação Q1, explicada anteriormente. Em seguida a operação de `Salva` (operação 160), faz a gravação da resposta positiva no banco de dados.

As três operações de token seguintes são apenas para limpeza dos dados e inserção de



um campo auxiliar (operações de 160 a 162), para diminuir a quantidade de dados combinados na operação *join\_token* (operação 150). Isso se faz necessário pois como para algumas questões, como as de módulo 3, é possível que diversas fontes de dados sejam examinadas, sendo assim, temos que fazer um merge do fluxo superior com o inferior, para garantir que as respostas positivas sejam gravadas primeiro. Visto que, se o regex for encontrado uma vez, a resposta é positiva, nenhuma, é desconhecida.

**Figura 53:** Relação de respostas possíveis para cada tipo de pergunta múltipla escolha.

The screenshot shows a software interface with a window titled 'Select values'. In the top right corner, there is a 'Step name' field containing the text 'select\_fields'. Below this, there are three tabs: 'Select & Alter', 'Remove', and 'Meta-data'. The 'Select & Alter' tab is active, and below it is a table labeled 'Fields :'. The table has five columns: '#', 'Fieldname', 'Rename to', 'Length', and 'Precision'. It contains six rows of data.

#	Fieldname	Rename to	Length	Precision
1	questionID			
2	formRecordId			
3	crfFormId			
4	listTypeID			
5	regex			
6	descricao			

O fluxo inferior, de respostas inconclusivas, é idêntico ao superior, apenas com o *join\_token* (operação 150), unindo os dois fluxos como explicado acima, e a operação de Fim (operação 153), simbolizando o fim da sub-transformação.

### 3.5.3 Resolução de questão de extração (*model\_extract*)

Algumas questões presentes no FRC-OMS, questionam o valor de algumas informações ou dados, de forma que sua resposta não está presente em uma gama de opções, como as perguntas de múltipla escolha, mas necessitam que a informação seja extraída da fonte de dados e armazenada como resposta a questão.

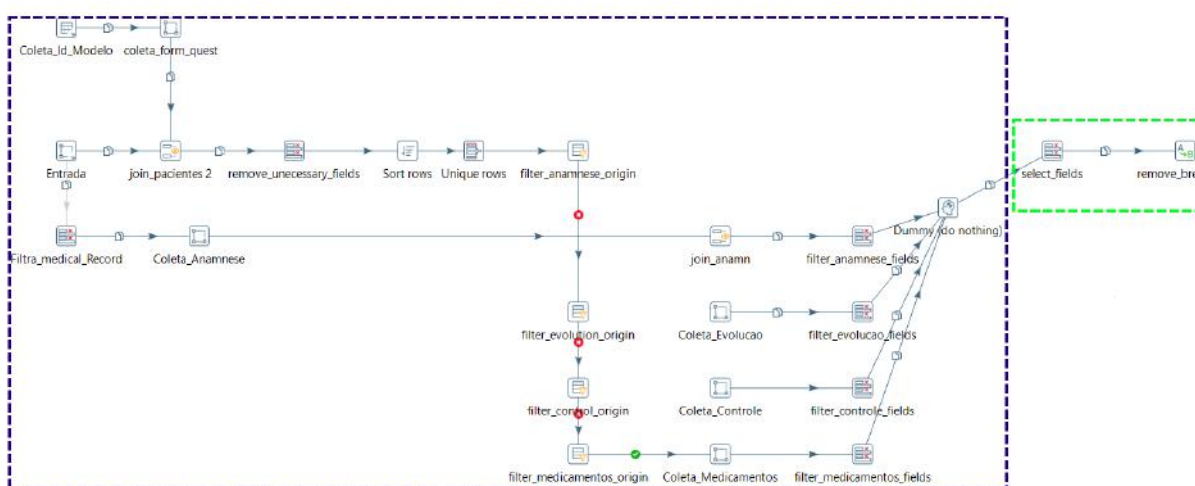
Perguntas como a temperatura corporal, a taxa de oxigenação, peso ou milimetragem de uma medicação, são exemplos de questões que necessitam de extração de dados (valores). A transformação *model\_extract* é responsável por responder a maioria das

perguntas que possuem essa característica, algumas questões, como as que precisam da quantidade de dias de internação e outras semelhantes, são respondidas em transformações específicas que serão explicadas mais à frente.

A arquitetura desta transformação é muito semelhante à transformação que resolve questões múltiplas escolhas. A transformação pode ser dividida em duas partes, sendo a primeira responsável pela coleta dos dados nas fontes de informações e a segunda com a lógica de extração da informação. Como não há opções predeterminadas, não há a necessidade de buscar as opções de cada questão, como ocorreu na transformação explicada anteriormente.

Na Figura 54 é possível ver, destacado em azul, a lógica de extração dos dados provenientes das fontes de informação e em verde, parte da lógica para extração da resposta da questão da fonte de dados. O *Coleta\_Control* é a única fonte de dados que não está presente na transformação das questões múltiplas escolhas, mas possui exatamente a mesma lógica já explicada para a extração de evolução e medicamentos, logo não será debatido novamente neste capítulo. O restante da lógica não possui alteração em relação a transformação *model\_yes\_unk*.

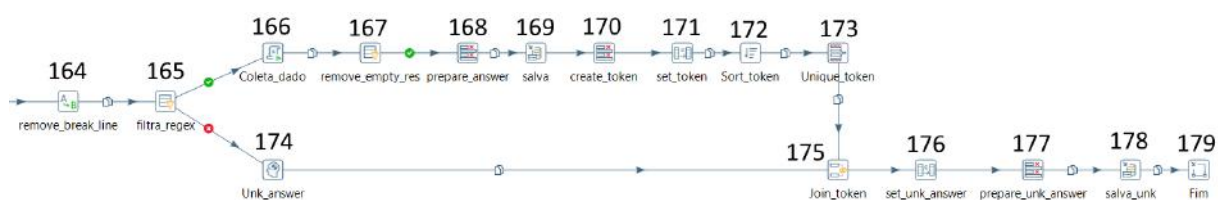
**Figura 54:** Lógica para extração dos dados das tabelas de informações do HUGG.



Na Figura 55 é possível ver toda a lógica de extração da informação. Com um início muito semelhante ao visto na transformação de questões múltiplas escolha, o fluxo primeiro é tratado na operação de *remove\_break\_line* (operação 164) para remoção da quebra de

linha do texto obtido das fontes de informação na parte anterior da transformação. Em seguida, o fluxo passa pela operação de *filtra\_regex* (operação 165). Esse regex inicial é utilizado para saber se a fonte de informação detém ou não o dado que estamos procurando. Essa ação é necessária para satisfazer uma limitação da plataforma, que não possui suporte para realizar extração de dados inexistentes na fonte de informação. Desta forma, antes de um registro passar pela transformação de extração, precisamos garantir que a informação existe para evitar possíveis erros durante a execução do ETL.

**Figura 55:** Lógica da coleta de respostas de questões com extração de texto livre.



Registros que não possuem a informação vão para a operação *Unk\_answer* (operação 174), que utilizando a mesma lógica de token explicada na transformação anterior, salva respostas com o valor "Desconhecido", após a gravação dos resultados que encontraram dados, na parte superior do fluxo.

Registros que encontraram a informação seguem o fluxo até a operação de *Coleta\_dado* (operação 166), essa operação é um script que realiza uma busca e extração através de dois regex cadastrados no banco, o *regex\_delimitador* e o *regex\_limpeza*. O primeiro insere um delimitador imediatamente antes do valor a ser extraído através de uma operação de *replace* no texto obtido das fontes de dados. Logo em seguida um novo *replace* é feito, retirando todo o texto que existe até o delimitador. Por último, o *regex\_limpeza* é utilizado na última operação de *replace* para retirar todo o texto após o dado a ser extraído. Como resultado final, apenas a informação que queremos existe no campo de texto.

Por limitações da ferramenta, não é possível utilizar um regex para fazer a extração de uma fonte de texto, apenas podemos procurar para encontrar se o texto se enquadra ou não no regex, e efetuar operações de *replace*. Por esse motivo, foi preciso implementar o script *javascript* para simular uma operação de busca e extração como mostrado na Figura 56.

**Figura 56:** Script de extração utilizado na transformação *model\_Extract*.



```

Step name Coleta_dado

Java script:
Script 1
//Script here
var regex = new RegExp(regex_delimitador);
descricao = descricao.replace(regex, 'DELIMITADOR2806');
var regex2 = new RegExp('^.*DELIMITADOR2806');
descricao = descricao.replace(regex2, '');
var regex3 = new RegExp(regex_limpeza);
descricao = descricao.replace(regex3, '');

```

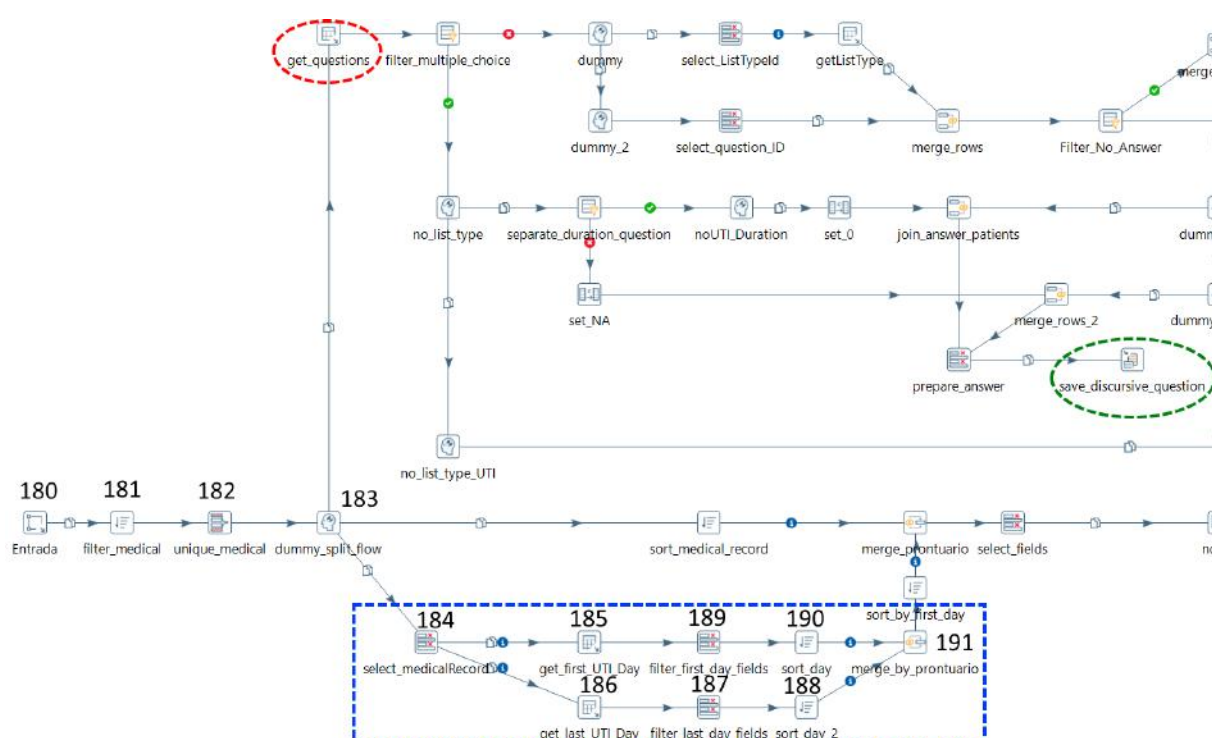
Após a operação que realiza a coleta do dado, há um filtro para garantir que dados vazios não estejam sendo salvos no banco de dados final. Na sequência são feitas as mesmas operações encontradas na transformação de questões múltiplas escolhas como a preparação da resposta, gravação no banco de dados e a gravação das respostas inconclusivas.

### 3.5.4 Duração e dias na UTI (*UCI\_getAndSave*)

Algumas questões do módulo 3 são relativas a datas e durações, especialmente a entrada, saída e dias em que o paciente ficou internado em uma unidade de terapia intensiva. Como pode ser observado, questões nesse sentido não possuem qualquer semelhança às lógicas explicadas nas duas transformações anteriores, de extração de dados e resolução de perguntas múltiplas escolhas. Regex, principal ferramenta dessas duas transformações, não pode ser utilizado para os cálculos dessas questões. Por estes motivos, foi criada uma nova transformação, a *UCI\_getAndSave*, com o intuito de responder 4 questões principais quando possível: O dia de entrada na UTI, dia de saída da UTI, se houve passagem pela UTI e duração de dias do paciente na UTI.

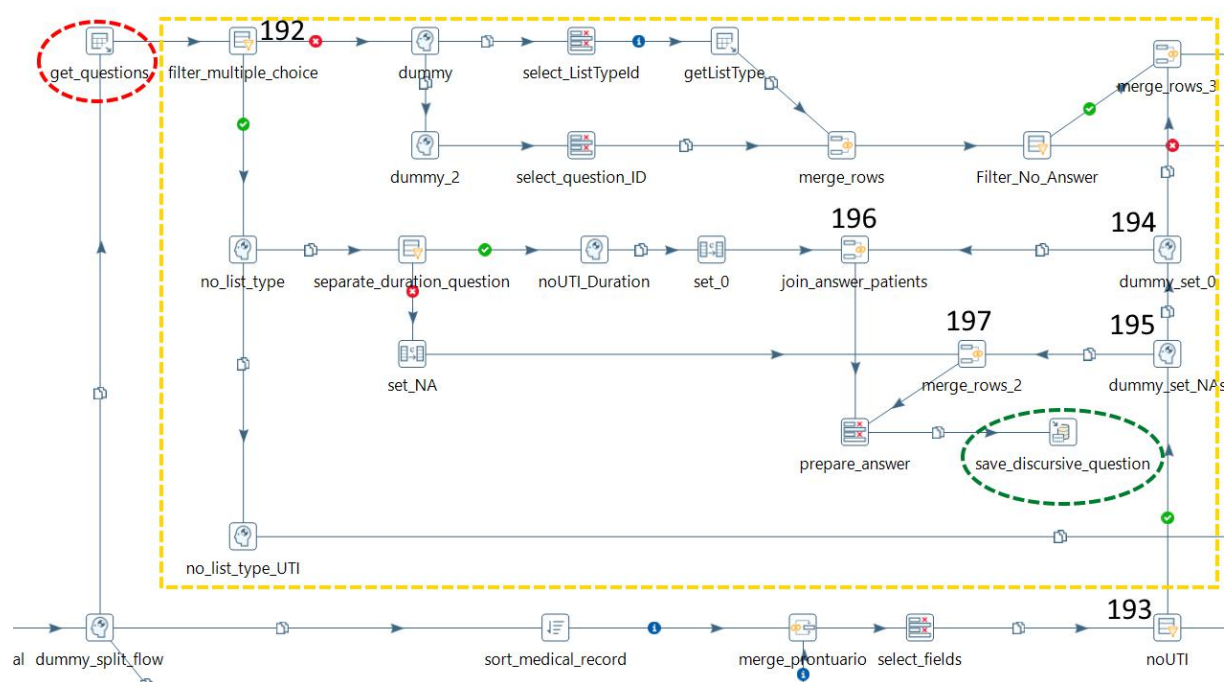
Por ser uma transformação com maior complexidade que as anteriores, vamos passar por seus setores focando na resolução de cada uma das 4 questões em que ela se propõe a responder. Na Figura 57 é possível observar a primeira parte da transformação, que se inicia na operação Entrada (operação 180) na parte inferior direita, seguindo por 2 operações de ordenação e limpeza (operações 181 e 182), para retirar registros duplicados. Ao chegar na operação *Dummy\_split\_flow* (operação 183), o fluxo vai para a lógica evidenciada pelo quadrado azul e pela operação circulada em vermelho.

Figura 57: Parte 1 da Arquitetura da transformação *UCI\_GetSave*.



Na lógica destacada em azul, encontramos as operações responsáveis pela obtenção do primeiro e do último dia de internação. Com um *select\_medicalRecord* (operação 184) como primeira operação, deixando apenas o Id do paciente, é feita uma busca pelo primeiro registro do paciente na UTI e o último (operações 185 e 186), o retorno de cada uma das duas queries é tratado nas operações de seleção seguinte e de ordenação (operações 187, 188, 189 e 190). O fluxo se mergeia através da *merge\_by\_prontuario* (operação 191), unindo o primeiro e último dia do paciente em um único registro, retornando ao fluxo principal e seguindo para o filtro na UTI, responsável por dividir os pacientes que estiveram na UTI daqueles que não. Para os que não tiveram na UTI, o fluxo segue para o quadrado destacado em amarelo na Figura 58.

Figura 58: Parte 2 da arquitetura da transformação *UCI\_GetSave*.



A operação *Get\_Questions* envolvida pelo tracejado vermelho é responsável por buscar as questões que serão respondidas nesta transformação. Assim como os dados de primeiro e último dia da UTI de cada paciente, ela também alimenta o conjunto de operações evidenciado em amarelo pela Figura 58. A primeira operação após a extração das questões é a *Filter\_multiple\_choices* (operação 192) que separa as questões que são múltiplas escolhas das que não são. Como apenas 3 questões são discursivas, sendo elas a duração de dias na UTI, primeiro dia e último, para os pacientes que não passaram na UTI devemos colocar 0, N/A e N/A respectivamente.

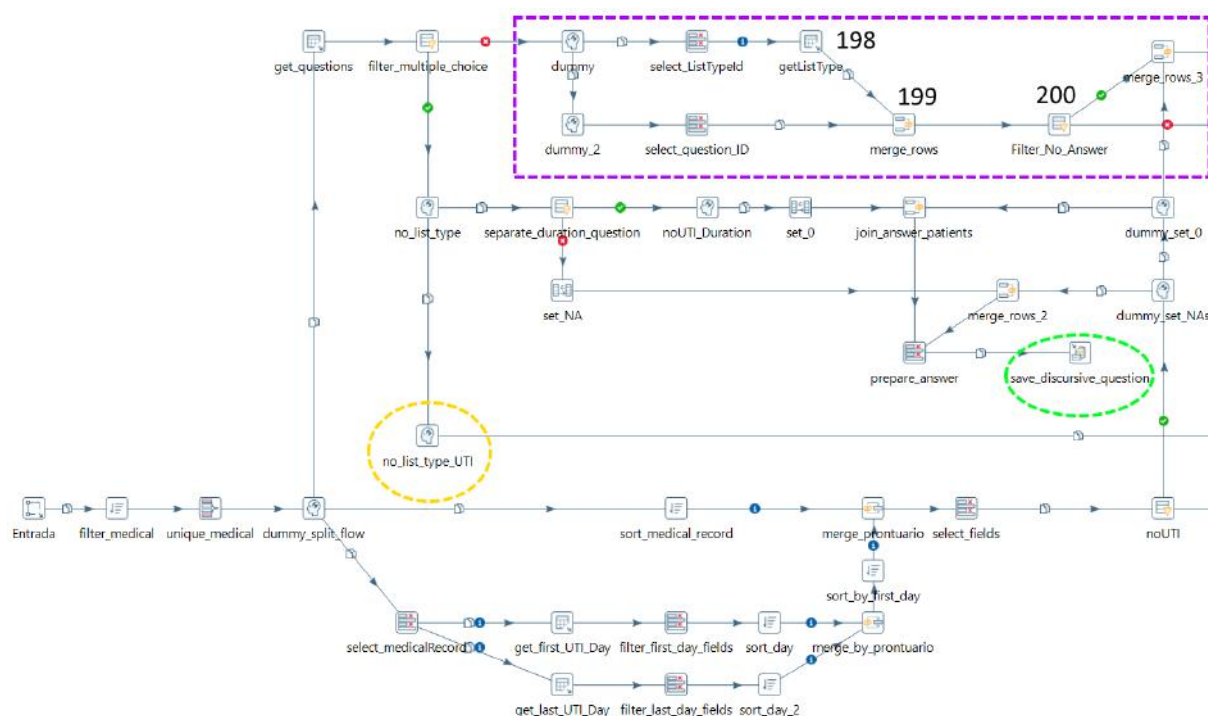
A operação *noUTI* (operação 193) seleciona os registros dos pacientes que não passaram pela UTI, eles são levados a duas operações *dummies*, *dummy\_set\_0* e *dummy\_set\_NAs* (operações 194 e 195 respectivamente) que vão ser mergeadas nas operações de merge *join\_answer\_patients* (operação 196), que contará com em registrar 0 dias de duração na UTI para esses pacientes, e na operação *Join rows (cartesian product)* (operação 197) que responde às questões de primeira e último dia na UTI como N/A, salvando seus resultados na operação salva, evidenciadas em verde na Figura 58.

As questões de duração de dias na UTI e primeiro e último dia na unidade de terapia

intensiva, também devem ser respondidas para pacientes que de fato adentraram na unidade durante algum período de suas internações, para isso há uma operação *dummy*, evidenciada em amarelo na Figura 59, que irá enviar essas questões para a próxima parte da lógica.

As questões de múltiplas escolhas precisam que suas opções sejam coletadas no banco de dados, para termos os id's das respostas positivas e negativas, assim como aconteceu na transformação *model\_yes\_unk* explicada no início deste capítulo. A parte da lógica evidenciada em roxo na Figura 59 simboliza essa operação. É recolhido o id do *ListType* (operação 198), feita a busca por essas opções, que posteriormente são mergeadas ao fluxo novamente na operação *merge\_rows* (operação 199) e filtradas em *Filter\_No\_Answer* (operação 200) para separar as respostas positivas das negativas.

**Figura 59:** Parte 3 da arquitetura da transformação *UCI\_GetSave*.



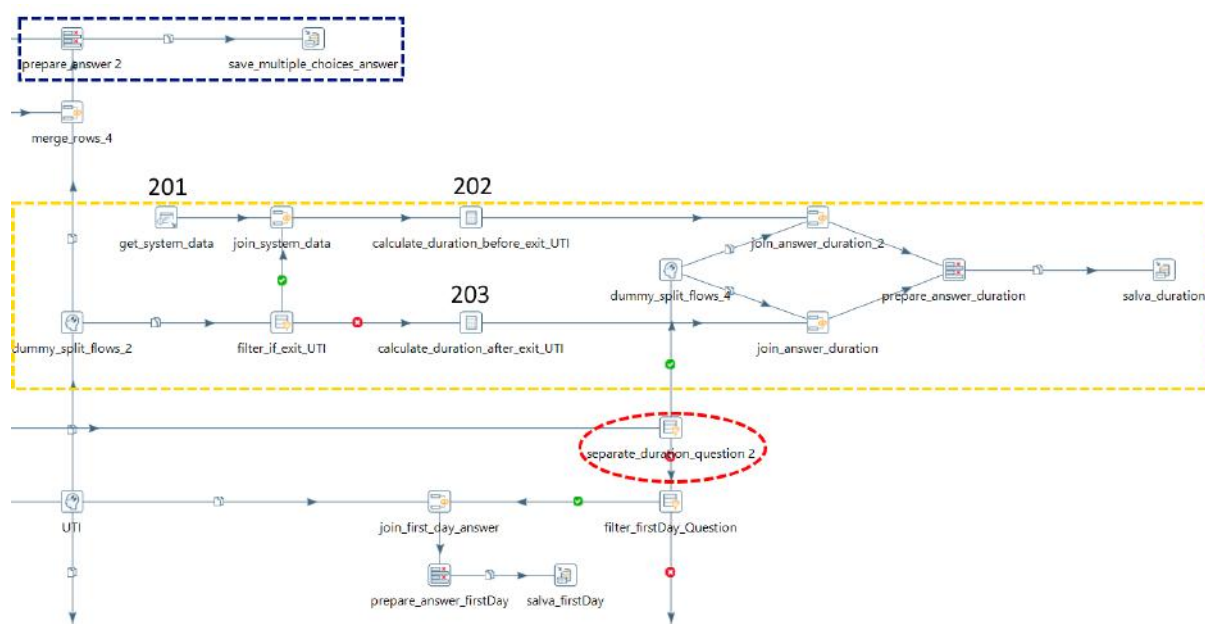
Na Figura 60 podemos ver a segunda parte da lógica desta transformação. A parte superior, evidenciada em azul marinho, grava como “Não” a resposta da questão se o paciente esteve na UTI, ela vem do fluxo anterior que identifica os pacientes que não tiveram passagem pela unidade de terapia intensiva.

A lógica destacada em amarelo trata de calcular a duração da estadia do paciente na

UTI, através dos campos de primeiro último dia, é feito um cálculo a partir da data atual do sistema, coletada na operação *get\_system\_data* (operação 201), e calculada nas operações *calculate\_duration\_before\_exit\_UTI* (operação 202) e *calculate\_duration\_after\_exit\_UTI*, (operação 203) até ter a resposta salva na operação *salva\_duration* ao final da transformação, como visto na figura.

A operação evidenciada em vermelho é responsável por filtrar a questão que está sendo respondida. Como explicado anteriormente, colhemos todas as questões que serão respondidas nesta transformação e as filtramos para que cada parte da lógica seja responsável por responder à questão correta. Na operação *separate\_duration\_question 2*, é filtrada a questão de duração total na UTI, que é respondida pela lógica evidenciada em amarelo.

**Figura 60:** Parte 4 da arquitetura da transformação *UCI\_GetSave*.



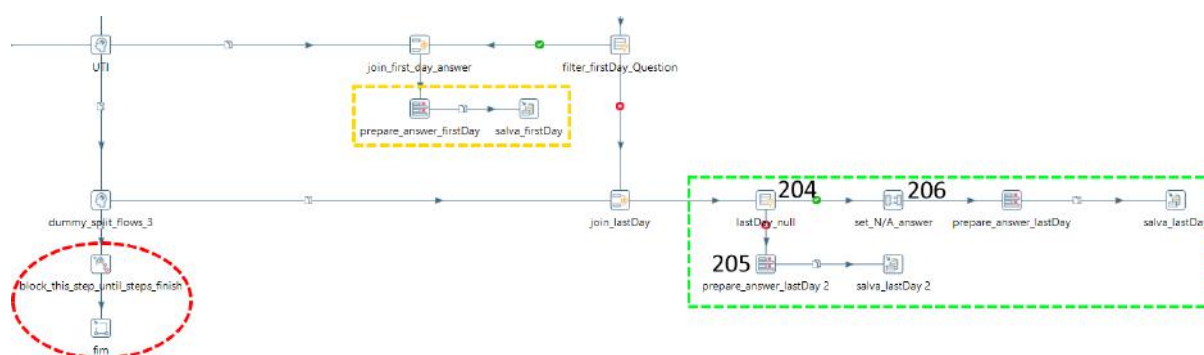
Na Figura 61 é possível observar a parte final da transformação. Evidenciado em amarelo temos as operações responsáveis por responder o primeiro dia de internação do paciente. Como todos os dados deste fluxo neste ponto da lógica são de pacientes que tiveram passagem pela UTI, temos a certeza de que todos os pacientes detêm essa informação.



A parte da lógica evidenciada em verde na Figura 61 possui a função de responder qual foi o último dia de internação do paciente na unidade de terapia intensiva. Como nem todos os pacientes que passaram por uma UTI detêm data de saída da mesma, é preciso verificar se essa informação não é nula na operação *lastDay\_null* (operação 204). Caso a data esteja presente, o registro segue para a operação *prepare\_answer\_lastDay 2* (operação 205) que irá limpar os campos para a gravação no banco, caso contrário, o fluxo segue para a operação *Set field value* (operação 206) que irá preencher a resposta como N/A, visto que o paciente adentrou a UTI mas ainda não saiu da unidade até a data de geração do relatório.

A operação evidenciada em vermelho na Figura 61 é responsável por travar o fluxo de dados enquanto as operações de gravação no banco não sejam finalizadas. Isso é necessário pois como a transformação é uma sub-transformação da transformação central, se essa operação de espera não for incluída, ela pode ser finalizada no fluxo principal antes de todas as questões serem respondidas. Desta forma, garantimos que todas as questões da qual a transformação *UCI\_getSave* é responsável foram respondidas.

**Figura 61:** Parte 5 da arquitetura da transformação *UCI\_GetSave*.

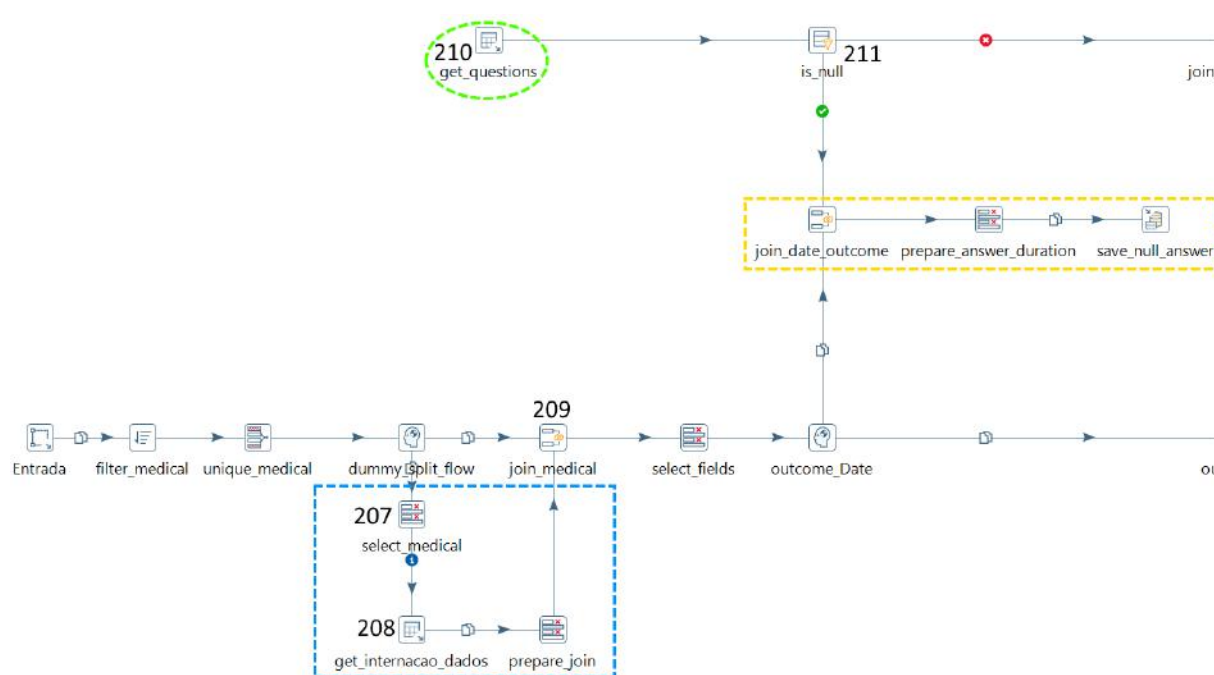


### 3.5.5 Informações de saída (*Outcome*)

As questões do módulo 3 relativas à saída do paciente, seja por alta, transferência ou falecimento são tratadas na transformação de *outcome*. Esta transformação tem o papel de responder a 2 questões com essas características, sendo elas estado de saída do paciente (alta, falecimento, transferência) e data de saída, se houver.

Na Figura 62 é possível ver a primeira parte da transformação. Evidenciado em azul temos a lógica que faz a busca pelos dados de internação do paciente. Primeiro isolando o id do paciente através da operação *select\_medical* (operação 207) que irá servir de insumo para a busca efetuada em *get\_internacao\_dados* (operação 208). Essa informação volta para o fluxo padrão dos registros através do *join\_medical* (operação 209) e é mergeada com a questão de data de saída, obtida na operação evidenciada em verde, *Get\_Questions* e filtrada por *is\_null* (operações 210 e 211 respectivamente), que verifica se a questão é discursiva. Nesta transformação, a data de alta é a única questão não múltipla escolha. A questão então é salva na lógica evidenciada em amarelo.

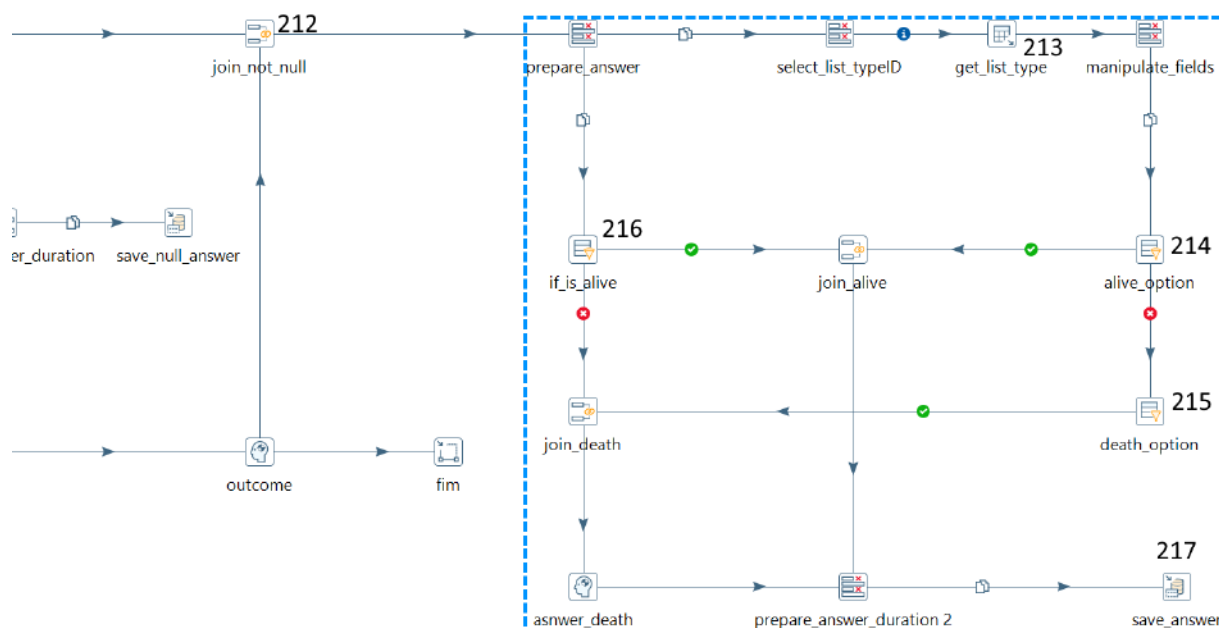
**Figura 62:** Arquitetura inicial da transformação *Outcome*.



A Figura 63 apresenta as 2 outras questões e os dados dos pacientes são unidos na operação *Join rows (cartesian product)* (operação 212) e seguem para a segunda parte da lógica. Evidenciado em azul, temos a resposta da questão do estado do paciente no momento da alta. Através da operação *get\_list\_type* (operação 213) é recolhido as opções de resposta, que são filtradas nas operações *alive\_option* e *death\_option* (operações 214 e 215), separando as respostas de alta e falecimento. Na operação *if\_is\_alive* (operação 216) é verificado se o

paciente faleceu ou não. Cada paciente é enviado para o fluxo correspondente a seu estado de saúde no momento da alta. Por fim, a resposta é salva no banco através da operação salva 2 (operação 217).

**Figura 63:** Lógica de resolução do estado de saúde do paciente no momento da alta/óbito.



### 3.5.6 Questões que referenciam outras questões (*get\_old\_questions*)

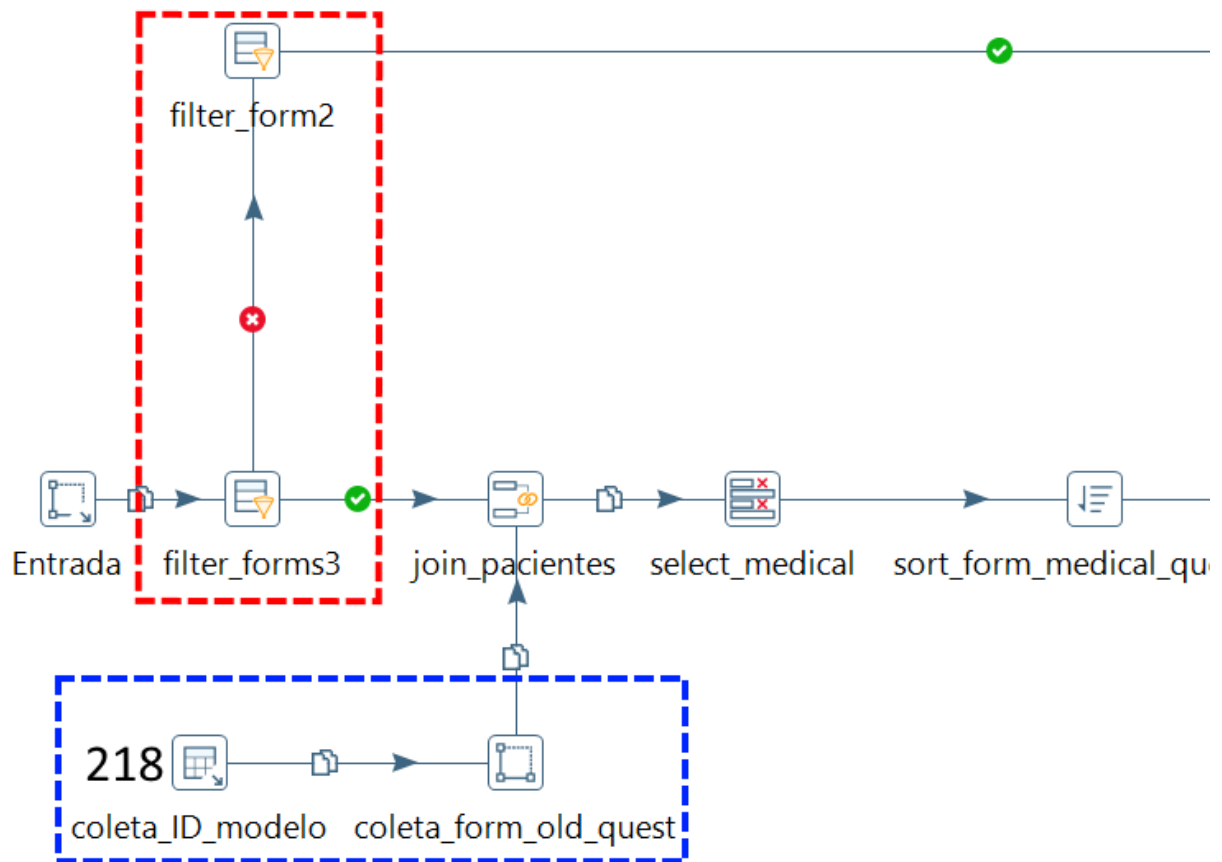
Algumas questões do módulo 3 referenciam questões já respondidas durante a geração dos módulos 1 e 2, como por exemplo se o paciente apresentou febre em algum momento durante seu período de internação. Para evitar que reprocessemos todas as informações dos pacientes para coletar um dado já tratado, há a transformação *get\_old\_questions* que colhe a resposta da questão dos módulos 1 e 2 para verificar se o paciente apresentou ou não determinada situação ou característica durante seu período hospitalizado.

Na Figura 64 é possível ver a primeira parte da lógica da transformação que colhe os dados das questões anteriores. A lógica evidenciada em azul marinho, na parte inferior, é responsável por coletar as questões que se enquadram neste modelo, a partir da operação *Coleta\_Id\_Modelo* (operação 218). O grupo de filtros evidenciados em vermelho são responsáveis por dividir os dados de entrada em perguntas do módulo 2, que servirão de fonte de informação, e do módulo 3 que serão validadas com as encontradas na lógica em azul. Todas

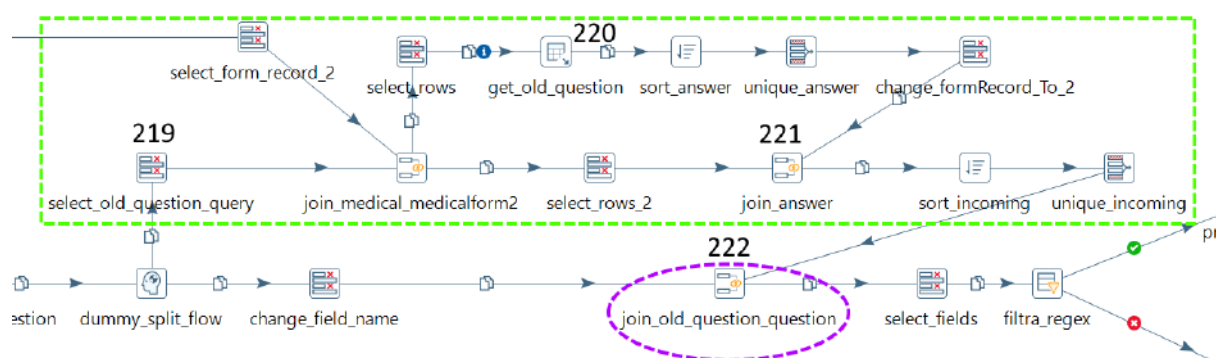
as questões encontradas pelo *Coleta\_Id\_Modelo* (operação 218) seguirão no fluxo, as demais serão descartadas.

A lógica em verde na Figura 64 é responsável por coletar a resposta da questão antiga, que foi gerada pelas transformações anteriores. As questões do módulo 3 chegam pela transformação *select\_old\_question\_query* (operação 219) e são unidas as questões do módulo 2 na operação de merge *join\_medical\_medicalform2*. A partir daí, é feita uma busca pela resposta dessa questão na operação *get\_old\_question* (operação 220). O resultado da busca volta ao fluxo na operação *Join rows (cartesian product) 2* (operação 221), unindo a questão do módulo 3 a resposta da sua semelhante, do módulo 2. O resultado disto é tratado, para evitar resultados duplicados e unido novamente no merge *join\_old\_question\_question* (operação 222). Neste ponto da lógica temos as informações do paciente, da questão a ser respondida e a resposta da questão relacionada a ela, tudo no mesmo registro no fluxo.

**Figura 64.A:** Arquitetura inicial da transformação *get\_old\_questions*.

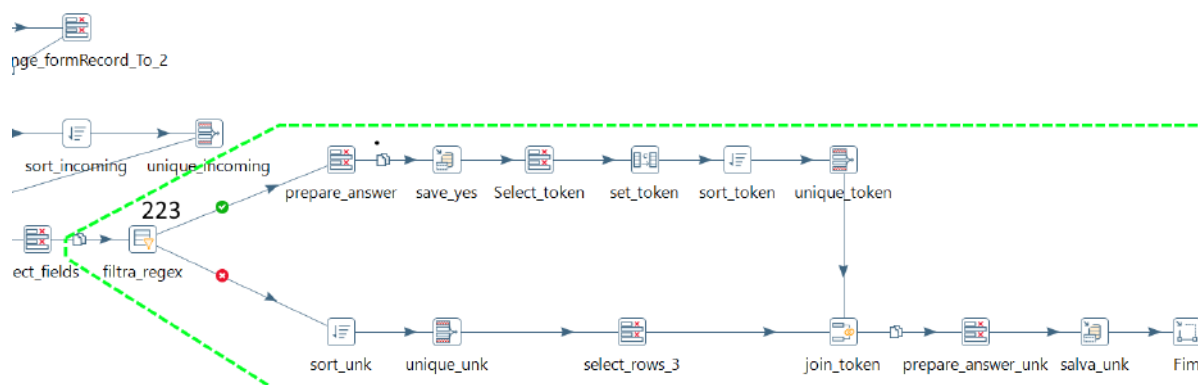


**Figura 64.B:** Arquitetura inicial da transformação *get\_old\_questions*.



Na Figura 65 podemos ver a parte final da lógica da transformação, muito semelhante ao encontrado na transformação *model\_yes\_unk*, a grande diferença fica no filtro *filtra\_regex* (operação 223) que verifica se a resposta é igual a uma negativa, como N/A, Desconhecido ou Não. Respostas diferentes destas vão para o fluxo de cima e a resposta colhida na questão relacionada é a mesma a ser salva no módulo 3. Caso contrário, o fluxo vai para a parte inferior, salvando a questão como desconhecida ou N/A.

**Figura 65:** Lógica de resposta da transformação *get\_old\_questions*.



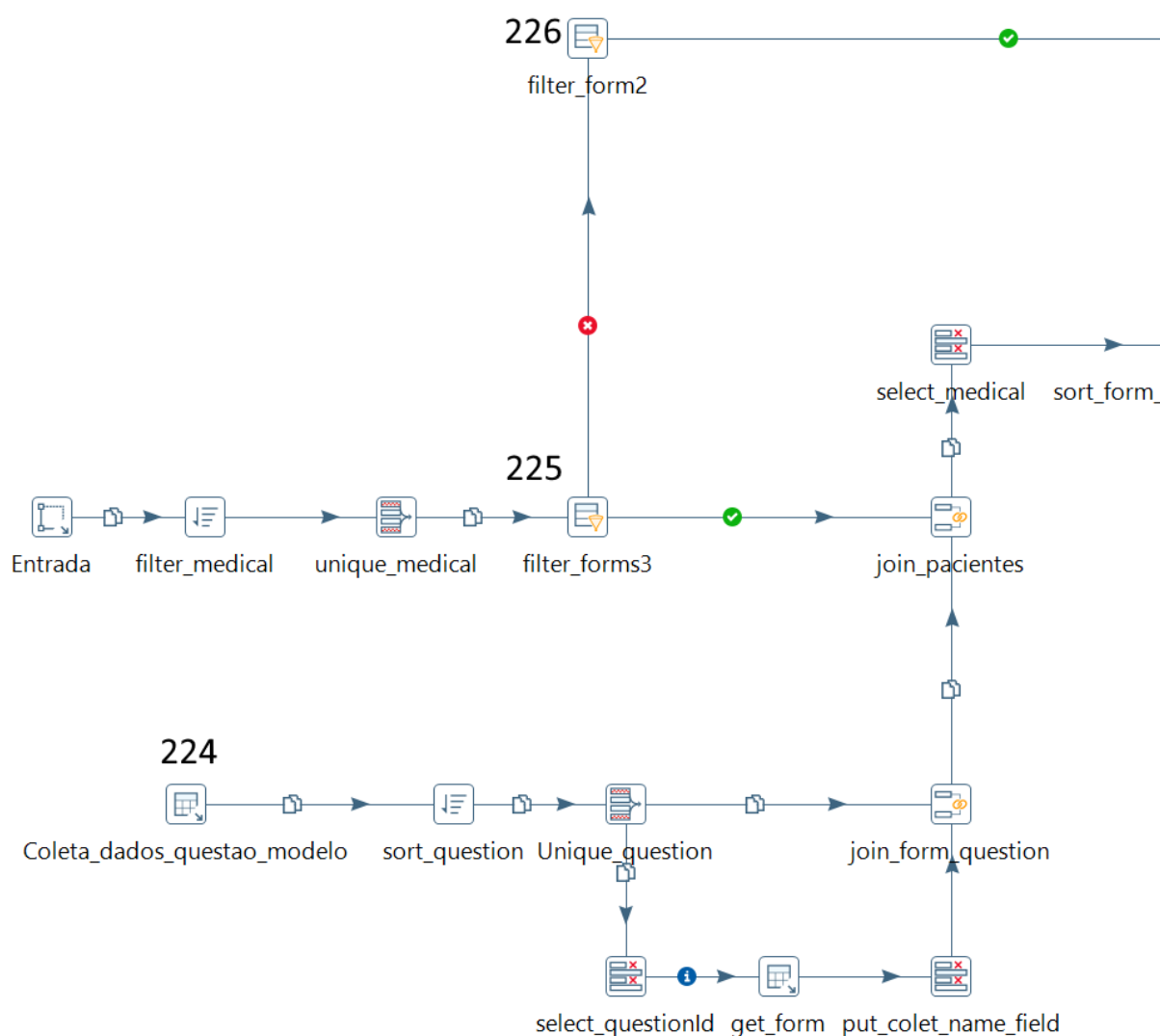
### 3.5.7 Questões com duração de dias (*DurationDays*)

Algumas questões do módulo 3 se dedicam a detalhar quantos dias o paciente apresentou certa condição. Muitas dessas questões envolvem perguntas já respondidas no módulo 2, por exemplo, por quanto tempo um paciente apresentou febre. Uma forma de responder a este questionamento é rever todas as fontes de dados para encontrar a primeira e última vez onde houve febre. Porém, como temos questionários de acompanhamento para cada dia de internação do paciente, e temos uma questão se o paciente apresentou febre naquele dia, basta buscarmos

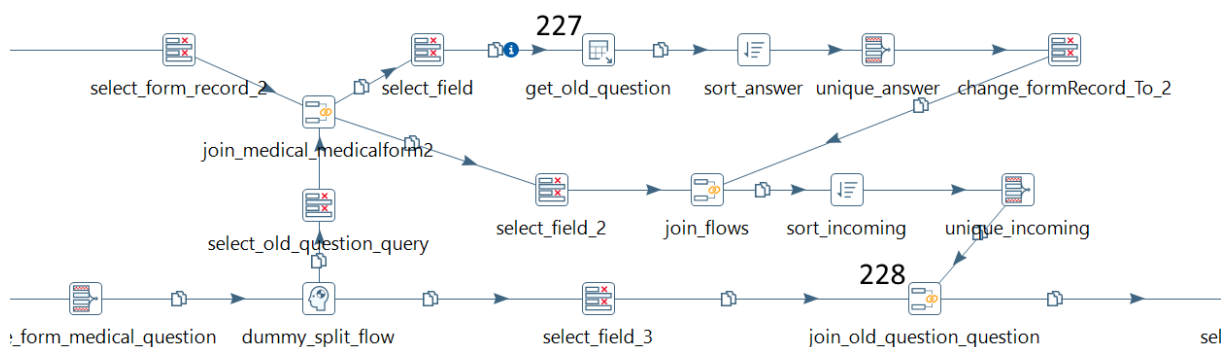
os questionários e examinarmos qual foi o primeiro e último questionário onde essa resposta foi dada como "Sim". Todas as perguntas que se caracterizam por essa descrição são respondidas na transformação *DurationDays*.

Assim como pode ser visto na Figura 66, o início da lógica da transformação é muito semelhante ao encontrado na transformação *get\_old\_questions*. As questões que serão resolvidas são recolhidas da operação *Coleta\_dados\_questao\_modelo* (operação 224), há duas operações de filtragem *filter\_form3* e *filter\_form2* (operações 225 e 226) que dividem as questões provenientes do modelo 2 e 3. Há a busca pela informação da questão do modelo 2 na operação *get\_old\_question* (operação 227) e as informações encontradas são mergeadas novamente ao fluxo principal na operação *join\_old\_question\_question* (operação 228).

**Figura 66.A:** Arquitetura inicial da transformação *DurationDays*.



**Figura 66.B:** Arquitetura inicial da transformação *DurationDays*.



As únicas diferenças são encontradas na segunda parte da lógica da transformação. Conforme pode ser visto na Figura 67. Os dados chegam até a operação *Filter rows* (operação 228) que verifica se foi encontrada ocorrência da questão antiga. Por exemplo, caso o paciente não tenha registro de febre em nenhum dia de internação, o fluxo vai para a parte amarela, inserindo a resposta de 0 dias como resposta à pergunta. Caso tenha havido ocorrência da questão procurada nos registros do paciente, o fluxo segue para a lógica evidenciada em verde na figura.

A data dos registros do módulo com a resposta positiva são obtidos em *get\_form\_date* (operação 229) e são retornados ao fluxo principal na operação *get\_first\_and\_last\_days* (operação 230) que divide o fluxo em dois. O fluxo superior ordena e obtém o último registro do módulo com a informação, enquanto o fluxo inferior faz a ordenação inversa, encontrando o primeiro registro com a informação. Esses dados então são examinados na operação *calculate\_duration* (operação 231) que calcula a diferença de dias entre eles, obtendo a informação que é convertida para inteiro na operação *transform\_to\_integer* (operação 232).

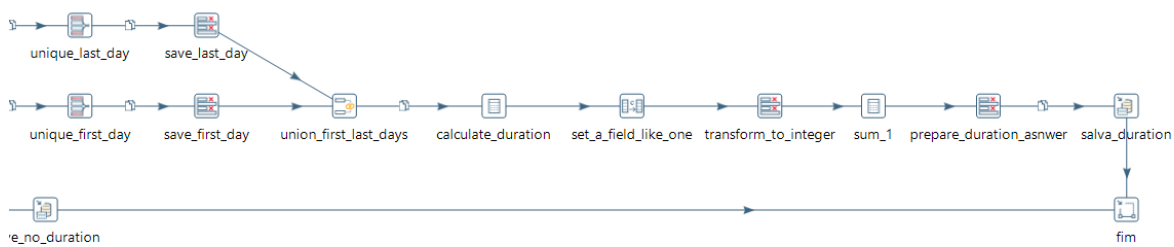
**Figura 67:** Lógica de cálculo da duração de dias em que o paciente mostrou determinado comportamento ou característica.



A parte final da lógica salva no banco de dados a quantidade de dias calculada entre o primeiro e último registro de acompanhamento calculado anteriormente como resposta à questão. O fluxo de respostas zeradas, ou seja, de pacientes que não tiveram nenhum dia

contabilizado, também é direcionado para a operação de saída da sub-transformação, como pode ser visto na Figura 68.

**Figura 68:** Parte final da arquitetura da transformação *DurationDays*





## 4 CONCLUSÃO

Através das implementações descritas neste trabalho foi possível atingir o objetivo proposto, ou seja, criar um ETL capaz de coletar as informações provenientes do banco de dados do HUGG, tratar esses dados para responder às questões do FRC-OMS e armazenar essas respostas no banco de dados da Fiocruz.

Vale frisar que a metodologia utilizada para a maior parte desta solução foi o regex, para validação e busca das informações necessárias para cada questão. Utilizar essa metodologia possibilitou padronizar a lógica de resolução de um conjunto de perguntas com o padrão, agilizando significativamente a velocidade de desenvolvimento e simplificando a manutenção das configurações. Perguntas que não seguiram o padrão, se utilizaram de lógicas individuais, mas representaram a minoria das questões cobertas.

O ETL foi executado com um total de 20 pacientes, durante seu desenvolvimento e teste. Como cada paciente ficou em média, 10 dias internado, houve centenas de registros examinados, resultando em 232 registros de módulos, com 1.621 perguntas respondidas. Das 303 questões presentes no FRC-OMS, 210 foram cobertas pela solução desenvolvida neste trabalho. A Figura 69 mostra uma parte dos dados gerados.

**Figura 69:** Resultados gerados pelo ETL.

	abc medicalRecord	abc hospitalUnitName	123 formrecordID	abc Modulo	abc QuestionGroup	123 questionID	abc question	abc answer
1	7600000	Hospital Universitário Gaffrée e Guinle - HUGG	12,005	Admission form	Demographics	110	Pregnant	N/A
2	7600000	Hospital Universitário Gaffrée e Guinle - HUGG	12,005	Admission form	Demographics	113	Gestational weeks assess	N/A
3	7600000	Hospital Universitário Gaffrée e Guinle - HUGG	12,005	Admission form	Co-morbidities	56	Chronic pulmonary dise	Unknown
4	7600000	Hospital Universitário Gaffrée e Guinle - HUGG	12,005	Admission form	Co-morbidities	53	Chronic kidney disease	Unknown
5	7600000	Hospital Universitário Gaffrée e Guinle - HUGG	12,005	Admission form	Co-morbidities	54	Chronic liver disease	Unknown
6	7600000	Hospital Universitário Gaffrée e Guinle - HUGG	12,005	Admission form	Co-morbidities	55	Chronic neurological dis	Unknown
7	7600000	Hospital Universitário Gaffrée e Guinle - HUGG	12,005	Admission form	[NULL]	166	Country	Brazil
8	7600000	Hospital Universitário Gaffrée e Guinle - HUGG	12,025	Discharge/death	Medication	34	Antimalarial agent	Yes
9	7600000	Hospital Universitário Gaffrée e Guinle - HUGG	12,057	Follow-up	Medication	34	Antimalarial agent	Unknown
10	7600000	Hospital Universitário Gaffrée e Guinle - HUGG	12,058	Follow-up	Medication	34	Antimalarial agent	Yes
11	7600000	Hospital Universitário Gaffrée e Guinle - HUGG	12,059	Follow-up	Medication	34	Antimalarial agent	Yes
12	7600000	Hospital Universitário Gaffrée e Guinle - HUGG	12,063	Follow-up	Medication	34	Antimalarial agent	Unknown
13	7600000	Hospital Universitário Gaffrée e Guinle - HUGG	12,064	Follow-up	Medication	34	Antimalarial agent	Unknown

O trabalho se deu em três partes, sendo elas a concepção da arquitetura e decisão das tecnologias utilizadas, a elaboração da primeira versão da solução e o desenvolvimento da versão final. O resultado final é de um conjunto de transformações, explicadas no capítulo anterior, e da exportação do banco, com as informações geradas pela execução do ETL, assim como sua configuração.

## 4.1 TRABALHOS FUTUROS

É importante notar que as questões que não foram cobertas nesta implementação, podem e devem servir de insumo para trabalhos futuros que porventura possam expandir a solução desenvolvida. Utilizar outras tecnologias como aprendizagem de máquina e métodos aperfeiçoados podem aumentar a gama de perguntas incluídas no processo.

Uma das possíveis vertentes de estudo é a IA implementada para aumentar a acurácia e possibilitar responder perguntas que necessitem de interpretação do que foi escrito pela junta médica durante a internação do paciente. Questões com o escopo de negação, como por exemplo “não houve uso de ventilação mecânica” são bons exemplos de perguntas que são escassamente cobertas na versão atual e que podem se beneficiar da inteligência artificial para serem levadas em consideração.

Para este tipo de implementação, será necessária a criação de uma árvore de terminologia. Além disso, recomenda-se o emprego de pesquisadores da área da saúde, que possam traduzir e ajudar a criar uma base de ML, que possibilite o aprendizado através de um banco de treinamento.

Outros pontos que podem ser trabalhados futuramente incluem: (i) o aumento da gama de dados analisados; (ii) a inserção de novos hospitais e fontes de informação; e (iii) o uso do software para o cadastro de informações provenientes de outros formulários, a partir da atualização das configurações de execução presentes no banco de dados.

## REFERÊNCIAS

- Alves, A. R. **Surtos de COVID-19 no leste e no centro da Europa fazem casos diários duplicarem na UE.** Disponível em <https://oglobo.globo.com/mundo/surtos-de-COVID-19-no-leste-no-centro-da-europa-fazem-casos-diarios-duplicarem-na-ue-25263902>. Acessado em 28 de Outubro, 2021.
- Baddini, B., e Fernandes, D. **Primeira pessoa é vacinada contra COVID-19 no Brasil.** Disponível em <https://www.cnnbrasil.com.br/nacional/primeira-pessoa-e-vacinada-contra-COVID-19-no-brasil/>. Acessado em 25 de Outubro, 2021.
- Cardim, M. E., e Lima, B. **Lockdown avança pelo país e chega a 18 cidades de cinco estados.** Disponível em <https://www.correiobraziliense.com.br/app/noticia/brasil/2020/05/08/interna-brasil,852582/lockdown-avanca-pelo-pais-e-chega-a-18-cidades-de-cinco-estados-veja.shtml>. Acessado em 24 de Outubro, 2021.
- Campos, M.L.M., Borges, V.; Lopes, G. R.; Cavalcanti, M. C.; Moreira J.; Cruz, S. M. S. (2021). **VODAN BR – uma plataforma de apoio para dados COVID-19 seguindo os princípios FAIR.** In: SALES, Luana Farias; VEIGA, Viviane dos Santos; HENNING, Patrícia; SAYÃO, Luís Fernando (org.). *Princípios FAIR aplicados à gestão de dados de pesquisa.* Rio de Janeiro: Ibict, 2021. p. 253 - 270. DOI: 10.22477/9786589167242.cap18.
- da Saúde, O. P.-A. **OMS declara emergência de saúde pública de importância internacional por surto de novo coronavírus.** Disponível em <https://www.paho.org/pt/news/30-1-2020-who-declares-public-health-emergency-novel-coronavirus>. Acessado em 20 de Outubro, 2021.
- De Sá, D. M. **Especial COVID-19: Os historiadores e a pandemia.** Disponível em <http://www.coc.fiocruz.br/index.php/pt/todas-as-noticias/1853-especial-COVID-19-os-historiadores-e-a-pandemia.html>. Acessado em 20 de Outubro, 2021.
- Fette, G. et al. **Information Extraction from Unstructured Electronic Health Records and Integration into a Data Warehouse.** In: GMDS 2012. 1 jan. 2012. Disponível em:

[https://www.researchgate.net/publication/264238057\\_Information\\_Extraction\\_from\\_Unstructured\\_Electronic\\_Health\\_Records\\_and\\_Integration\\_into\\_a\\_Data\\_Warehouse](https://www.researchgate.net/publication/264238057_Information_Extraction_from_Unstructured_Electronic_Health_Records_and_Integration_into_a_Data_Warehouse). Acessado em: 01 nov. 2022

Ford, E. et al. **Extracting information from the text of electronic medical records to improve case detection: a systematic review.** AMIA. Anais...Journal of the American Medical Informatics Association Advance Access: OXFORD, fevereiro 2016.

G1. **Brasil registra 2.349 mortes em 24 horas, novo recorde desde início da pandemia; média móvel também aumenta.** Disponível em <https://g1.globo.com/bemestar/coronavirus/noticia/2021/03/10/brasil-registra-2349-mortes-em-24-horas-novo-recorde-desde-inicio-da-pandemia-media-movel-tambem-aumenta.ghtml>. Acessado em 28 de Outubro, 2021.

G1. **Brasil vive 'início de 2ª onda' de COVID por falta de testes, de política centralizada e de isolamento social, apontam pesquisadores.** Disponível em <https://g1.globo.com/bemestar/coronavirus/noticia/2020/11/23/brasil-vive-inicio-de-2a-onda-de-COVID-por-falta-de-testes-de-politica-centralizada-e-de-isolamento-social-apontam-pesquisadores.ghtml>. Acessado em 25 de Outubro, 2021.

G1. **Com nova metodologia, número de casos de coronavírus na china vai a 59,8 mil; mortes somam 1,3 mil.** Disponível em <https://g1.globo.com/ciencia-e-saude/noticia/2020/02/13/com-nova-metodologia-numero-de-casos-de-coronavirus-na-china-vai-a-598-mil-mortes-somam-13-mil.ghtml>. Acessado em 26 de Outubro, 2021.

G1. **Mortes e casos de coronavírus nos estados.** Disponível em <https://especiais.g1.globo.com/bemestar/coronavirus/estados-brasil-mortes-casos-media-movel/?ga=2.95146524.1811220905.1636081925-55a974b3-192d-2835-4c04-dee87da8ebf9>. Acessado em 28 de Outubro, 2021.

G1. **Vacinação contra a COVID: 55,89% da população está totalmente imunizada; 72,97% tomou a 1ª dose.** Disponível em <https://g1.globo.com/saude/coronavirus/vacinas/noticia/2021/11/06/vacinacao-contra-a-COVID-5589percent-da-populacao-esta-totalmente-imunizada-7297percent-tomou-a-1a-dose.ghtml>. Acessado em 28 de Outubro, 2021.

Henning, P. C., e etal. **Go fair e os princípios fair: o que representam para a expansão dos dados de pesquisa no âmbito da ciência aberta.** Em *Questão*, Porto Alegre E-ISSN 1808-5245, nº especial ConfOA (2019), 389–412.

Holton, K. **Imunização no reinounido: Mulher de 90 anos é 1ª vacinada contra COVID-19.** Disponível em <https://www.cnnbrasil.com.br/internacional/britanica-de-90-anos-e-primeira-a-receber-a-vacina-da-pfizer-fora-dos-testes/>. Acessado em 24 de Outubro, 2021.

Jacobsen, A., Kaliyaperumal, R., da Silva Santos, L. O. B., Mons, B., Schultes, E., Roos, M., e Thompson, M. **A generic workflow for the data fairification process.** *Data Intelligence 2* (2020), 56–65.

Oliveira, Lanna Mayra Silva. **Caracterização do conceito de modularidade no desenvolvimento de linguagens de programação.** 2017. 58 f. Monografia (Graduação em Sistemas de Informação) – Instituto de Ciências Exatas e Aplicadas, Universidade Federal de Ouro Preto, João Monlevade, 2017.

Madeiro, C. **COVID-19 já é a maior causa de mortes no brasil registrada em um único ano.** Disponível em <https://noticias.uol.com.br/saude/ultimas-noticias/redacao/2020/09/15/com-133-mil-obitos-COVID-ja-tem-recorde-como-causa-morte-no-pais-em-um-ano.htm>. Acessado em 24 de Outubro, 2021.

Mons, B. **The VODAN in: support of a fair-based infrastructure for COVID-19.** *European Journal of Human Genetics*, 28 (2020), 724–727.

Morales, J. **Qual foi a vacina desenvolvida mais rápido na história?** Disponível em <https://guiadoestudante.abril.com.br/estudo/qual-foi-a-vacina-desenvolvida-mais-rapido-na-historia/>. Acessado em 28 de Outubro, 2021.

Meystre, S. M. et al. **Extracting Information from Textual Documents in the Electronic Health Record: A Review of Recent Research.** *MIA Yearbook of Medical Informatics* 2008.

Moreira, A., e Pinheiro, L. **Oms declara pandemia de coronavírus.** Disponível em <https://g1.globo.com/bemestar/coronavirus/noticia/2020/03/11/oms-declara-pandemia-de->

coronavirus.html. Acessado em 22 de Outubro, 2021.

**SANAR. Linha do tempo do coronavírus no brasil.** Disponível em <https://www.sanarmed.com/linha-do-tempo-do-coronavirus-no-brasil>. Acessado em 24 de Outubro, 2021.

Veiga, V., Campos, M. L., Silva, C. R. L. da, Henning, P., & Moreira, J. (2021). **VODAN BR: a gestão de dados no enfrentamento da pandemia coronavírus.** Páginas a&b: Arquivos E Bibliotecas, 51–58. Obtido de <https://ojs.letras.up.pt/index.php/paginasueb/article/view/10220>.

**TARGET TRUST. Curso Pentaho Business Intelligence. 2021.** Disponível em: <https://www.targettrust.com.br/curso/pentaho-business-intelligence/>. Acesso em: 09 abr. 2023.

Know Solution. **Pentaho BI: conheça as vantagens de utilizar essa solução.** Disponível em: <https://www.knowsolution.com.br/pentaho-bi-conheca-as-vantagens-de-utilizar-essa-solucao/>. Acesso em: 09 abr. 2023.

## ANEXO A – EXTRATO DE ATAS DE REUNIÕES

### Decisões e Propostas Apresentadas

1 – Referente ao Mapeamento das informações do AGHU para o FGPC-OMS:

1.1 Quanto aos diferentes tipos de qualificações de Anamnese e de Evolução registradas no Prontuário:

- Foram identificados para os pacientes de COVID, os seguintes tipos de qualificação: “Medicina”, “Enfermagem”, “Fisioterapia”, “Serviço Social”, “Nutrição”, “Fonoaudiologia”, “Terapia Ocupacional” e “Psicologia”.
- Uma análise preliminar das questões do FGPC-OMS permitiu à equipe identificar, como prioritários para a obtenção de dados, os tipos referentes a “Medicina”, “Enfermagem” e “Fisioterapia”.
- Um aspecto importante para a análise dos dados não estrutura é o registro de uma mesma informação como, por exemplo, dos sinais vitais por todos os profissionais. Esse registro faz parte da avaliação que cada profissional faz do paciente.

1.2 Interpretação e mapeamento para o FGPC-OMS:

1.2.1 Organização dos Módulos do FGPC-OMS versus Registros do Prontuário:

O FGPC-OMS foi apresentado ao grupo, sendo ressaltada sua estruturação em três módulos. O primeiro é responsável pela coleta dos dados do paciente na sua admissão; o segundo para registro do acompanhamento do paciente, enquanto ele estiver internado; e o terceiro para consolidar as informações do atendimento e registrar o desfecho do paciente.

O FGPC-OMS é organizado em perguntas e respostas padronizadas ou de texto livre, que devem ser obtidas a partir dos dados registrados no prontuário eletrônico AGHU.

Numa avaliação preliminar, considerando os registros de anamnese, evolução, medicamentos, controles e exames foram estabelecidos os seguintes tratamentos:

- Módulo de Admissão: Para cada atendimento do paciente haverá um módulo de admissão, registrando: a data de internação; os resultados de exames referentes a primeira coleta após a admissão; os medicamentos administrados no dia da admissão; e o primeiro controle registrado como sintomas da admissão;
- Módulo de Desfecho: Para cada atendimento com alta/óbito/transferência haverá um módulo de desfecho, contemplando as informações da conclusão do atendimento.
- Módulo de Acompanhamento: Para **cada dia de internação** do paciente, após a data de sua admissão, haverá **apenas um** módulo de acompanhamento, onde serão registradas, de forma consolidada, as informações referentes ao dia em questão.

1.2.2 Organização proposta para o Módulo de Acompanhamento:

Para a criação do módulo de acompanhamento, considerando um para cada dia da internação, foram estabelecidos os tratamentos apresentados a seguir:

- Para o registro dos sinais vitais serão empregadas as informações registradas em controle às 12:00 hs. Caso não exista o registro, será empregado o primeiro lançamento após às 12:00hs;
- Para registro de resultados de exames: serão empregadas as informações referentes à primeira coleta de exame do dia;
- Para medicamentos: serão relacionados todos os medicamentos fornecidos ao paciente ao longo do dia.

1.2.3 Propostas para Mapeamento das Informações

A seguir estão relacionadas propostas preliminares para obtenção dos dados para cada módulo do FGPC-OMS. Essas propostas deverão ser evoluídas nas próximas reuniões do Grupo de Trabalho.

Uma atenção especial deverá ser dada para termos como “NEGA” ou “SEM” nos registros de anamnese e evolução, evitando o registro de respostas incorretas para o FGPC-OMS. Exemplos de registros identificados no prontuário: “nega HAS ou DM”; “sem dispnéia FR 20irpm, ventilando sob cateter nasal 2l/min SAT 100%”.

Para os agrupamentos de questões que apresentarem a questão do tipo “Outros:” como, por exemplo, “Comorbidades” e “Sintomas na Admissão”, será necessário um levantamento junto aos profissionais de saúde que atuam com os pacientes com COVID-19.

Para a questão de “Sinais Vitais” referente à AVPU (Alert, Verbal, Pain, Unresponsive) será analisada a possibilidade de mapeamento com as siglas LOTE (Lucido Orientado no Tempo e Espaço), BEG (Bom Estado Geral) e REG (Regular Estado Geral), associado com outros termos na evolução.

a) Módulo de Admissão:

Emprego de dados estruturados referentes a controles, exames e medicamentos e possibilidade de análise dos dados não estruturados, ou seja, dos registros textuais dos profissionais de saúde, na anamnese ou na primeira evolução para obtenção das informações relacionadas a seguir:

Questões do Formulário	Proposta de Obtenção
<p><b>1a. CLINICAL INCLUSION CRITERIA</b></p> <p>One or more of these during this illness   A history of self-reported feverishness or measured fever of <math>\geq 38^{\circ}\text{C}</math>. <input type="checkbox"/>Yes <input type="checkbox"/>No</p> <p>  Cough <input type="checkbox"/>Yes <input type="checkbox"/>No</p> <p>  Dyspnoea (shortness of breath) OR Tachypnoea* <input type="checkbox"/>Yes <input type="checkbox"/>No</p> <p>  Clinical suspicion despite not meeting criteria above <input type="checkbox"/>Yes <input type="checkbox"/>No</p> <p><small>* Respiratory rate <math>\geq 50</math> breaths/min for <math>&lt; 1</math> year; <math>\geq 40</math> for 1-4 years; <math>\geq 30</math> for 5-12 years; <math>\geq 20</math> for <math>\geq 13</math> years</small></p>	<p>Para o item 1.a</p> <p>Buscar na anamnese termos como dispnéia, tosse, febre</p>
<p><b>1b. DEMOGRAPHICS</b></p> <p>Sex at birth <input type="checkbox"/>Male <input type="checkbox"/>Female <input type="checkbox"/>Not specified Date of birth [ ]</p> <p>If date of birth is unknown, record: Age [ ] [ ] [ ] years OR [ ] [ ] months OR [ ] [ ] days</p> <p>Health care worker? <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown Laboratory worker? <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</p> <p>Pregnant? <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown <input type="checkbox"/>N/A If yes: Gestational weeks assessment [ ] [ ] weeks</p> <p><small>If currently pregnant or recently pregnant (delivery within 21 days of symptom onset), complete Pregnancy CRF</small></p>	<p>Para o item 1.b</p> <p>Informações sobre sexo, idade e data de nascimento serão obtidos do paciente.</p> <p>Informações referentes às questões “Profissional de Saúde”, “Profissional de laboratório” e “Gravidez” deverão ser analisadas na anamnese</p>

<p><b>1c. DATE OF ONSET AND ADMISSION VITAL SIGNS (reportable data at presentation/admission)</b></p> <p>Symptom onset (date of first/earliest symptom) [ ]</p> <p>Admission date at this facility [ ]</p> <p>Temperature [ ] [ ] [ ] [ ] <math>^{\circ}\text{C}</math> Heart rate [ ] [ ] [ ] [ ] beats/min</p> <p>Respiratory rate [ ] [ ] breaths/min</p> <p>BP [ ] [ ] [ ] [ ] [ ] [ ] (systolic) [ ] [ ] [ ] [ ] (diastolic) mmHg Severe dehydration <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</p> <p>Sternal capillary refill time <math>&gt; 2</math> seconds <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</p> <p>Oxygen saturation: [ ] [ ] [ ] % on <input type="checkbox"/>Room air <input type="checkbox"/>Oxygen therapy <input type="checkbox"/>Unknown A V P U (circle one)</p> <p>Glasgow Coma Score (GCS/15) [ ] [ ] [ ] Malnutrition <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</p> <p>Mid-upper arm circumference [ ] [ ] [ ] cm Height [ ] [ ] [ ] [ ] cm Weight [ ] [ ] [ ] [ ] kg</p>	<p>Para o item 1.c</p> <p>- Verificar melhor forma de se obter na anamnese as informações referentes à:</p> <p>a) Data de início dos sintomas</p> <p>b) Desidratação severa, Desnutrição</p> <p>c) AVPU</p> <p>d) Escala de Glasgow</p>
<p><b>1d. CO-MORBIDITIES (existing at admission) (Link = Unknown)</b></p> <p>Chronic cardiac disease <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Link Diabetes <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Link</p> <p>(not hypertension) Hypertension <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Link Current smoking <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Link</p> <p>Chronic pulmonary disease <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Link Tuberculosis (active) <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Link</p> <p>Asthma <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Link Tuberculosis (previous) <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Link</p> <p>Chronic kidney disease <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Link Asplenia <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Link</p> <p>Chronic liver disease <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Link Malignant neoplasm <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Link</p> <p>Chronic neurological disorder <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Link Other <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Link</p> <p>If yes, specify _____</p> <p>HIV <input type="checkbox"/>Yes (on ART) <input type="checkbox"/>Yes (not on ART) <input type="checkbox"/>No <input type="checkbox"/>Unknown ART regimen: _____</p>	<p>Para o item 1.d</p> <p>→ Necessário verificar o siglário empregado pelo HUGG e identificar as comorbidades estabelecidas no Formulário, bem como outros termos que possam ser empregados para registro das informações.</p> <p>Exemplo de siglas identificadas em análise preliminar:</p> <p>HAS – Hipertensão</p> <p>DM – Diabetes</p> <p>HIV</p>
<p><b>1e. PRE-ADMISSION AND CHRONIC MEDICATION Were any of the following taken within 14 days of admission?</b></p> <p>Angiotensin converting enzyme inhibitors (ACE inhibitors)? <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</p> <p>Angiotensin II receptor blockers (ARBs)? <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</p> <p>Non-steroidal anti-inflammatory (NSAID)? <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</p> <p>Antiviral? <input type="checkbox"/>Chloroquine/hydroxychloroquine <input type="checkbox"/>Azithromycin <input type="checkbox"/>oseltamivir/fluoravir <input type="checkbox"/>Other: _____</p>	<p>Para o item 1.e</p> <p>- Levantamento junto aos profissionais responsáveis pelo processo de admissão dos principais medicamentos que são informados pelos pacientes. Os termos apresentados serão inseridos em mecanismos de busca.</p>



1f. SIGNS AND SYMPTOMS ON ADMISSION (Link = Unknown)			
History of fever	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk	Lower chest indrawing	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk
Cough	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk	Headache	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk
with sputum production	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk	Altered consciousness/confusion	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk
with haemoptysis	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk	Seizures	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk
Sore throat	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk	Abdominal pain	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk
Runny nose	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk	Vomiting/nausea	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk
Wheezing	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk	Diarrhoea	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk
Chest pain	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk	Conjunctivitis	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk
Muscle aches	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk	Skin rash	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk
Joint pain (arthritis)	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk	Skin ulcers	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk
Fatigue/malaise	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk	Lymphadenopathy	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk
Loss of taste	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk	Inability to walk	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk
Loss of smell	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk	Bleeding	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk
Shortness of breath	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk	If bleeding, specify site(s):	
Stroke: ischaemic stroke	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk		
Stroke: intracerebral haemorrhage	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk		
Other:	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unk		
If yes, specify:			

Para o item 1.f

Similar ao item 1.d – verificar possibilidades adotadas pelos profissionais do HUGG, ou seja, siglas empregadas ou termos mais comuns para o registro das informações no prontuário.

1g. MEDICATION On the day of admission, did the patient receive any of the following	
Oral/enteric fluids?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
Intravenous fluids?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
Antiviral?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown <b>If yes:</b> <input type="checkbox"/> Ribavirin <input type="checkbox"/> Lopinavir/Ritonavir <input type="checkbox"/> Neuraminidase inhibitor
	<input type="checkbox"/> Interferon alpha <input type="checkbox"/> Interferon beta <b>Other, specify:</b> _____
Corticosteroid?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown <b>If yes, route:</b> <input type="checkbox"/> Oral <input type="checkbox"/> Intravenous <input type="checkbox"/> Inhaled <b>If yes, please provide agent and maximum daily dose:</b> _____
Antibiotic?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown <b>If yes, specify:</b> _____
Antifungal agent?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
Antimalarial agent?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown <b>If yes, specify:</b> _____
Experimental agent?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown <b>If yes, specify:</b> _____
Non-steroidal anti-inflammatory (NSAID)	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
Angiotensin converting enzyme inhibitors (ACE inhibitors)	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
Angiotensin II receptor blockers (ARBs)	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
Systemic anticoagulation	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown

Para o item 1.g

Como mencionado anteriormente, serão analisados todos os medicamentos administrados ao paciente no dia da sua admissão, correlacionando-os ao tipo estabelecido no formulário.

1h. SUPPORTIVE CARE On the day of admission, did the patient receive any of the following:	
ICU or high dependency unit admission?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
Oxygen therapy?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown <b>If yes, complete all below</b>
	<b>O<sub>2</sub> flow:</b> <input type="checkbox"/> 0-5 L/min <input type="checkbox"/> 6-10 L/min <input type="checkbox"/> 11-15 L/min <input type="checkbox"/> > 15 L/min <input type="checkbox"/> Unknown
	<b>Source of oxygen:</b> <input type="checkbox"/> Pipet <input type="checkbox"/> Cylinder <input type="checkbox"/> Concentrator <input type="checkbox"/> Unknown
	<b>Interface:</b> <input type="checkbox"/> Nasal prongs <input type="checkbox"/> HF nasal cannula <input type="checkbox"/> Mask <input type="checkbox"/> Mask with reservoir <input type="checkbox"/> CPAP/BIPAP/MSV <input type="checkbox"/> Unknown
Non-invasive ventilation? (e.g. BIPAP/CPAP)	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
Invasive ventilation (any)?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
<b>If yes, what were the following values closest to 08:00:</b>	
	PEEP (cm H <sub>2</sub> O) _____, FiO <sub>2</sub> (%) _____, Plateau pressure (cm H <sub>2</sub> O) _____, PaCO <sub>2</sub> _____, PaO <sub>2</sub> _____
Extracorporeal (ECMO) support?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
Prone position?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
Inotropes/vasopressors?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown

Para o item 1.h

- Será verificado se o paciente foi admitido para UTI ou Enfermaria-COVID;

- Quanto à terapia de oxigênio (evolução) – verificar a melhor forma de sua obtenção.

- Fonte de oxigênio quando necessária será sempre “PIPED”
- Quanto à interface: Nasal prongs → “cateter tipo óculos” ou O2 sob cateter nasal; HF nasal cannula → “cateter de alto fluxo” ou CNAF; Mask → “Macronebulização” ou MNBZ; Mask with reservoir → MNBZ com reservatório ou “máscara de venturi”;

Quanto a ventilação invasiva:

- Se em controles houver registro de PEEP e FiO<sub>2</sub> → marcar como sim
- Plateau pressure, PaCO<sub>2</sub> e PaO<sub>2</sub> estão em evolução – verificar melhor forma para obtenção

Demais questões precisam ser investigadas quanto a forma de apresentação na evolução

11. LABORATORY RESULTS ON ADMISSION (record units if different from those listed)					
Parameter	Value*	Units	Parameter	Value*	Units
Haemoglobin	<input type="checkbox"/> g/L <input type="checkbox"/> g/dL		Creatinine	<input type="checkbox"/> mg/L <input type="checkbox"/> µmol/L	
WBC count	<input type="checkbox"/> /mm <sup>3</sup> <input type="checkbox"/> /mm <sup>3</sup> ( $\times 10^9/L$ )		Sodium	<input type="checkbox"/> mEq/L <input type="checkbox"/> mmol/L	
Haematocrit	<input type="checkbox"/> %		Potassium	<input type="checkbox"/> mEq/L <input type="checkbox"/> mmol/L	
Platelets	<input type="checkbox"/> /mm <sup>3</sup> ( $\times 10^9/L$ )		Procalcitonin	<input type="checkbox"/> ng/mL <input type="checkbox"/> µg/L	
APTT/APTR	<input type="checkbox"/> seconds		CRP	<input type="checkbox"/> mg/L	
PT (seconds)	<input type="checkbox"/> seconds		LDH	<input type="checkbox"/> IU/L	
INR			Creatine kinase	<input type="checkbox"/> IU/L <input type="checkbox"/> U/KAT/L	
ALT/SGPT	<input type="checkbox"/> IU/L		Troponin	<input type="checkbox"/> ng/mL <input type="checkbox"/> µg/L	
AST/SGOT	<input type="checkbox"/> IU/L		ESR	<input type="checkbox"/> mm/hour	
Total bilirubin	<input type="checkbox"/> mg/L <input type="checkbox"/> µmol/L		D-dimer	<input type="checkbox"/> ng/mL <input type="checkbox"/> µg/L	
Urea (BUN)	<input type="checkbox"/> g/L <input type="checkbox"/> mg/dL <input type="checkbox"/> mmol/L		Ferritin	<input type="checkbox"/> ng/mL <input type="checkbox"/> µg/L	
Lactate	<input type="checkbox"/> mg/dL <input type="checkbox"/> mmol/L		IL-6	<input type="checkbox"/> pg/mL	

Para o item 1.i  
Como mencionado anteriormente, serão registrados os resultados laboratoriais referentes ao primeiro exame realizado pelo paciente no HUGG.

b) Módulo de Acompanhamento:

Emprego de dados estruturados referentes a controles, exames e medicamentos e possibilidade de análise dos dados não estruturados, ou seja, dos registros textuais dos profissionais de saúde, na evolução para obtenção das informações relacionadas a seguir:

Questões do Formulário	Proposta de Obtenção
<p><b>2a. VITAL SIGNS</b> (record most abnormal value between 00:00 to 24:00)</p> <p>Temperature <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> °C Heart rate <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> beats per min Respiratory rate <input type="checkbox"/> <input type="checkbox"/> breaths/min            BP <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> (systolic) <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> (diastolic)/mmHg Severe dehydration <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown            Sternal capillary refill time &gt; 2 seconds <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown A V P U (circle one)            Oxygen saturation on <input type="checkbox"/> Room air <input type="checkbox"/> Oxygen therapy <input type="checkbox"/> Unknown GCS/15 <input type="checkbox"/> <input type="checkbox"/></p>	<p>Para o item 2.a Como apresentado anteriormente, os sinais vitais serão obtidos do registro de controles realizados às 12:00hs ou o primeiro realizado após esse horário. - Verificar como obter as informações referentes à Desidratação severa, AVPU, tempo de recarga capilar e escala de Glasgow (CGS/15)</p>

<p><b>2b. DAILY CLINICAL FEATURES</b> (Yes/No/Unknown)</p> <p>Cough <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown            and sputum production            Sore throat <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown            Chest pain <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown            Shortness of breath <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown            Loss of smell <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown            Loss of taste <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown</p> <p>Confusion <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown            Squares <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown            Vomiting/nausea <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown            Diarrhoea <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown            Conjunctivitis <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown            Myalgia <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown            Other, specify _____ <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown</p>	<p>Para o item 2.b Similar aos itens 1.d e 1.f – verificar possibilidades adotadas pelos profissionais do HUGG, ou seja, siglas empregadas ou termos mais comuns para o registro das informações no prontuário (evolução).</p>
---	--

2c. LABORATORY RESULTS (record units if different from those listed)					
Parameter	Value*	Units	Parameter	Value*	Units
Haemoglobin	<input type="checkbox"/> g/L <input type="checkbox"/> g/dL		Creatinine	<input type="checkbox"/> mg/L <input type="checkbox"/> µmol/L	
WBC count	<input type="checkbox"/> /mm <sup>3</sup> <input type="checkbox"/> /mm <sup>3</sup> ( $\times 10^9/L$ )		Sodium	<input type="checkbox"/> mEq/L <input type="checkbox"/> mmol/L	
Haematocrit	<input type="checkbox"/> %		Potassium	<input type="checkbox"/> mEq/L <input type="checkbox"/> mmol/L	
Platelets	<input type="checkbox"/> /mm <sup>3</sup> ( $\times 10^9/L$ )		Procalcitonin	<input type="checkbox"/> ng/mL <input type="checkbox"/> µg/L	
APTT/APTR	<input type="checkbox"/> seconds		CRP	<input type="checkbox"/> mg/L	
PT (seconds)	<input type="checkbox"/> seconds		LDH	<input type="checkbox"/> IU/L	
INR			Creatine kinase	<input type="checkbox"/> IU/L <input type="checkbox"/> U/KAT/L	
ALT/SGPT	<input type="checkbox"/> IU/L		Troponin	<input type="checkbox"/> ng/mL <input type="checkbox"/> µg/L	
AST/SGOT	<input type="checkbox"/> IU/L		ESR	<input type="checkbox"/> mm/hour	
Total bilirubin	<input type="checkbox"/> mg/L <input type="checkbox"/> µmol/L		D-dimer	<input type="checkbox"/> ng/mL <input type="checkbox"/> µg/L	
Urea (BUN)	<input type="checkbox"/> g/L <input type="checkbox"/> mg/dL <input type="checkbox"/> mmol/L		Ferritin	<input type="checkbox"/> ng/mL <input type="checkbox"/> µg/L	
Lactate	<input type="checkbox"/> mg/dL <input type="checkbox"/> mmol/L		IL-6	<input type="checkbox"/> pg/mL	

Para o item 2.c  
Serão registrados os resultados laboratoriais do paciente, com base na primeira coleta da data do acompanhamento.

<p><b>2d. MEDICATION</b> At any time during this 24-hour hospital day, did the patient receive:</p> <p>Oral/ogastric fluids? <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown Intravenous fluids? <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown            Antiviral? <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown If yes: <input type="checkbox"/> Ribavirin <input type="checkbox"/> Lopinavir/Ritonavir <input type="checkbox"/> Neuraminidase inhibitor  <input type="checkbox"/> Interferon alpha <input type="checkbox"/> Interferon beta (Other, specify _____)            Corticosteroid? <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown If yes, route: <input type="checkbox"/> Oral <input type="checkbox"/> Intravenous <input type="checkbox"/> Inhaled            If yes, please provide agent and maximum daily dose: _____            Antibiotic? <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown If yes, specify _____            Antifungal agent? <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown            Antimalarial agent? <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown If yes, specify _____            Experimental agent? <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown If yes, specify _____            Non-steroidal anti-inflammatory (NSAID) <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown            Angiotensin converting enzyme inhibitors (ACE inhibitors) <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown            Angiotensin II receptor blockers (ARBs) <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown            Systemic anticoagulation <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown</p>	<p>Para o item 2.d Serão analisados e mapeados todos os medicamentos administrados para o paciente na data do acompanhamento.</p>
---	---

<p><b>2e. SUPPORTIVE CARE:</b> At any time during this 24-hour hospital day, did the patient receive:  <b>ICU or high dependency unit admission?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown  Date of ICU/HDU admission [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] <input type="checkbox"/>Unknown  ICU/HDU discharge date [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] <input type="checkbox"/>Not discharged yet <input type="checkbox"/>Unknown  <b>Oxygen therapy?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown <b>If yes, complete all below:</b>  <b>O<sub>2</sub> flow:</b> <input type="checkbox"/>1-5 L/min <input type="checkbox"/>6-10 L/min <input type="checkbox"/>11-15 L/min <input type="checkbox"/> &gt; 15 L/min <input type="checkbox"/>Unknown  <b>Source of oxygen:</b> <input type="checkbox"/>Piped <input type="checkbox"/>Cylinder <input type="checkbox"/>Concentrator <input type="checkbox"/>Unknown  <b>Interface:</b> <input type="checkbox"/>Nasal prongs <input type="checkbox"/>HF nasal cannula <input type="checkbox"/>Mask <input type="checkbox"/>Mask with reservoir <input type="checkbox"/>CPAP/NIV mask <input type="checkbox"/>Unknown  <b>Non-invasive ventilation?</b> (e.g. BiPAP, CPAP) <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown  <b>Invasive ventilation (any)?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown  <b>If yes, what were the following values closest to 08:00:</b>  PEEP (cm H<sub>2</sub>O) _____ PFC (%) _____ Plateau pressure (cm H<sub>2</sub>O) _____ PaCO<sub>2</sub> _____ PaO<sub>2</sub> _____  <b>Extracorporeal (ECMO) support?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown  <b>Prone position?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown  <b>Inotropes/vasopressors?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown  <b>Renal replacement therapy (RRT) or dialysis?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</p>	<p>Para o item 2.e  Serão aplicadas as mesmas análises realizadas no item 1.h, considerando a data do acompanhamento.</p>
---	---

c) Módulo de Desfecho:

Emprego de dados estruturados referentes a controles, exames e medicamentos; consolidação das informações obtidas ao longo da internação; e registro de informações não estruturadas da evolução:

Questões do Formulário	Proposta de Obtenção
<p><b>3a. DIAGNOSTIC/PATHOGEN TESTING</b>  <b>Chest X-ray/CT performed?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown <b>If yes, infiltrates present?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown  <b>Was pathogen testing done during this illness episode?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown <b>If yes, complete all below:</b>  <b>Influenza virus:</b> <input type="checkbox"/>Positive <input type="checkbox"/>Negative <input type="checkbox"/>Not done <b>If positive, type:</b> _____  <b>Coronavirus:</b> <input type="checkbox"/>Positive <input type="checkbox"/>Negative <input type="checkbox"/>Not done <b>If positive:</b> <input type="checkbox"/>MERS-CoV <input type="checkbox"/>SARS-CoV-2 <input type="checkbox"/>Other _____  <b>Other respiratory pathogen:</b> <input type="checkbox"/>Positive <input type="checkbox"/>Negative <input type="checkbox"/>Not done <b>If positive, specify:</b> _____  <b>Viral haemorrhagic fever:</b> <input type="checkbox"/>Positive <input type="checkbox"/>Negative <input type="checkbox"/>Not done <b>If positive, specify virus:</b> _____  <b>Other pathogen of public health interest detected:</b> <b>If yes, specify:</b> _____  <b>Falciparum malaria:</b> <input type="checkbox"/>Positive <input type="checkbox"/>Negative <input type="checkbox"/>Not done  <b>Non-falciparum malaria:</b> <input type="checkbox"/>Positive <input type="checkbox"/>Negative <input type="checkbox"/>Not done  <b>HIV:</b> <input type="checkbox"/>Positive <input type="checkbox"/>Negative <input type="checkbox"/>Not done</p>	<p>Para o item 3.a  - Para identificação de Raio-x ou TC → verificar evolução do paciente.  Buscar padrão de registro quanto à presença ou não do vírus  - Quanto aos testes de patógenos: considerar para os pacientes na enfermaria e UTI COVID-19 – Positivo para Coronavírus do tipo SARS-CoV-2.  → Registrar demais testes como “Não Realizado”</p>

<p><b>3b. COMPLICATIONS:</b> At any time during hospitalization, did the patient experience:</p> <table border="1"> <tr><td>Shock</td><td><input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</td><td>Bacteremia</td><td><input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</td></tr> <tr><td>Sepsis</td><td><input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</td><td>Shedding</td><td><input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</td></tr> <tr><td>Meningitis/encephalitis</td><td><input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</td><td>Endocarditis</td><td><input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</td></tr> <tr><td>Anaemia</td><td><input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</td><td>Myocarditis/pericarditis</td><td><input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</td></tr> <tr><td>Cardiac arrhythmia</td><td><input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</td><td>Acute renal injury</td><td><input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</td></tr> <tr><td>Cardiac arrest</td><td><input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</td><td>Pancreatitis</td><td><input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</td></tr> <tr><td>Pneumonia</td><td><input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</td><td>Liver dysfunction</td><td><input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</td></tr> <tr><td>Irritocollitis</td><td><input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</td><td>Cardiomyopathy</td><td><input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</td></tr> <tr><td>Acute respiratory distress syndrome (ARDS)</td><td><input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</td><td>Other</td><td><input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</td></tr> <tr><td>Stroke: ischaemic stroke</td><td><input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</td><td><b>If yes, specify:</b></td><td></td></tr> <tr><td>Stroke: intracerebral haemorrhage</td><td><input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</td><td></td><td></td></tr> </table>	Shock	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Bacteremia	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Sepsis	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Shedding	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Meningitis/encephalitis	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Endocarditis	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Anaemia	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Myocarditis/pericarditis	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Cardiac arrhythmia	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Acute renal injury	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Cardiac arrest	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Pancreatitis	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Pneumonia	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Liver dysfunction	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Irritocollitis	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Cardiomyopathy	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Acute respiratory distress syndrome (ARDS)	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Other	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Stroke: ischaemic stroke	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	<b>If yes, specify:</b>		Stroke: intracerebral haemorrhage	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown			<p>Para o item 3.b  De forma similar aos itens 1.d, 1.f e 2.b  → verificar possibilidades adotadas pelos profissionais do HUGG, ou seja, siglas empregadas ou termos mais comuns para o registro das informações no prontuário (evolução).  Verificar com profissionais da UTI complicações comuns que não estão lançadas na relação da OMS.</p>
Shock	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Bacteremia	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown																																										
Sepsis	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Shedding	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown																																										
Meningitis/encephalitis	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Endocarditis	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown																																										
Anaemia	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Myocarditis/pericarditis	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown																																										
Cardiac arrhythmia	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Acute renal injury	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown																																										
Cardiac arrest	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Pancreatitis	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown																																										
Pneumonia	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Liver dysfunction	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown																																										
Irritocollitis	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Cardiomyopathy	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown																																										
Acute respiratory distress syndrome (ARDS)	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Other	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown																																										
Stroke: ischaemic stroke	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	<b>If yes, specify:</b>																																											
Stroke: intracerebral haemorrhage	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown																																												

<p><b>3c. MEDICATION:</b> While hospitalized or at discharge, were any of the following administered:</p> <p><b>Oral/enteric fluids?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown <b>Intravenous fluids?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</p> <p><b>Antiviral?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown <b>If yes:</b> <input type="checkbox"/>Ribavirin <input type="checkbox"/>Oseltamivir/Tamiflu <input type="checkbox"/>Neuraminidase inhibitor <input type="checkbox"/>Interferon alpha <input type="checkbox"/>Interferon beta <input type="checkbox"/>Other, specify: _____</p> <p><b>Corticosteroid?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown <b>If yes, route:</b> <input type="checkbox"/>Oral <input type="checkbox"/>Intravenous <input type="checkbox"/>Inhaled <b>If yes, specify agent and maximum daily dose:</b> _____</p> <p><b>Antibiotic?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown <b>If yes, specify:</b> _____</p> <p><b>Antifungal agent?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown <b>If yes, specify:</b> _____</p> <p><b>Antimalarial agent?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown <b>If yes, specify:</b> _____</p> <p><b>Experimental agent?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown <b>If yes, specify:</b> _____</p> <p><b>Non-steroidal anti-inflammatory (NSAID)</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown <b>If yes, specify:</b> _____</p> <p><b>Systemic anticoagulation</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</p>	<p>Para o item 3.c  Serão considerados todos os medicamentos administrados ao paciente durante sua hospitalização.</p>
---	--

<p><b>3d. SUPPORTIVE CARE:</b> At any time during hospitalization, did the patient receive/undergo:</p> <p><b>ICU or high dependency unit admission?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown <b>If yes, total duration:</b> _____ days  <b>Date of ICU admission:</b> [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] <input type="checkbox"/>N/A  <b>Date of ICU discharge:</b> [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] <input type="checkbox"/>In ICU at outcome <input type="checkbox"/>N/A  <b>Oxygen therapy?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown <b>If yes, complete all:</b> <b>Total duration:</b> _____ days  <b>O<sub>2</sub> flow:</b> <input type="checkbox"/>1-5 L/min <input type="checkbox"/>6-10 L/min <input type="checkbox"/>11-15 L/min <input type="checkbox"/> &gt; 15 L/min  <b>Source of oxygen:</b> <input type="checkbox"/>Piped <input type="checkbox"/>Cylinder <input type="checkbox"/>Concentrator  <b>Interface:</b> <input type="checkbox"/>Nasal prongs <input type="checkbox"/>HF nasal cannula <input type="checkbox"/>Mask <input type="checkbox"/>Mask with reservoir <input type="checkbox"/>CPAP/NIV mask  <b>Non-invasive ventilation?</b> (e.g. BiPAP, CPAP) <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown <b>If yes, total duration:</b> _____ days  <b>Invasive ventilation (any)?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown <b>If yes, total duration:</b> _____ days  <b>Extracorporeal (ECMO) support?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown <b>If yes, total duration:</b> _____ days  <b>Prone position?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown <b>If yes, total duration:</b> _____ days  <b>Inotropes/vasopressors?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown <b>If yes, total duration:</b> _____ days  <b>Renal replacement therapy (RRT) or dialysis?</b> <input type="checkbox"/>Yes <input type="checkbox"/>No <input type="checkbox"/>Unknown</p>	<p>Para o item 3.d  - Data de admissão e saída da UTI será obtida pela movimentação entre UTI/Enfermaria ou UTI/Desfecho.  - Demais informações serão coletadas de forma similar aos itens 1.h e 2.e, considerando a duração como o total de dias onde cada item foi identificado.</p>
--	--

<p><b>3e. OUTCOME</b></p> <p><b>Outcome:</b> <input type="checkbox"/>Discharged alive <input type="checkbox"/>Hospitalized <input type="checkbox"/>Transfer to other facility <input type="checkbox"/>Death <input type="checkbox"/>Palliative discharge <input type="checkbox"/>Unknown</p> <p><b>Outcome date:</b> [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] <input type="checkbox"/>Unknown</p> <p><b>If discharged alive, ability to self-care at discharge versus before illness:</b> <input type="checkbox"/>Same as before illness <input type="checkbox"/>Worse <input type="checkbox"/>Better <input type="checkbox"/>Unknown</p>	<p>Para o item 3.e  - Registrar a data do fim do atendimento  - Verificar como deverá ser tratada a saída (Outcome) e estado do paciente</p>
--	--

## ANEXO B – TABELA DE MEDICAMENTOS

Cod Medicamento	Medicamento	Associação
13544	ACIDO ACETILSALICILICO COMP	AINE e Antiagregante Plaquetário
12750	ACIDO FOLICO COMP	Antianêmico
400810	ACIDO POLIACRILICO (CARBOMER) GELEIA OFTALMICA, 10 G	Lubrificante oftálmico
13650	ACIDO VALPROICO CAP	Antiepiléptico
277810	AGUA DESTILADA INJ - SIST FECHADO (MPP)	Diluyente
13676	ALBUMINA HUMANA 20% SOL. INJ. - FR 50 ML	Hemoderivado
400817	ALFAEPOETINA (ERITROPOETINA) SOL. INJ. 1 ML	Fator estimulante de Hemácias
13684	aloPURinol COMP	Antigotoso: inibidor da biossíntese de ácido úrico.
13692	AMICACINA - SOL. INJ. - AMP 2 ML	Antibiótico
13757	amioDARONA COMP	Antiarrítmico
400824	AMOXICILINA + CLAV DE POTASSIO PO LIOFILO P/ INJ	Antibiótico
13811	AMPICILINA - PO LIOFILO P/ INJ	Antibiótico
400837	ANLODIPINO COMP	Antihipertensivo: Bloqueador de Canais de Cálcio
293617	ATENOLOL COMP	Antihipertensivo: Bloqueador Seletivo de Receptores beta 1 adrenérgicos
136107	ATENOLOL COMP	Antihipertensivo: Bloqueador Seletivo de Receptores beta1 adrenérgicos
269929	AZITromicina COMP	Antibiótico
157678	AZITromicina PO P/ SUSP. ORAL	Antibiótico
181315	BISACODIL COMP	Laxante
288949	BROMOPRIDA INJ. - AMP 2 ML	Antiemético
13978	CAPTOPRIL COMP	Antihipertensivo: Inibidor da ECA
401118	carBAMazepina COMP	Antiepiléptico
400858	CARBONATO DE CALCIO COMP	Repositor Eletrolítico (Minerais)
280632	carVEDilol COMP	Antihipertensivo: Bloqueador não Seletivo de Receptores beta e alfa adrenérgicos
16969	cefTRIAxona LIOFILO P/ INJ.	Antibiótico
402571	CISATRACURIO - AMP 10ML	Bloqueador Neuromuscular
293627	CISATRACURIO - AMP 5ML (MPP)	Bloqueador Neuromuscular
224669	CLARITROMICINA LIOFILO P/ INJ.	Antibiótico
14184	CLONAZEPAM COMP	Ansiolítico e Anticonvulsivante (benzodiazepínico)
13404	CLONAZEPAM SOL. ORAL - FR 20 ML	Ansiolítico e Anticonvulsivante (benzodiazepínico)
271412	cloNIDina AMP (MPP)	Antihipertensivo: Agonista Receptores Alfa-2 adrenérgicos
248843	cloNIDina COMP	Antihipertensivo: Agonista Receptores Alfa-2 adrenérgicos
273025	CLOPIDOGREL COMP (MPP)	Antiagregante Plaquetário
400925	CLORETO DE POTASSIO 10% - AMP (MPP)	Solução
400927	CLORETO DE POTASSIO 6% XAROPE	Repositor Eletrolítico (Mineirais)
223190	CLOROQUINA	Agente antimalárico
290234	DARUNAVIR COMP	Inibidor de Protease (tratamento antirretroviral)
16390	DEXAmetasona COMP	Anti-inflamatório Esteroidal
14575	DEXAmetasona SOL. INJ 4 MG/ML - AMP 2,5 ML	Anti-inflamatório Esteroidal
14583	DEXCLORFENIRAMINA COMP	Antihistamínico
401032	DEXMEDETOMIDINA SOL. INJ. AMP 2 ML (MPP)	Ansiolítico e Sedativo
401034	DEXTROCETAMINA AMP 2 ML (MPP)	Anestésico
14664	DIAZEPAM 10 MG - AMP 2 ML (MPP)	Ansiolítico e Anticonvulsivante (benzodiazepínico)
14826	DIPIRONA AMP 2 ML	AINE
14800	DIPIRONA COMP	AINE
401971	DOLUTEGRAVIR COMP	Inibidor de Integrase (tratamento antirretroviral)
142077	ENALAPRIL COMP	Antihipertensivo: Inibidor da ECA
242845	ENOXAPARINA SODICA SOL. INJ. 0,2 ML (MPP)	Antitrombótico/Anticoagulante
242853	ENOXAPARINA SODICA SOL. INJ. 0,4 ML (MPP)	Antitrombótico/Anticoagulante

269466	ENOXAPARINA SODICA SOL. INJ. 0,6 ML (MPP)	Antitrombótico/Anticoagulante
266299	ENOXAPARINA SODICA SOL. INJ. 0,8 ML (MPP)	Antitrombótico/Anticoagulante
14923	ESPIRONOLACTONA COMP	Diurético
15040	FENITOINA SOL. INJ. - AMP 5 ML	Anticonvulsivante / Antiepilético
401086	fentaNILA - INJ. - F/A 10 ML (MPP)	Analgésico Opióide
289140	FITOMENADIONA (VITAMINA K1) - SOL. INJ. 1 ML IM/SC	Antídoto de Anticoagulante
151041	FLUCONAZOL 200MG SOL. INJ. - BOLSA 100 ML	Antifúngico
401098	FLUTICASONA 60 DOSES, AEROSSOL ORAL	Anti-inflamatório Esteroidal
401103	FUROSEMIDA SOL. INJ. 2 ML	Diurético
15237	GLIBENCLAMIDA COMP (MPP)	Antidiabético: sulfoniluréias
400873	GLICONATO DE CALCIO 10% SOL. INJ	Repositor Eletrolítico
400874	GLICOSE 10% SOL. INJ. SISTEMA FECHADO	Diluyente para Solução Parenteral
15334	GLICOSE 50% SOL. INJ. - AMP (MPP)	Solução Hiperglicemiante
400881	haloPERIdol SOL ORAL - FRASCO 20 ML	Antipsicótico
15407	haloPERIdol SOL. INJ. INTRAMUSCULAR - AMP 1 ML	Antipsicótico
16349	HEPARINA AMP 0,25 ML SC (MPP)	Anticoagulante
202878	HidrALAZINA COMP	Vasodilatador
15458	HidrALAZINA SOL. INJ. - AMP 1 ML	Vasodilatador
281190	HidroCLOROTiazida COMP	Diurético
15474	HIDROCORTISONA PO PARA SOLUCAO	Anti-inflamatório Esteroidal
259748	HIDROXICLOROQUINA COMP	Antimalárico
400886	HIOSCINA (ESCOPOLAMINA) INJ. - AMP 1 ML	Antiespasmódico
264202	INSULINA HUMANA NPH 10ML (MPP)	Antidiabético
400903	INSULINA HUMANA REGULAR 10 ML (MPP)	Antidiabético
265217	IVERMECTINA COMP	Antiparasitário
148261	LACTULOSE XPE - FR 120 ML	Laxante
401119	LEVOflaxacino 500 MG - SOL. INJ. - BOLSA 100 ML	Antibiótico
289019	LevoTIROXina COMP	Hormônio Sintético Tireoídiano
289017	LevoTIROXina COMP	Hormônio Sintético Tireoídiano
275319	LOSARTANA COMP	Antihipertensivo: Bloqueador de Receptor de Angiotensina II
980200	MEDICAMENTO NAO CADASTRADO	N/A
213780	MEROPENEM PO LIOFILO P/ INJ	Antibiótico
218480	MEROPENEM PO LIOFILO P/ INJ.	Antibiótico
247278	METADONA COMP (MPP)	Analgésico Opióide
264164	METFORMINA COMP	Antidiabético: biguanidas
136123	METILDOPA COMP	Antihipertensivo: Agonista Receptores Alfa-2 adrenérgicos
401120	METILPREDNISOLONA PO PARA SOL. INJ. + DILUENTE EXCLUSIVO	Anti-inflamatório Esteroidal
16420	METILPREDNISOLONA PO PARA SOL. INJ. + DILUENTE EXCLUSIVO	Anti-inflamatório Esteroidal
16101	METOCLOPRAMIDA SOL. INJ. - AMP 2 ML	Antiemético
16128	METRONIDAZOL COMP	Antibiótico e Antiparasitário
400957	MIDAZOLAM SOL. INJ. - AMP 10 ML (MPP)	Hipnótico e Sedativo
16179	MORFINA INJ - AMP 1 ML (MPP)	Analgésico Opióide
400967	MORFINA INJ. - AMP 2ML (CENTRO CIRURGICO) (MPP)	Analgésico Opióide
400974	NISTATINA CREME	Antifúngico
400976	NitroPRUSSIATO SOL. INJ 2 ML OU PO P/ SOL 50 MG (MPP)	Vasodilatador / Antianginoso
400977	NOREpinefrina (NORADRENALINA) SOL. INJ. - AMP 4 ML (MPP)	Vasopressor/Vasoconstritor
17442	OLEO MINERAL LIQUIDO (PETROLATO)	Laxante
136085	OMEPRAZOL CAP	Inibidor de bomba de prótons
400984	OMEPRAZOL PO P/ INJ. + DILUENTE EXCLUSIVO	Inibidor de bomba de prótons
149454	ONDANSETRONA 4 MG/2 ML INJETAVEL	Antiemético
141097	ONDANSETRONA 8 MG/4 ML INJETAVEL	Antiemético
271851	OSELTAMIVIR CAP	Antiviral
248665	PIPERACILINA 4,0 G + TAZOBACTAM 0,5 G INJ.	Antibiótico
272925	POLIMIXINA B PO PARA SOLUCAO P/ INJ.	Antibiótico
17680	PredniSONA COMP	Anti-inflamatório Esteroidal
17671	PredniSONA COMP	Anti-inflamatório Esteroidal
17752	PROMETAZINA COMP	Antihistamínico
401045	PROPOFOL 1%, EMULSAO INJETAVEL FR 20 ML (MPP)	Anestésico e Hipnótico
270226	QUEtiapina COMP	Antipsicótico
401044	RINGER COM LACTATO, SOL. INJ. SIST FECHADO	Repositor Eletrolítico
273166	RITONAVIR COMP	Antiviral
135950	SALBUTAMOL AEROSSOL 200 DOSES	Broncodilatador: Agonista Receptores Beta-2 Adrenérgicos
271617	SINVASTATINA COMP	Hipolipemiantes (estatina)
400982	SORO FISIOLÓGICO (CLORETO DE SODIO 0,9%) INJ. SIST. FECH	Diluyente / Repositor Hídrico
400985	SORO FISIOLÓGICO (CLORETO DE SODIO 0,9%) SOL. INJ. SIST FECH	Diluyente / Repositor Hídrico
400980	SORO FISIOLÓGICO (CLORETO DE SODIO 0,9%) SOL. INJ. SIST FECH	Diluyente / Repositor Hídrico
400983	SORO FISIOLÓGICO (CLORETO DE SODIO 0,9%) SOL. INJ. SIST FECH	Diluyente / Repositor Hídrico
400877	SORO GLICOSADO (GLICOSE 5%) SOL. INJ. SIST FECHADO	Diluyente para Solução Parenteral
201553	SORO GLICOSADO (GLICOSE 5%) SOL. INJ. SIST FECHADO	Diluyente para Solução Parenteral
15270	SORO GLICOSADO (GLICOSE 5%) SOL. INJ. SIST FECHADO	Diluyente para Solução Parenteral
18406	SULFAMETOXAZOL + TRIMETOP 40 MG/ML+8 MG/ML SUSP. ORAL	Antibiótico
12858	SULFAMETOXAZOL + TRIMETOP 400 MG+80 MG COMP	Antibiótico
108987	SULFAMETOXAZOL + TRIMETOP 80 MG/ML+16 MG/ ML SOL. INJ.	Antibiótico
109223	SULFATO DE MAGNESIO 10% SOL. INJ.- AMP (MPP)	Repositor Eletrolítico
18015	SULFATO DE MAGNESIO 50% SOL. INJ - AMP (MPP)	Repositor Eletrolítico

279816	TIGECICLINA	Antibiótico
290232	TRAMADOL 100 MG - AMP 2 ML (MPP)	Analésico Opióide
401072	VANCOMICINA PO P/ INJECÃO	Antibiótico
235369	VARFARINA COMP (MPP)	Anticoagulante
136239	VASOpessina 1 ML AMP	Vasopressor/Vasoconstritor
401079	VITAMINAS DO COMPLEXO B COMP	Vitaminas

**ANEXO C – DICIONÁRIO DE DADOS DO HUGG**

<b>Dados Demográficos</b>	
<b>Coluna</b>	<b>Descrição</b>
Prontuário	Número do Prontuário
Data de Nascimento	DD/MM/AAA
Idade	em anos
Sexo	F/M
Bairro	Bairro do paciente

<b>Internação</b>	
<b>Coluna</b>	<b>Descrição</b>
Atendimento	Número de Atendimento do referido prontuário para tratamento do COVID-19
Prontuário	Número do Prontuário
Unidade Funcional	Nome do local onde o paciente
Data de Internação	Data do início da internação referente ao atendimento específico
Data de Saída	Data da saída do hospitalar referente ao atendimento específico
Data de Óbito	Data referente ao óbito, caso

	tenha ocorrido
CID	CID associado ao motivo da internação do paciente Pode haver mais de um
Descricao CID	Descrição do CID na internação do paciente

<b>Anamnese</b>	
<b>Coluna</b>	<b>Descrição</b>
Prontuário	Número do Prontuário
Atendimento	Número de Atendimento do referido prontuário para tratamento do COVID-19
Data de criação	Data em que a anamnese foi realizada (DD/MM/AAA hh:mm:ss)
Data de alteração	Data em que a anamnese foi realizada



	(DD/MM/AAA hh:mm:ss)
Qualificação Profissional	Qualificação do Profissional de saúde responsável pelos dados da anamnese (ex: Médico Enfermagem, Fisioterapia, Serviço Social, Nutrição)
Anamnese	Registro de dados não estruturados avaliados pelos profissionais de saúde

<b>Evolução</b>	
<b>Coluna</b>	<b>Descrição</b>
Atendimento	Número de Atendimento do referido prontuário para tratamento do COVID-19
Prontuário	Número do Prontuário
Unidade Funcional	Nome do local onde o paciente está internado
Data da internação	Data em que o paciente foi internado (DD/MM/AAAA hh:mm:ss)

Data de evolução	Data em que a evolução do paciente foi verificada (DD/MM/AAA hh:mm:ss)
Data de alteração	Data de alteração de uma evolução
Qualificação Profissional	Qualificação do Profissional de saúde responsável pelos dados da evolução
Evolução	Registro de dados não estruturados avaliados pelos profissionais de saúde

<b>Medicamentos</b>	
<b>Coluna</b>	<b>Descrição</b>
Atendimento	Número de Atendimento do referido prontuário para tratamento do COVID-19
Prontuário	Número do Prontuário
Data da prescrição	Data em que o medicamento foi prescrito (DD/MM/AAAA hh:mm:ss)

Medicamento	Código do medicamento ministrado ao paciente durante o atendimento
Descrição Medicamento	Descrição do medicamento ministrado
Dosagem	Dose prescrita do medicamento
Tipo Medicamento	Indica se o medicamento é um antibiótico, antifúngico, anti-inflamatório, dentre outros

<b>Controles</b>	
<b>Coluna</b>	<b>Descrição</b>
Atendimento	Número de Atendimento do referido prontuário para tratamento do COVID-19
Prontuário	Número do Prontuário
Data Coleta	Data em que o medicamento foi prescrito (DD/MM/AAAA hh:mm:ss)
Tipo Controle	Identifica o tipo de controle medido - Sinal vital, Monitorização Hemodinâmica, Monitorização Ventilatória, Monitorização Neurológica, Monitorização Metabólica

Controle	Controle realizado
Sigla Controle	Sigla do controle realizado Exemplo: FC - Frequência Cardíaca
Medição	Dose prescrita do medicamento
Unidade de Medida	Unidade de medida correspondente ao controle. Para Temperatura - °C