# Arquitetura de Dados e Infraestrutura de Conhecimento para Sistemas Multi-Agentes em Culinária Computacional

## 1. Introdução à Culinária Computacional e Sistemas Multi-Agentes

A intersecção entre a ciência de dados, a inteligência artificial e a gastronomia deu origem ao campo emergente da Culinária Computacional. Para o desenvolvimento de uma equipe de agentes autônomos capazes de gerar receitas, raciocinar sobre substituições de ingredientes e entender a viabilidade física e química de um prato, a simples alimentação de modelos de linguagem com textos de receitas é insuficiente. O desafio central reside na construção de uma representação digital fidedigna dos ingredientes — um "Gêmeo Digital" (Digital Twin) do alimento — que capture não apenas sua identidade semântica, mas também suas propriedades visuais, nutricionais, físicas e relacionais. Este relatório técnico detalha exaustivamente as fontes de dados, ontologias e estratégias de arquitetura necessárias para fundar tal sistema, atendendo à necessidade crítica de um banco de imagens e informações granulares para cada ingrediente.

### 1.1 O Desafio da Multimodalidade na Geração de Receitas

A geração de receitas por agentes de IA transcende a mera concatenação probabilística de palavras. Um agente competente deve operar em um espaço multimodal. Quando um usuário solicita uma receita, ou quando um agente de inventário visualiza o interior de uma geladeira, o sistema deve ser capaz de ancorar (grounding) o conceito abstrato de "tomate" em dados concretos. Isso implica que o banco de dados subjacente deve fornecer suporte para diferentes modalidades de raciocínio. O raciocínio visual exige imagens de alta fidelidade para reconhecimento e segmentação; o raciocínio nutricional exige dados laboratoriais precisos sobre macro e micronutrientes; e o raciocínio culinário exige conhecimento sobre sabor, textura e compatibilidade química. A ausência de qualquer um desses pilares resulta em "alucinações culinárias", onde receitas podem ser semanticamente plausíveis, mas fisicamente impossíveis ou nutricionalmente desastrosas.

A complexidade aumenta quando consideramos a natureza dinâmica dos alimentos. Um ingrediente não é uma entidade estática; ele possui estados (cru, cozido, frito, processado) e variações taxonômicas (maçã Fuji vs. maçã Gala) que alteram drasticamente suas propriedades. A maioria dos datasets genéricos falha em capturar essas nuances, tratando "batata" como uma entidade única, ignorando que uma batata *russet* é ideal para assar, enquanto uma batata *yukon gold* é preferível para purês devido ao teor de amido. Para uma equipe de agentes especializada, essa granularidade não é opcional; é o diferencial entre um gerador de texto genérico e uma ferramenta gastronômica profissional.

### 1.2 Requisitos para uma Equipe de Agentes Colaborativos

Ao projetar uma equipe de agentes, assume-se uma divisão de responsabilidades onde cada agente requer uma "visão" diferente do mesmo dado. A arquitetura de dados deve, portanto, servir múltiplas *views* sobre o mesmo objeto "Ingrediente".

O **Agente Nutricionista** requer precisão decimal em dados quantitativos. Ele não se beneficia de imagens artísticas, mas depende de tabelas de composição química validadas, como as fornecidas pelo USDA, para calcular a densidade calórica e o perfil de vitaminas.1 Para este agente, o banco de dados é uma matriz numérica.

O **Agente Visual ou de Inventário** opera no espectro oposto. Sua função é reconhecer ingredientes no mundo real ou gerar representações visuais do prato final. Ele necessita de datasets de imagens robustos, variados em iluminação, ângulo e estado de processamento, como os encontrados no Nutrition5k ou Fruits-360.3 Para este agente, o dado é um tensor de pixels.

O **Agente Chef ou Criativo** atua na camada semântica e ontológica. Ele precisa saber que "leite de amêndoa" é um substituto funcional para "leite de vaca" em receitas veganas, mas que se comporta de maneira diferente sob fervura. Isso requer o uso de ontologias estruturadas como FoodOn e grafos de conhecimento como FoodKG, que codificam relações hierárquicas e funcionais entre ingredientes.5

Este relatório analisará como integrar essas fontes díspares — governamentais, acadêmicas, crowdsourced e comerciais — em um pipeline coeso que satisfaça as demandas de todos os agentes do sistema.

## 2. A Base da Verdade Nutricional: USDA FoodData Central

Para qualquer aplicação séria de culinária computacional, a precisão dos dados nutricionais é inegociável. Datasets de imagens encontrados em repositórios abertos frequentemente contêm metadados nutricionais estimados ou incorretos. Portanto, a estratégia recomendada é dissociar a fonte da verdade nutricional da fonte visual, utilizando o **USDA FoodData Central (FDC)** como a espinha dorsal bioquímica do sistema.

### 2.1 Análise Estrutural do USDA FoodData Central

O USDA FoodData Central é um sistema integrado de dados de composição de alimentos e nutrientes hospedado pelo Departamento de Agricultura dos Estados Unidos.2 Diferente de tabelas simples, o FDC é composto por cinco tipos de dados distintos, cada um servindo a um propósito específico na modelagem de ingredientes.

A **Foundation Foods** representa o nível mais granular e moderno de dados. O foco aqui é na variabilidade. Em vez de fornecer apenas uma média, este dataset inclui metadados sobre o número de amostras, localização, data de coleta e métodos analíticos utilizados.2 Para um agente de pesquisa, isso permite entender que o teor de açúcar em uma fruta pode variar dependendo da safra ou região.

O **SR Legacy (Standard Reference)**, embora não seja mais atualizado desde 2018, continua sendo a referência padrão para commodities e ingredientes básicos não processados.2 É aqui que se encontra a "base da verdade" para itens como farinha de trigo, ovos crus, cortes de carne in natura e vegetais comuns. Para a geração de receitas a partir do zero (scratch cooking), o SR Legacy é a fonte primária de metadados.

O **FNDDS (Food and Nutrient Database for Dietary Studies)** é projetado para converter alimentos consumidos em gramas de nutrientes. É particularmente útil para pratos compostos e processados comuns na dieta americana, facilitando o trabalho de agentes que precisam estimar o valor nutricional de refeições completas.2

A base de **Branded Foods** é fornecida pela indústria alimentícia e contém dados de rótulos de produtos empacotados.2 Com mais de 300.000 itens, é essencial para agentes que lidam com ingredientes específicos de marcas (ex: um molho de tomate específico ou um tipo de cereal), embora os dados sejam menos detalhados em micronutrientes do que os da Foundation Foods.

### 2.2 Acesso Programático e Implementação Técnica

Para integrar o USDA FDC ao fluxo de trabalho dos agentes, o uso da API oficial ou dos arquivos de dados brutos é necessário. A API é RESTful, gratuita e requer uma chave de API (API Key) obtida através de registro simples.7

A estrutura de resposta da API e dos arquivos JSON baixáveis é hierárquica. Um alimento é identificado por um fdcId (Food Data Central ID). A consulta a um fdcId retorna um objeto contendo:

* **Descrição:** Nome do alimento (ex: "Broccoli, raw").
* **Food Nutrients:** Uma lista de nutrientes, onde cada nutriente possui seu próprio ID, nome, quantidade e unidade de medida (ex: ID 1003 para Proteína, valor em gramas).3
* **Food Portions:** Dados cruciais para converter medidas caseiras ("1 xícara") em gramas, permitindo que os agentes "entendam" as quantidades descritas em receitas textuais.

Existem bibliotecas Python desenvolvidas pela comunidade, como usda-fdc e python-usda, que encapsulam as chamadas de API e facilitam a manipulação dos dados.10 Estas ferramentas permitem a criação de scripts de *scraping* ético e *caching* local, essenciais para evitar latência e respeitar os limites de taxa da API (rate limits).12 O uso de scripts como os encontrados no repositório usda-fdc-data permite baixar, limpar e estruturar os dados em bancos locais (como PostgreSQL ou MongoDB), criando um repositório de alta performance para consulta pelos agentes.13

### 2.3 Limitações Visuais e Estratégia de Mitigação

Uma limitação crítica do USDA FDC para o objetivo do usuário é a **ausência quase total de imagens**.14 O banco de dados é focado em química, não em visão computacional. Embora existam menções a um motor de busca de fotos do USDA, a integração programática entre o fdcId e imagens de alta qualidade é inexistente ou falha para a maioria dos itens.14

Isso impõe um requisito arquitetural: o sistema deve operar com uma chave de ligação (join key) que conecte o registro nutricional do USDA a um registro visual externo. O nome do alimento (em inglês) frequentemente serve como essa chave, mas requer normalização (ex: mapear "Apple, raw, with skin" do USDA para a classe "Apple Red Delicious" de um dataset de imagens). Essa "Estratégia de Fusão" é o cerne da construção do banco de dados multimodal proposto.

## 3. O Componente Visual: Datasets de Imagens e Grounding

Para dotar os agentes de capacidade visual, é necessário recorrer a datasets acadêmicos e repositórios de imagens especializados. A análise revelou três fontes primárias que, quando combinadas, cobrem o espectro de ingredientes crus a pratos montados e produtos industrializados.

### 3.1 Nutrition5k: Profundidade e Massa para Regressão Nutricional

O dataset **Nutrition5k** 3 representa o estado da arte em dados para culinária computacional. Desenvolvido por pesquisadores do Google, ele aborda diretamente a relação entre a aparência visual de um alimento e seus macronutrientes.

#### 3.1.1 Metodologia de Coleta e Riqueza de Dados

Diferente de datasets raspados da web, o Nutrition5k foi coletado em ambiente controlado (cafeterias). O processo envolveu a captura incremental de vídeos de pratos sendo montados. Para cada ingrediente adicionado ao prato, o sistema registrou o peso exato (massa) e capturou imagens RGB e de profundidade (RGB-D) de múltiplos ângulos.3

Este método gera metadados de valor inestimável para agentes de IA:

* **Massa por Ingrediente:** O dataset informa não apenas que há "brócolis" no prato, mas que há exatamente "45 gramas de brócolis". Isso permite treinar agentes para estimar *volume* e *densidade* a partir de pixels, uma habilidade crítica para monitoramento de dieta e geração de imagens realistas.3
* **Visão Multiangular:** Os vídeos rotativos fornecem vistas de 360 graus, permitindo que agentes de visão aprendam a reconhecer ingredientes oclusos ou parcialmente visíveis.3
* **Dados de Profundidade:** A inclusão de canais de profundidade (Depth) permite diferenciar entre uma fatia fina de queijo e um bloco denso, resolvendo ambiguidades que confundem modelos puramente 2D.16

#### 3.1.2 Estrutura dos Metadados

Os dados são fornecidos em arquivos CSV estruturados. O arquivo dish\_metadata contém os totais do prato (calorias, gordura, proteína, carboidratos). O arquivo ingredient\_metadata detalha cada componente, ligando o ID do ingrediente (ingr\_ID) aos valores nutricionais por grama derivados do USDA.3 Isso fecha o ciclo entre a imagem e a base da verdade nutricional, tornando o Nutrition5k um recurso de "pedra de roseta" para traduzir pixels em nutrientes.

### 3.2 Fruits-360: Inventário Granular de Hortifruti

Enquanto o Nutrition5k foca em pratos compostos, o **Fruits-360** 4 oferece uma granularidade excepcional para ingredientes crus, especificamente frutas, vegetais e nozes.

#### 3.2.1 Variedade e Captura

O dataset contém mais de 90.000 imagens cobrindo 131 classes distintas.20 A característica definidora é a captura em plataforma giratória, resultando em imagens do mesmo objeto sob todos os ângulos de rotação.22 Isso fornece robustez aos agentes de reconhecimento visual, permitindo identificar uma fruta independentemente de sua orientação espacial.

A granularidade das classes é um diferencial significativo. O dataset não possui apenas uma classe "Maçã", mas distingue entre "Maçã Braeburn", "Maçã Crimson Snow", "Maçã Golden", "Maçã Granny Smith", entre outras.22 Para um Agente Chef, essa distinção é vital: ele pode recomendar visualmente uma maçã *Granny Smith* para uma torta (devido à acidez e firmeza) em oposição a uma *Red Delicious*.

#### 3.2.2 Integração de Metadados

O Fruits-360 é primariamente um dataset de imagens. Os metadados nutricionais não são nativos ou são limitados em alguns forks do projeto.24 Para utilizar este dataset na equipe de agentes, é necessário criar uma tabela de mapeamento que vincule cada nome de pasta (classe) ao fdcId correspondente no USDA. Por exemplo, mapear a classe Banana para o item USDA correspondente a bananas cruas. Isso enriquece a imagem estática com todo o poder analítico do USDA.

### 3.3 Open Food Facts: O Universo dos Produtos Processados

Para ingredientes que vêm em embalagens (farinhas, molhos, laticínios, enlatados), o **Open Food Facts (OFF)** 25 é o recurso definitivo. Trata-se de um banco de dados colaborativo (crowdsourced) massivo, cobrindo milhões de produtos globalmente.

#### 3.3.1 Escala e Acesso a Imagens

O OFF disponibiliza imagens cruas das embalagens, tabelas nutricionais e listas de ingredientes. Estas imagens são hospedadas na AWS e acessíveis via URLs previsíveis baseadas no código de barras do produto.29 Isso permite que os agentes acessem não apenas a "foto do produto", mas também a "foto da verdade" (a tabela nutricional impressa), permitindo validação cruzada via OCR (Reconhecimento Óptico de Caracteres).

#### 3.3.2 Metadados Avançados: NOVA e Eco-Score

Além dos nutrientes, o OFF fornece classificações calculadas que são essenciais para agentes focados em saúde e sustentabilidade:

* **Classificação NOVA:** Categoriza os alimentos pelo grau de processamento industrial, do Grupo 1 (não processado) ao Grupo 4 (ultraprocessado).32 Isso permite que um Agente de Saúde filtre ou penalize ingredientes ultraprocessados na geração de receitas.
* **Eco-Score:** Avalia o impacto ambiental do produto.
* **Tags de Alérgenos e Dietas:** O sistema normaliza ingredientes para detectar alérgenos (glúten, leite) e adequação a dietas (vegano, kosher).34

#### 3.3.3 Desafios de Qualidade de Dados

Por ser alimentado por usuários, o OFF sofre de inconsistências e ruído.36 Um produto pode ter dados nutricionais incompletos ou fotos borradas. Para uso profissional, recomenda-se utilizar o dump de dados completo (exportação diária) e aplicar filtros de qualidade (ex: descartar produtos sem nutriscore\_grade ou com dados nutricionais flagrantemente errados) antes de ingeri-los no banco de dados da equipe de agentes.

## 4. Semântica e Raciocínio: Ontologias e Grafos de Conhecimento

Dados isolados (imagens e números) não conferem "entendimento" aos agentes. Para que um agente saiba que pode substituir *manteiga* por *óleo de coco* em uma receita vegana, ele precisa de uma estrutura de conhecimento semântico. Ontologias e grafos de conhecimento fornecem essa camada lógica.

### 4.1 FoodOn: A Ontologia Padrão da Indústria

**FoodOn** 5 é uma ontologia construída para harmonizar a terminologia alimentar global. Ela fornece uma hierarquia de classes rigorosa baseada na *Basic Formal Ontology* (BFO), garantindo interoperabilidade semântica.

#### 4.1.1 Estrutura Hierárquica e Relações

No FoodOn, cada ingrediente é um nó em uma árvore complexa. "Leite de vaca" não é apenas uma string; é uma subclasse de "leite mamífero", que é subclasse de "secreção corporal", e também está ligado a "produto lácteo".

* **Relações Funcionais:** A ontologia define propriedades como derives from (deriva de), has ingredient (tem ingrediente), part of (parte de) e has quality (tem qualidade - cor, textura).40
* **Aplicação para Agentes:** Se o Agente Criativo precisa gerar uma receita "sem glúten", ele não precisa de uma lista negra manual de milhares de produtos. Ele pode consultar a ontologia FoodOn para identificar todos os grãos descendentes da classe *Triticeae* (trigo, centeio, cevada) e excluí-los automaticamente, bem como todos os produtos que possuem a relação derives from com esses grãos. Isso confere robustez e escalabilidade ao raciocínio do agente.

### 4.2 FoodKG: O Grafo de Conhecimento Integrado

O **FoodKG** 6 é um recurso que integra receitas, nutrientes e ontologias em um grafo de conhecimento unificado. Ele representa um avanço significativo para sistemas de recomendação alimentar.

#### 4.2.1 Mapeamento WordNet e ImageNet

Uma das características mais poderosas do FoodKG para a construção de bancos de imagens é o seu mapeamento com o **WordNet** e o **ImageNet**.42 O FoodKG utiliza os *offsets* (identificadores numéricos de sinônimos) do WordNet para ligar conceitos de ingredientes a *synsets* (conjuntos de sinônimos) no ImageNet.

* **Solução para Imagens Ausentes:** Como o ImageNet é um dos maiores repositórios de imagens categorizadas do mundo, essa ligação permite que os agentes recuperem milhares de imagens para um ingrediente genérico (ex: "morango") através de seu *synset*, mesmo que o banco de dados local não tenha uma foto específica. Isso é feito através de triplas RDF onde o predicado #ImgURLs aponta para os recursos visuais externos.44

#### 4.2.2 Consultas SPARQL e Raciocínio Complexo

O FoodKG permite o uso da linguagem de consulta SPARQL para realizar perguntas complexas que bancos SQL tradicionais lutariam para responder.46

* **Exemplo de Query:** "Encontre receitas que contenham 'carne bovina' (ou seus derivados), tenham menos de 500 calorias, não contenham 'amendoim' e sejam classificadas como 'culinária mexicana'".
* **Embeddings Vetoriais (AGROVEC):** O FoodKG utiliza modelos de *graph embedding* (como AGROVEC) para calcular similaridade semântica.44 Isso permite que os agentes encontrem substitutos não óbvios baseados na proximidade vetorial no grafo, sugerindo ingredientes que ocupam papéis funcionais ou de sabor semelhantes.

## 5. Enriquecimento Dinâmico: APIs Comerciais e Gratuitas

Para complementar os datasets estáticos e fornecer funcionalidades em tempo real (como preços, disponibilidade de mercado ou processamento de linguagem natural avançado), a integração de APIs externas é estratégica.

### 5.1 Spoonacular API: O "Canivete Suíço" da Culinária

A API **Spoonacular** 49 é amplamente reconhecida como uma das mais completas para engenharia de alimentos.

* **Busca e Informações de Ingredientes:** O endpoint GET /food/ingredients/{id}/information fornece uma riqueza de dados para cada ingrediente, incluindo imagens padronizadas, custos estimados, informações nutricionais e até o corredor provável de supermercado onde o item pode ser encontrado.49
* **Busca Baseada em Inventário:** O endpoint "Search Recipes by Ingredients" é fundamental para o Agente de Inventário. Ele permite enviar uma lista de itens disponíveis (ex: o que foi reconhecido visualmente na geladeira) e receber receitas que maximizam o uso desses itens e minimizam compras adicionais.51
* **Pricing e Limites:** A API oferece um plano gratuito (Free Tier) que permite testar e desenvolver, geralmente limitado a 150 requisições/dia ou um sistema de pontos (50 pontos/dia). Planos pagos escalam conforme o volume, oferecendo SLAs de disponibilidade.52 Para uma equipe de agentes em produção, o caching de respostas é essencial para controlar custos.

### 5.2 Edamam API: Processamento de Linguagem Natural (NLP)

A **Edamam** 55 destaca-se na interpretação de texto não estruturado.

* **Nutrition Analysis API:** Este serviço permite enviar o texto de uma receita inteira ou uma lista de ingredientes em linguagem natural (ex: "1 xícara de arroz e dois ovos cozidos") e receber uma análise nutricional estruturada e detalhada, incluindo dietas (Low-Carb, High-Protein) e alérgenos.55 Isso é vital para agentes que precisam "ler" receitas da web ou entender comandos de voz de usuários.
* **Database:** A Edamam possui acesso a um banco de dados de quase 900.000 alimentos genéricos e produtos empacotados.55
* **Modelos de Negócio:** O plano gratuito é robusto para desenvolvedores, mas o uso comercial exige atribuição clara ("Powered by Edamam") e está sujeito a limites de taxa. Planos empresariais oferecem acesso ilimitado e suporte dedicado.56

### 5.3 TheMealDB: Imagens e Simplicidade

O **TheMealDB** 58 é um projeto open-source de alta qualidade focado na simplicidade e no visual.

* **Thumbnails de Ingredientes:** Uma das maiores utilidades para a construção de interfaces visuais para os agentes é o acesso direto a imagens de ingredientes em alta qualidade através de URLs previsíveis (ex: https://www.themealdb.com/images/ingredients/{Ingrediente}.png).58
* **Acesso:** Totalmente gratuito para uso básico, com uma versão para apoiadores (Patreon) que oferece recursos adicionais.58 É uma excelente fonte para popular rapidamente a camada visual do banco de dados sem custos de API ou complexidade de scraping.

## 6. Arquitetura de Integração e Pipeline ETL

A construção efetiva do banco de dados requer uma arquitetura de "Data Lakehouse" que ingira, processe e sirva esses dados heterogêneos.

### 6.1 Estrutura do Pipeline de Dados (ETL)

O processo de Extração, Transformação e Carga (ETL) deve seguir uma lógica de camadas para garantir a integridade dos dados.

**Tabela 1: Estratégia de Camadas de Dados**

| **Camada** | **Fonte de Dados Principal** | **Responsabilidade** | **Formato de Armazenamento** |
| --- | --- | --- | --- |
| **Bronze (Raw)** | USDA CSVs, OFF Dumps, Nutrition5k CSVs | Armazenamento imutável dos dados brutos baixados. | Object Storage (S3/MinIO) - JSON/CSV |
| **Prata (Refined)** | Scripts Python (usda-fdc, openfoodfacts) | Limpeza, normalização de unidades (tudo para gramas/ml), validação de tipos. | Banco SQL (PostgreSQL) ou NoSQL (MongoDB) |
| **Ouro (Aggregated)** | Fusão (Join) de Fontes | Criação do "Objeto Ingrediente Universal". Mapeamento de IDs (USDA ID <-> OFF Barcode <-> FoodOn URI). | Grafo (Neo4j) ou Document Store Indexado (Elasticsearch) |

#### 6.1.1 Processo de Fusão e Enriquecimento

O maior desafio técnico é a fusão. O USDA tem o ID 1109003 para "Cenoura". O Fruits-360 tem a pasta Carrot. O TheMealDB tem a imagem Carrot.png.

1. **Normalização de Nomes:** Utilizar técnicas de NLP (Lemmatization, Fuzzy Matching) para alinhar os nomes em inglês.
2. **Mapeamento Ontológico:** Utilizar o FoodKG para conectar o termo "Carrot" ao seu conceito no WordNet e, consequentemente, às suas propriedades no USDA e imagens no ImageNet.
3. **Preenchimento de Lacunas (Gap Filling):**
   * *Sem imagem?* Buscar no TheMealDB ou usar a URL do ImageNet via FoodKG.
   * *Sem nutrição?* Buscar o "pai" mais próximo na ontologia FoodOn (ex: se não houver dados para "Maçã Cripps Pink", usar a média de "Maçã Crua" do USDA).

### 6.2 Representação de Dados para Agentes (JSON Schema)

Para facilitar a comunicação entre os agentes, recomenda-se padronizar o objeto "Ingrediente" em um formato JSON rico.

JSON

{  
 "id\_sistema": "uuid-v4-unico",  
 "metadados\_identidade": {  
 "nome\_display": "Cenoura Crua",  
 "nomes\_alternativos":,  
 "usda\_fdc\_id": 1109003,  
 "foodon\_uri": "http://purl.obolibrary.org/obo/FOODON\_00001002",  
 "nova\_group": 1  
 },  
 "componente\_visual": {  
 "imagem\_principal": "s3://imagens/ingredientes/cenoura\_padrao.jpg",  
 "imagens\_variacoes": [  
 {"fonte": "Fruits-360", "url": "...", "estado": "cru"},  
 {"fonte": "Nutrition5k", "url": "...", "estado": "processado\_cubos"}  
 ],  
 "cor\_dominante\_hex": "#F57F17"  
 },  
 "componente\_nutricional": {  
 "base\_referencia": "100g",  
 "energia\_kcal": 41,  
 "macronutrientes": {  
 "proteina\_g": 0.93,  
 "carboidratos\_g": 9.58,  
 "lipidios\_g": 0.24  
 },  
 "micronutrientes\_chave":  
 },  
 "propriedades\_fisicas": {  
 "densidade\_estimada\_g\_ml": 0.64,  
 "estado\_padrao": "solido"  
 }  
}

### 6.3 Estratégias de Recuperação de Imagens e Scraping

Para o preenchimento inicial das imagens, scripts de *scraping* direcionado podem ser utilizados. Repositórios como ComprehensiveFoodDatabase no GitHub fornecem scripts prontos que varrem a web buscando imagens para itens do USDA, utilizando processamento paralelo (threaded processing) para acelerar a aquisição.62 Contudo, é vital respeitar os termos de serviço dos sites alvo e priorizar fontes com APIs abertas ou licenças permissivas (Creative Commons) para evitar passivos legais.

## 7. Conclusão e Roteiro Estratégico

A construção de um banco de imagens e informações de ingredientes para uma equipe de agentes de IA é um projeto de integração de dados em larga escala. Não existe uma única "bala de prata" que forneça todas as informações necessárias. A solução robusta reside na orquestração inteligente de fontes especializadas:

1. **USDA FoodData Central** como a autoridade inquestionável para a química e nutrição.
2. **Nutrition5k e Fruits-360** para prover a inteligência visual e volumétrica necessária para reconhecimento e estimativa.
3. **Open Food Facts** para cobrir o vasto universo de produtos industrializados com dados de saúde (NOVA/Eco-Score).
4. **FoodOn e FoodKG** para tecer a malha semântica que permite aos agentes raciocinar como chefs, entendendo substituições e hierarquias.
5. **APIs (Spoonacular/Edamam)** para enriquecimento dinâmico e tratamento de linguagem natural.

Ao implementar o pipeline ETL descrito e adotar o modelo de dados unificado, sua equipe de agentes terá acesso a um "cérebro culinário" que transcende listas de strings, operando com uma compreensão profunda, visual e científica dos alimentos. Este alicerce de dados é o que diferenciará um gerador de receitas genérico de uma ferramenta de IA gastronômica verdadeiramente capaz e inovadora.

#### Referências citadas

1. Food Nutrition Dataset - Kaggle, acessado em dezembro 14, 2025, <https://www.kaggle.com/datasets/utsavdey1410/food-nutrition-dataset>
2. USDA FoodData Central, acessado em dezembro 14, 2025, <https://fdc.nal.usda.gov/>
3. google-research-datasets/Nutrition5k: Detailed visual + ... - GitHub, acessado em dezembro 14, 2025, <https://github.com/google-research-datasets/Nutrition5k>
4. Fruits-360 dataset - Kaggle, acessado em dezembro 14, 2025, <https://www.kaggle.com/datasets/moltean/fruits>
5. FoodOntology/joint-food-ontology-wg - GitHub, acessado em dezembro 14, 2025, <https://github.com/FoodOntology/joint-food-ontology-wg>
6. FoodKG, acessado em dezembro 14, 2025, <https://foodkg.github.io/>
7. API Key Signup - USDA FoodData Central, acessado em dezembro 14, 2025, <https://fdc.nal.usda.gov/api-key-signup>
8. FoodData Central API Guide, acessado em dezembro 14, 2025, <https://fdc.nal.usda.gov/api-guide>
9. Downloadable Data - USDA FoodData Central, acessado em dezembro 14, 2025, <https://fdc.nal.usda.gov/download-datasets>
10. USDA Food Data Central (FDC) Python Client - Read the Docs, acessado em dezembro 14, 2025, <https://usda-fdc.readthedocs.io/en/stable/>
11. jdagdelen/python-usda: A fork of pygov, refreshed and focused only on USDA NDB API, acessado em dezembro 14, 2025, <https://github.com/jdagdelen/python-usda>
12. Rate limits for the REST API - GitHub Docs, acessado em dezembro 14, 2025, <https://docs.github.com/en/rest/using-the-rest-api/rate-limits-for-the-rest-api>
13. mkayeterry/usda-fdc-data: Repository for cleaning, processing, and wrangling USDA FoodData Central datasets. - GitHub, acessado em dezembro 14, 2025, <https://github.com/mkayeterry/usda-fdc-data>
14. How to get images for foods fetched on USDA FoodData Central API?, acessado em dezembro 14, 2025, <https://opendata.stackexchange.com/questions/21240/how-to-get-images-for-foods-fetched-on-usda-fooddata-central-api>
15. How to get images for foods fetched on USDA FoodData Central API? - Reddit, acessado em dezembro 14, 2025, <https://www.reddit.com/r/learnprogramming/comments/16s8gux/how_to_get_images_for_foods_fetched_on_usda/>
16. Nutrition5k: A Comprehensive Nutrition Dataset - Kaggle, acessado em dezembro 14, 2025, <https://www.kaggle.com/datasets/siddhantrout/nutrition5k-dataset>
17. Nutrition5k dataset - Kaggle, acessado em dezembro 14, 2025, <https://www.kaggle.com/datasets/gillesokhin/nutrition5k-dataset>
18. Predicting Nutritional Contents of Meals Using Deep Learning | by David Blanar | Medium, acessado em dezembro 14, 2025, <https://medium.com/@david.blanar/predicting-nutritional-contents-of-meals-using-deep-learning-c455e175b601>
19. Fruits-360 dataset - GitHub, acessado em dezembro 14, 2025, <https://github.com/fruits-360>
20. An Efficient Automatic Fruit-360 Image Identification and Recognition Using a Novel Modified Cascaded-ANFIS Algorithm - PMC - NIH, acessado em dezembro 14, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC9228155/>
21. Fruits-360 dataset - Kaggle, acessado em dezembro 14, 2025, <https://www.kaggle.com/datasets/moltean/fruits/versions/9>
22. Fruits 360 dataset - Mendeley Data, acessado em dezembro 14, 2025, <https://data.mendeley.com/datasets/rp73yg93n8/1>
23. Fruits-360 dataset - Mendeley Data, acessado em dezembro 14, 2025, <https://data.mendeley.com/datasets/rp73yg93n8/2>
24. Fruit Nutrition - Kaggle, acessado em dezembro 14, 2025, <https://www.kaggle.com/datasets/fathurrahmans/fruit-nutrition>
25. NutriGreen image dataset: a collection of annotated nutrition, organic, and vegan food products - PMC - NIH, acessado em dezembro 14, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC11002244/>
26. Data, API and SDKs - Open Food Facts, acessado em dezembro 14, 2025, <https://world.openfoodfacts.org/data>
27. Open Food Facts, acessado em dezembro 14, 2025, <https://world.openfoodfacts.org/>
28. Open Food Facts - world's largest open food database - EU-Citizen.Science, acessado em dezembro 14, 2025, <https://citizenscience.eu/project/430>
29. Open Food Facts Images - Registry of Open Data on AWS, acessado em dezembro 14, 2025, <https://registry.opendata.aws/openfoodfacts-images/>
30. How to download product images - Product Opener (Open Food Facts Server), acessado em dezembro 14, 2025, <https://openfoodfacts.github.io/openfoodfacts-server/api/how-to-download-images/>
31. Open Food Facts AWS images dataset, acessado em dezembro 14, 2025, <https://openfoodfacts.github.io/openfoodfacts-server/api/aws-images-dataset/>
32. Nova groups for food processing - Open Food Facts, acessado em dezembro 14, 2025, <https://world.openfoodfacts.org/nova>
33. Super Agers An Evidence Based Approach to Longevity Eric Topol - DOKUMEN.PUB, acessado em dezembro 14, 2025, <https://dokumen.pub/super-agers-an-evidence-based-approach-to-longevity-eric-topol.html>
34. Food Ingredients and Allergens - Kaggle, acessado em dezembro 14, 2025, <https://www.kaggle.com/datasets/uom190346a/food-ingredients-and-allergens>
35. Ingredients Extraction and Analysis - Open Food Facts wiki, acessado em dezembro 14, 2025, <https://wiki.openfoodfacts.org/Ingredients_Extraction_and_Analysis>
36. Beyond the Label: Decoding Food Processing with NOVA and Python | by Adel Basli, acessado em dezembro 14, 2025, <https://medium.com/@adelbasli/beyond-the-label-decoding-food-processing-with-nova-and-python-5ff12334d63e>
37. Introduction to Open Food Facts API documentation, acessado em dezembro 14, 2025, <https://openfoodfacts.github.io/openfoodfacts-server/api/>
38. FoodOntology/foodon: The core repository for the FOODON food ontology project. This holds the key classes of the ontology; larger files and the results of text-mining projects will be stored in other repos. - GitHub, acessado em dezembro 14, 2025, <https://github.com/FoodOntology/foodon>
39. FoodOn – FoodOn: A farm to fork ontology, acessado em dezembro 14, 2025, <https://foodon.org/>
40. FoodOn Relations, acessado em dezembro 14, 2025, <https://foodon.org/design/foodon-relations/>
41. FoodOn Structure, acessado em dezembro 14, 2025, <https://foodon.org/design/foodon-structure/>
42. FoodKG: A Tool to Enrich Knowledge Graphs Using Machine Learning Techniques - Frontiers, acessado em dezembro 14, 2025, <https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2020.00012/pdf>
43. FoodKG: A Semantics-Driven Knowledge Graph for Food Recommendation - Computer Science, acessado em dezembro 14, 2025, <http://www.cs.rpi.edu/~zaki/PaperDir/ISWC19.pdf>
44. FoodKG: A Tool to Enrich Knowledge Graphs Using Machine Learning Techniques - NIH, acessado em dezembro 14, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC7931944/>
45. FoodKG: A Tool to Enrich Knowledge Graphs Using Machine Learning Techniques, acessado em dezembro 14, 2025, <https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2020.00012/full>
46. Endpoint Access - FoodKG, acessado em dezembro 14, 2025, <https://foodkg.github.io/endpoint.html>
47. foodkg.github.io/whattomake.html at master · foodkg/foodkg.github.io · GitHub, acessado em dezembro 14, 2025, <https://github.com/foodkg/foodkg.github.io/blob/master/whattomake.html>
48. (PDF) FoodKG: A Tool to Enrich Knowledge Graphs Using Machine Learning Techniques, acessado em dezembro 14, 2025, <https://www.researchgate.net/publication/341008283_FoodKG_A_Tool_to_Enrich_Knowledge_Graphs_Using_Machine_Learning_Techniques>
49. Fetch Ingredients - Educative.io, acessado em dezembro 14, 2025, <https://www.educative.io/courses/plan-meals-and-recipes-with-the-spoonacular-api-in-javascript/fetch-ingredients>
50. Spoonacular - ingredient information API, acessado em dezembro 14, 2025, <https://www.worldindata.com/api/spoonacular-ingredient-information-api/>
51. Spoonacular API Documentation | API Specifications & Integration Guide - AllThingsDev, acessado em dezembro 14, 2025, <https://allthingsdev.co/apimarketplace/documentation/Spoonacular%20API/66750a5670009c3ab417c4ed>
52. Spoonacular API - APILayer, acessado em dezembro 14, 2025, <https://marketplace.apilayer.com/spoonacular-api>
53. [API] - What's for dinner tonight? Spoonacular has you covered! : r/tasker - Reddit, acessado em dezembro 14, 2025, <https://www.reddit.com/r/tasker/comments/h7pkdu/api_whats_for_dinner_tonight_spoonacular_has_you/>
54. Top 8 Nutrition APIs for Meal Planning 2024 - Eat Fresh Tech, acessado em dezembro 14, 2025, <https://www.eatfresh.tech/blog/top-8-nutrition-apis-for-meal-planning-2024>
55. Edamam - Food Database API, Nutrition API and Recipe API, acessado em dezembro 14, 2025, <https://www.edamam.com/>
56. Food and Grocery Database API - NLP and Image Recgnition - Edamam, acessado em dezembro 14, 2025, <https://developer.edamam.com/food-database-api>
57. Edamam APIs Terms of Service, acessado em dezembro 14, 2025, <https://www.edamam.com/terms/api/>
58. Free Recipe API - TheMealDB.com, acessado em dezembro 14, 2025, <https://www.themealdb.com/api.php>
59. Lard Ingredient API - TheMealDB.com, acessado em dezembro 14, 2025, <https://www.themealdb.com/ingredient/451-lard>
60. Migas Recipe API - TheMealDB.com, acessado em dezembro 14, 2025, <https://www.themealdb.com/meal/53086-migas-recipe>
61. TheMealDB.com - Free recipe API and database, acessado em dezembro 14, 2025, <https://www.themealdb.com/>
62. lxaw/ComprehensiveFoodDatabase: A comprehensive database of foods in the United States. - GitHub, acessado em dezembro 14, 2025, <https://github.com/lxaw/ComprehensiveFoodDatabase>