DATA ENGINEERING · MID-PROGRESS

GrocEyes

Shop what fits your body and beliefs

Personal shopper consultant with a restriction-aware product database and an incremental pipeline for food, snacks, cosmetics and household goods

Vinia Permata Kusuma - 20252530
Department of Industrial and Management Engineering, POSTECH

Shopping with a restriction is a daily audit

For some people, buying groceries is a burden: there is always something to check.

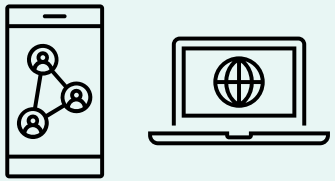
At the supermarket



With an allergy, a medical need or a belief, every package is read one by one to find the right product to buy.

GrocEyes → Scan once: a verdict and the ingredient that decided it

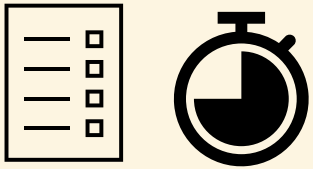
Online shopping



Even online it is hard: search and categories cannot filter for “eligible for me”, so the checking stays manual.

GrocEyes → A catalog of only the products your profile allows

No time to plan



Groceries cover our food and daily needs, but not everyone has time to plan what to buy or to cook at home.

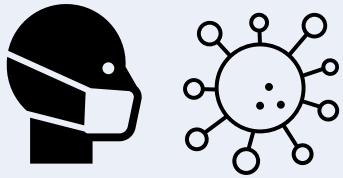
GrocEyes → Shopping list, home stock and meal plan that fit your profile

Different people, different restrictions

People restrict what they buy for three reasons: health, personal goals and belief. Unclear ingredients and warning labels make each of them hard, and suitable options can cost more.

1 Medical

health needs



- Food allergies (peanut, egg, milk...)
- Celiac, lactose intolerance
- Diabetes, hypertension, cancer
- Skin allergy
- Fragrance allergy (cosmetics)

Allergen labels confuse buyers; texture and nutrients matter too

GrocEyes → allergens and 'may contain' (level 3)

2 Diet & personal goals

gym, body, lifestyle



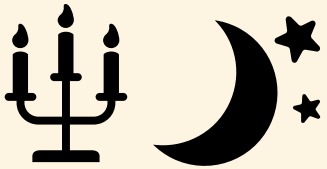
- Weight loss: fewer calories
- Gym, muscle gain: more protein
- Keto or low-carb
- Low sodium
- Plant-based by choice

Goals are personal and set in nutrients rather than ingredients

GrocEyes → plant-based and allergen-free now; calorie or protein goals need nutrition data (not this term)

3 Belief

religion and ethics



- Halal (Islam), kosher (Judaism)
- Vegan, vegetarian (ethics)
- Hindu*: no beef,vegan
- Buddhist*: no meat, vegetarian

*some branches

Muslim students pay more to keep halal; technical names hide animal origin

GrocEyes → halal, kosher and vegan with species-level resolution

Who carries this burden?

The proof: who they are, what they look for, and why choosing food is hard for them.

MEDICAL

WHO

5.5% of schoolchildren have a food allergy

national school survey 2022

WHY IT IS HARD

216 food-allergy injury reports in 2023, up from 62 in 2021

Korea Consumer Agency (CISS)

UNCLEAR LABELS

75.8% of tested foods carry a “may contain” warning

Korea Consumer Agency, 2018 (120 products)

DIET & PERSONAL GOALS

WHO

26.4% of adults do strength training

KNHANES 2025, KDCA

WHAT THEY WANT

59.2% weigh health when they choose food

Opensurvey 2025, ages 15–59, n = 1,500

HOW THEY CHOOSE

62.1% avoid or limit sugar, the top ingredient they cut

Opensurvey 2025, ages 15–59

BELIEF

WHO

≈ 200 k Muslims live in Korea

KMF estimate (150–260 k)

WHY IT IS HARD

2.7 / 5 how easy halal food is to buy, Muslim residents (5 = easy)

KREI survey 2015, n = 250

BIGGEST COMPLAINT

38.3% of Muslim visitors name food the #1 thing to improve

KTO Muslim visitor survey 2016

What these groups spend & opportunities

The money: the markets behind the same three groups, and what the Korean registry shelf offers them.

MEDICAL

₩3 T
care food for older adults (2025)
aT projection

₩237 B
special medical foods (2022)
MFDS reports, via Dailypharm

DIET & PERSONAL GOALS

₩6.04 T
health supplements (2024)
KHSA survey 2024

₩450 B
protein foods (2023), ≈ ₩800 B by 2026
aT estimate, via press 2024

BELIEF

$1.43 T
Muslim food spending worldwide (2023)
DinarStandard SGIE 2024/25; $1.9 T by 2028

$1.19 B
food exports to Muslim countries (2025)
MAFRA 2026, 57 OIC countries

On the shelf: what the registry data adds

Of 1,557,310 consumer food products in the Korean registry, 56% contain an animal ingredient and 9% contain wheat, barley or rye.

Current similar services

Apps scan for one concern or start from a recipe; none checks Korean products for all your needs.

Service	Where	What it does	What it misses for a restricted shopper in Korea
MUFKO (Muslim Friendly Korea)	Korea	AI label scan for haram ingredients; halal map	Halal only; no allergy or vegan profile, no alternatives
Yuka	15 countries, not Korea	Barcode health score for food and cosmetics	Not in Korea; one health score, not 'can I eat this'
Fig	US	Checks products against 2,800+ diets and allergies	US products only; no Korean registry; halal not listed
GrocEyes	**Korean registry + global**	**Profile → every eligible product**	**Several restrictions × tolerance level; evidence per verdict; food, cosmetics, household**

What GrocEyes adds: ① a strictness level per restriction ② the ingredient behind every verdict ③ shoppers' label photos fill the registry's gaps. Services as described on their own sites, checked 21 Oct 2026.

How a user moves through GrocEyes

Set your profile once; GrocEyes filters, you find and buy, and it plans your stock and next trip.



1

Profile

Restrictions (halal, nut, vegan, ...) and how strict: three levels



2

Filter

Shows only products your profile allows, with the reason for each one left out



3

Discover

Browse eligible categories, search online shops, or scan a product in the store



4

Use stock

Tell what is at home; get a meal schedule that uses it up



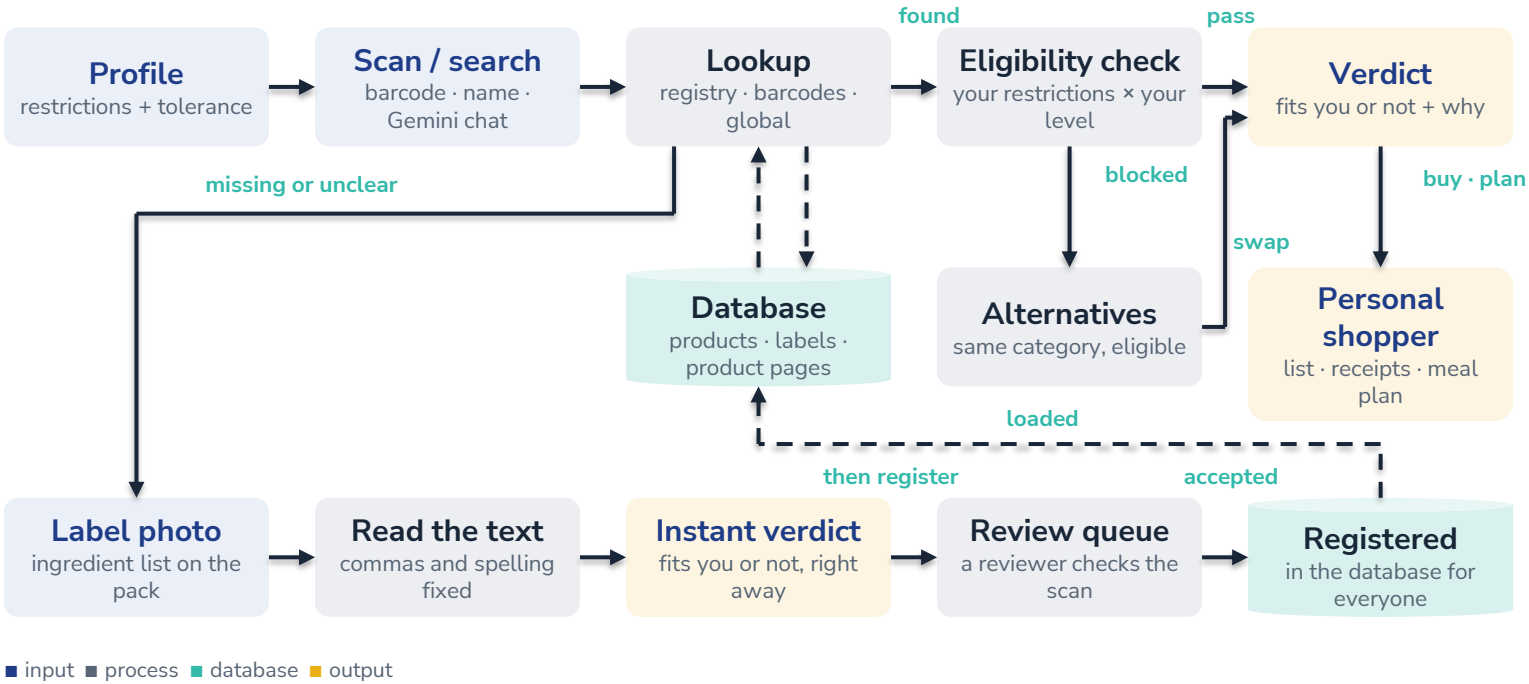
5

Next plan

Estimated next grocery day; change alerts in the final phase

Operating steps: the whole system

Three paths: known products are checked, blocked ones swapped, new ones answered and registered.



Tolerance levels: how strict should the filter be?

You choose how strict each restriction is; each level keeps the rules above it and adds one more.

Illustrative example · rice cracker: rice, sugar, 혼합제제 (additive mix, contents not listed). Pack: “made in a factory that also handles peanuts”.

1 Ingredients only

for: mild intolerance, preference

says NO when

a listed ingredient conflicts

OK

2 Nothing unclear

for: strict diet, halal, vegan

also says NO when

an ingredient is unclear

NO

3 No “may contain”

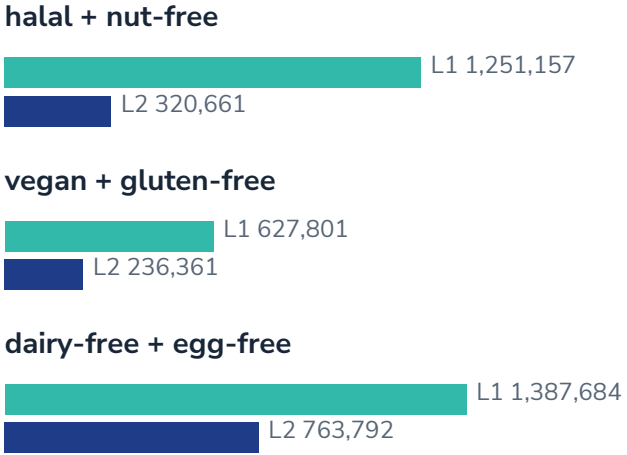
for: severe allergy, anaphylaxis

also says NO when

the pack says “may contain”

NO

Products left for the shopper



Of 1,557,310 Korean products. L1 / L2 = level 1 / level 2: stricter is safer, but leaves fewer products. Level 3 needs the pack's “may contain” line (global data or a label photo); the Korean registry does not record it, so these bars stop at level 2.

Data: what one row is, and what is collected

Only official APIs and open data: seven sources, about 6.2 million products collected so far.

Source	What it is	One row =	Received → kept
Korea food registry (MFDS C002)	every processed food reported in Korea	one product recipe	1,088,954
Korea livestock registry (C006)	meat, dairy and egg products	one product recipe	572,025
Ingredient dictionary (I1020)	official names, other names, species	one ingredient	24,887 → 23,535 names
Barcode registry (I2570, C005)	shelf barcode ↔ registry number	one barcode ↔ report link	313,038 → 104,653 barcodes linked to 61,032 products
Open Food Facts	global food, by the community	one product	4,848,211 → 4,501,893 (1.29 M with ingredients)
Open Beauty Facts	global cosmetics	one product	64,237 → 40,054 (16,300 with ingredients)
Open Products Facts	global household goods	one product	41,133 → 25,127 (2,786 with ingredients)

Korean registry: 1.66 M rows → 1.56 M sold in shops (94 k factory-only ingredients set aside). Official APIs and open data only; every raw file is kept.

End-to-end pipeline

Five stages, left to right: sources, collect and join, transform, database, output.

Sources	Collect + join	Transform	Database	Output
Korean food registry	Keep every file as received	Clean the text	PostgreSQL database	Eligible catalog
Barcode lists	Fingerprint each file	Split ingredient lists	Products, labels, users	Scan in store
Ingredient dictionary	Never overwritten	Recognise ingredients:	Product cards (JSON)	Product check
Global food, beauty, household data	Rebuild from here anytime	땅콩 → peanut		Personal shopper
Shopper label photos		Label: NO · CHECK · OK		Change alerts
		Names → English		

Preprocessing and labels: one real product

One real registry product shows the cleaning and labelling; the bars show the labels over all Korean foods.

One product: 땅콩조림 (braised peanuts) · Sosaengrokeol Food

1 As received the registry's ingredient line

땅콩, 물엿, 간장, 설탕, 정제소금, 콩기름(대두유)

2 Clean + split commas inside ( ) stay together

땅콩 | 물엿 | 간장 | 설탕 | 정제소금 | 콩기름(대두유)

3 Recognise 6 of 6 words matched to an ingredient

peanut | corn syrup | soy sauce | sugar | refined salt | soybean oil

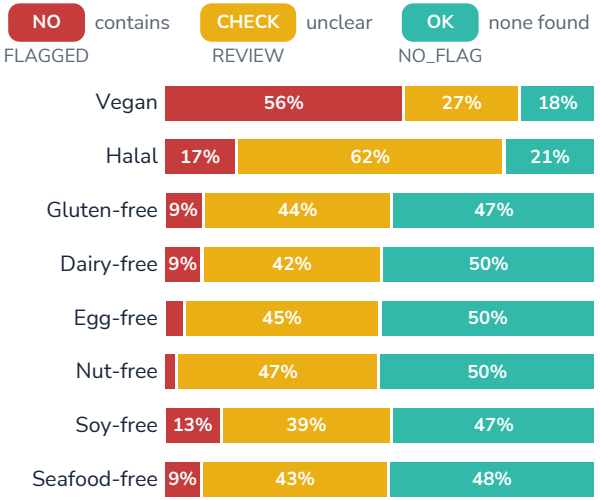
4 Label, with the reason one label per restriction

NO **nut-free** 땅콩 · **soy-free** 간장

OK **halal · vegan**

CHECK **gluten-free** 간장: soy sauce is often brewed with wheat

Labels over 1.56 M Korean foods



CHECK is common: labels often list a mix (혼합제제) without its contents. Labels come from rules, not people; accuracy is measured on a hand-checked sample in November.

Done on the full build: –10,129 old duplicates removed · 10.2 M ingredient words · 97.1% recognised · 49.7 M labels, each with its reason

Raw data: what each source gives

Each source gives different pieces; together they make one table of 6.2 million products.

Source	Rows received	Identity	Category & dates	Ingredient list	Ingredient dictionary	Barcode	Allergen · trace · claim
MFDS C002 · C006 product reports	1.66 M	report no. · name · company	food type · permit and change date	full list, in Korean	–	–	–
MFDS I1020 ingredient dictionary	24,887	–	–	–	official name · other names · class · Latin species · English name	–	–
MFDS I2570 · C005 shelf barcodes	313 K	report no. · name · company	–	–	–	barcode number	–
Open Food / Beauty / Products Facts	4.95 M	barcode · name · brand	category · countries sold	ingredient text, many languages	–	barcode	allergens · may contain · labels such as vegan
Shopper label scan	per scan	name · brand	category	read from the label photo	–	barcode photo	'contains …' · 'same factory …'
All sources together	**6.2 M products**	**one product id · name · company**	**category in English · market · change date**	**10.2 M ingredient words**	**each word → ingredient · species · plant or animal**	**barcode (GTIN)**	**allergen · may contain · claim**

Database design: SQL tables

Facts with a fixed shape go in SQL tables: one row is one thing, linked to the product.

Ingredients — 10.2 M rows

One row = one ingredient word in one product

땅콩 → peanut (2nd ingredient)

Barcodes — 297 K rows

One row = one shelf barcode → one product

8800241150005 → 프로토파낙스

Products — 6.2 M rows

One row = one product: registry, global or shopper scan

국산참깨강정 · Ongoeulhangwa · Snack / biscuit

every table links here by product id

Labels — 49.7 M rows

One row = one product × one restriction

nut-free → FLAGGED (땅콩)

Users & shopping — 8 tables

One row = one profile, stock item or list item

halal 2 · nut-free 3 · eggs 8 pcs

Why SQL here: every row has the same columns, keys keep the tables consistent, and one join answers “which products can I eat?”.

Measured: eligible catalog 90 s → 1.8 s after adding a per-product eligibility table · full build 28 min · database 17 GB

Database design: NoSQL documents

Data whose shape changes is kept whole as a JSON document, in the same PostgreSQL (JSONB).

Product card — 1.6 M documents

everything about one product, read by the app in one go

```
{"name": {"original": "국산참깨강정",
          "en": "Domestic Sesame Gangjeong"},
 "ingredients": [{"text": "땅콩", "en": "peanut"}, …],
 "labels": {"nut_free": {"label": "FLAGGED"}, …}}
```

Label evidence — 36.4 M labels

why a label was given: none, one or many reasons

```
[{"token": "땅콩",
  "reason": "peanut",
  "label": "FLAGGED"}]
```

Raw API page — every update

an API answer kept exactly as it arrived

```
{"C006": {"row": [
  {"PRDLST_NM": "저염훈제",
   "RAWMTRL_NM": "콩기름(대두유), 닭가슴살, …"},
  …]}}
```

Meal plan — per household

the AI's plan as returned, then checked against home stock

```
{"days": [{"date": "2026-10-21",
  "meals": [{"dish": "Egg Fried Rice",
    "uses": [{"item": "Eggs", "amount": 2}]}]}]}
```

Why documents here: the shape differs from product to product. Stored as JSONB and queried with SQL; the MongoDB version would be $unwind → $match → $group.

Prototype: check a product in the store

Scan or search one product: see if it fits your profile, why, and what to buy instead.

국산참깨강정

Domestic Sesame Gangjeong · Ongoeulhangwa · Snack / biscuit

Profile halal + nut-free: **NOT ELIGIBLE**

- **Halal: OK** nothing conflicting found
- **Nut-free: NO** 땅콩 → peanut

Ingredients (English)

corn syrup, peanut, sesame

Original: 물엿, 땅콩, 참깨

Eligible alternatives, same food type

1. **이브로콘스낵**
   Ibeuro Corn Snack · Donga Confectionery
2. **현미100 롱떡뻥 스틱**
   Brown Rice 100 Long Rice Puff Snack Stick · Olbareum
3. **부드러운 롱떡뻥 스틱**
   Soft Long Rice Puff Snack Stick · Olbareum

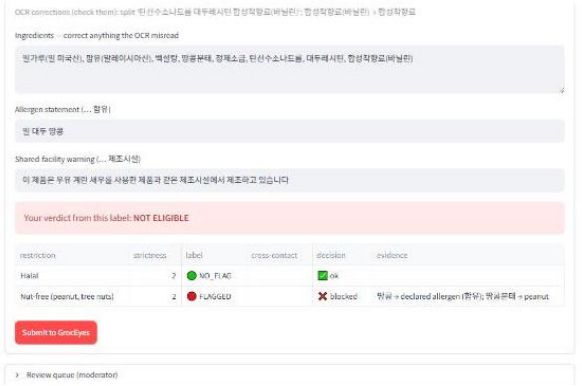


Same check in the app. The alternatives pass level 2; for level 3, check the pack for “may contain”.

Not in the database? Scan it, register it

A product we do not know is answered from its label photo, then reviewed and added for everyone.

1. **Scan** — Barcode typed or read from a photo, checked for typos
2. **Look up** — Registry, barcode links, global catalog, earlier user scans
3. **Photograph** — Missing, or marked CHECK: photo of the ingredient list
4. **Read + repair** — Text read from the photo; commas and spelling fixed (합성작향료 → 합성착향료)
5. **Verdict now** — Same rules as the database; “contains” and “same factory” lines are read too
6. **Review → load** — Waits for a reviewer; accepted scans join the database at the next build



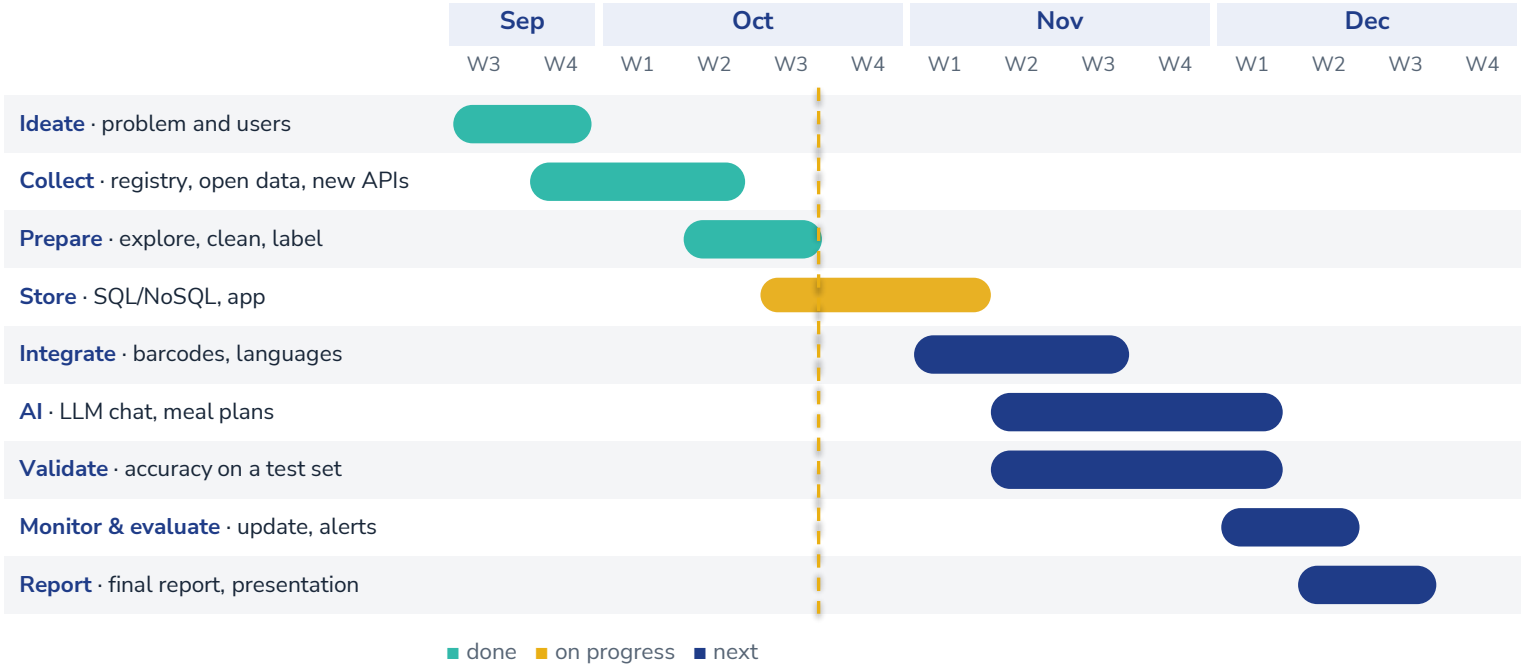
Live verdict from a scanned label, OCR corrections shown; Submit sends it to the review queue.

Why it matters

Only 61 k of 1.65 M registry products have a known shelf barcode, and 27–62% are CHECK; a label photo fills both gaps.

Progress and next steps

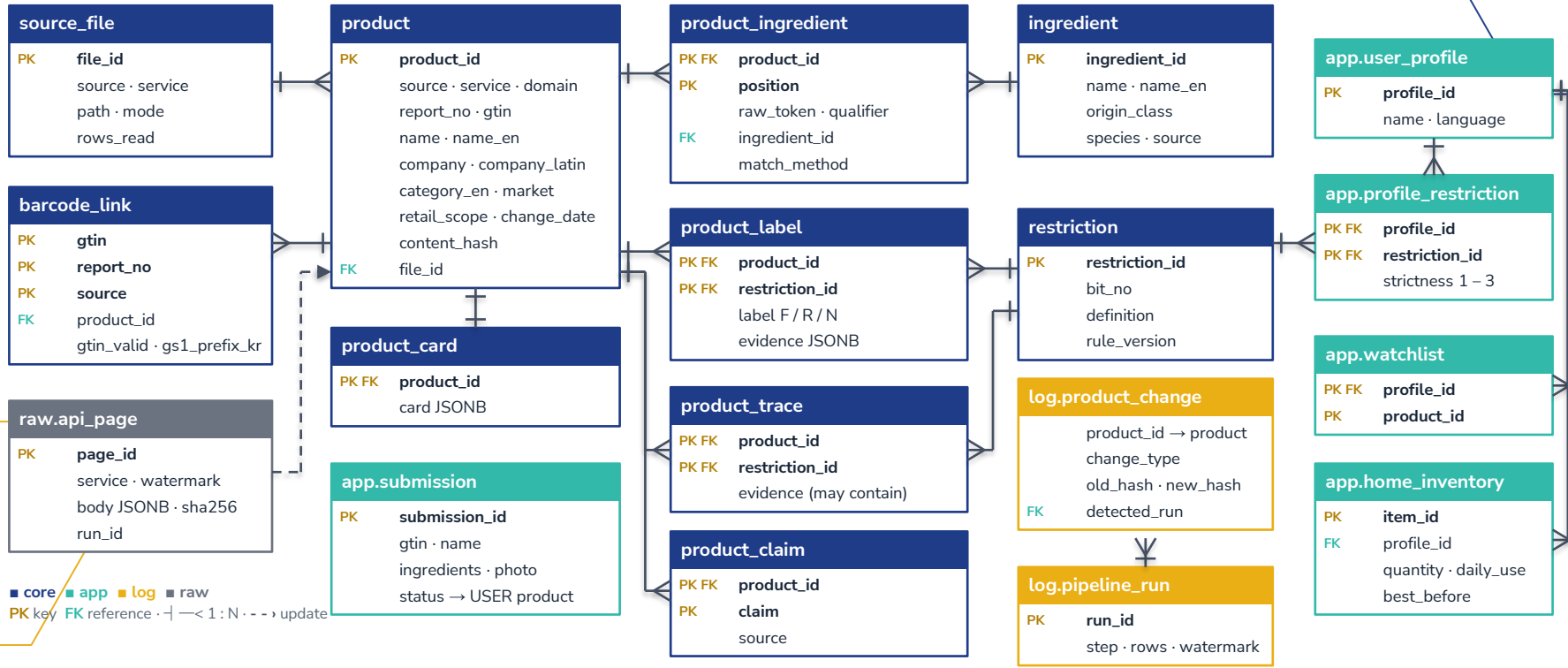
Collect, prepare, store and serve are done; AI, validation and monitoring come next.



Thank you

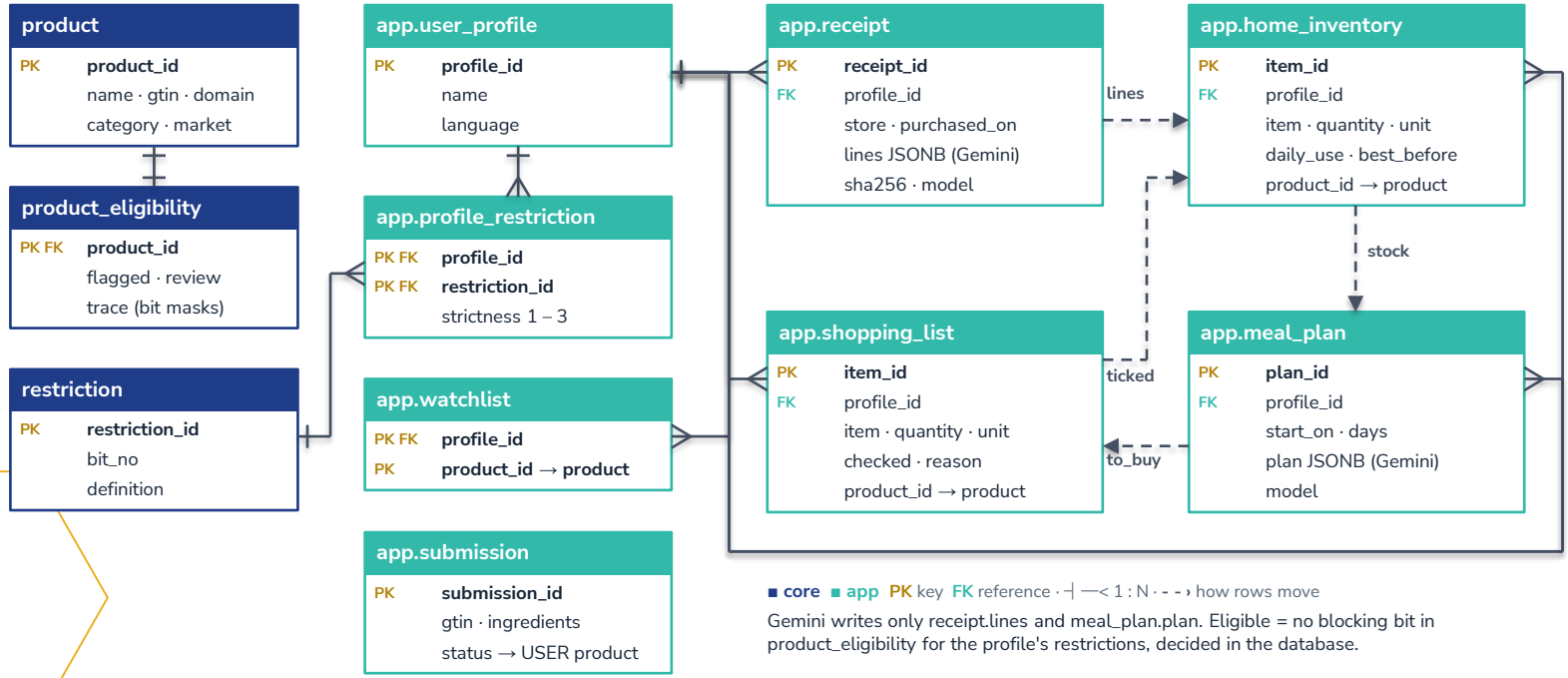
Appendix · ERD: tables and keys in one database

Appendix: every table and key; product is the hub. Eligibility and app tables: next slide.



Appendix · ERD: app schema (personal shopper)

Appendix: the personal-shopper tables; writes only receipt lines and meal plans.



■ core ■ app PK key FK reference · ┤ —< 1 : N · - - › how rows move

Gemini writes only receipt.lines and meal_plan.plan. Eligible = no blocking bit in product_eligibility for the profile's restrictions, decided in the database.

Sources

the research, statistics and data sources behind this deck.

Research

Handani (2021). Halal food awareness of Muslim students in Korea. Research Horizon 1(1).

Zurzolo et al. (2017). Precautionary allergen labelling. Allergy Asthma Clin Immunol 13, 42.

Feltz et al. (2024). Confusion about animal-derived ingredients. SSRN preprint.

Raheem et al. (2021). Texture-modified food for dysphagia. IJERPH 18(10), 5125.

Evert et al. (2019). Nutrition therapy for adults with diabetes. Diabetes Care 42(5), 731-754.

EU Cosmetics Regulation 1223/2009, Annex III fragrance allergens.

Economical Impact (slide 5)

Care food: aT estimates via Dong-A Business Review (2025).

Special medical foods: MFDS production reports via Dailypharm (Mar 2024).

Supplements: Korea Health Supplements Association, 2024 market survey.

Protein foods: aT estimates via Herald Business and Aju News (Aug 2024).

Muslim food: DinarStandard, State of the Global Islamic Economy 2024/25.

Exports to OIC countries: MAFRA press release (Jan 2026).

Market Research (slide 4)

Food allergy: 2022 schoolchildren survey, Allergy Asthma Immunol Res (2026).

Allergy injury reports: Ham, Oh & Kim, Food Law and Policy No. 5 (KCA CISS, 2021–2023).

Strength training: KDCA, KNHANES 2025.

Healthy choice, sugar: Opensurvey, beverage trend report 2025 (ages 15–59, n = 1,500).

Muslims: Korea Muslim Federation estimate via secondary reports.

Halal buying ease: KREI, 국내 할랄식품 시장 실태와 대응과제 (2015).

“May contain”: Korea Consumer Agency, allergen labelling report (Mar 2018).
Muslim visitors: KTO 2016 survey, via MNB (Feb 2017).

Data and services

MFDS Food Safety Korea Open API: C002, C006, I1020, I2570, C005 (Oct 2026).

Open Food / Beauty / Products Facts, ODbL 1.0.

Claude is used for this project to brainstorming and writing, but all idea comes from the author