TECHNOLOGY · DATA

Find the Data Your Best AI Ideas Depend On, and Give It Owners

An OmniTek Consulting implementation guide for VPs and Directors in pharma, biotech, and medical device companies.

READING TIME
About 8 minutes

WHO IT'S FOR
VPs and Directors

GOES WITH
The AI Maturity Quick Check

> Most AI ideas don't fail on the AI. They fail because nobody knows where the data lives, who can approve access, or whether it can be used this way at all. Mapping the two or three data sources behind your best ideas, and giving each one an owner, turns months of chasing into a few weeks.

SCOPE This guide covers AI used in how your company works, such as drafting, analysis, and operations. It doesn't cover AI that is part of a product you sell, such as AI-enabled device software or software as a medical device. That has its own regulatory path; check it with your Regulatory team.

Why it matters

- **Data access is the slowest step.** A pilot that could run in two weeks can wait months for someone to find the right system owner and agree to share an extract.
- **Not all data can be used the same way.** Trial participant data may be limited by what participants consented to. GxP records must keep their integrity. Some data can't leave a validated system at all.
- **Quality of the answer follows quality of the data.** An AI summary of inconsistent deviation records is a confident summary of inconsistent records.

Knowing your data before you choose a tool also tells you which tools are acceptable, which is why this guide pairs with the guide on getting a secure tool approved.

Five data classes

Sort every data source into one of five classes. The class decides which tools can touch it and what approvals you need. Use your company's existing classification if it has one; this maps onto most.

Class	Examples	AI rule of thumb
Public	Published papers, agency guidance, your public website	Any approved tool
Internal	Meeting notes, training material, internal SOPs	Approved enterprise tools only, with contract terms that stop vendors training on your data
Confidential	Aggregated unpublished study results (see below), pricing, manufacturing processes, partner contracts	Approved enterprise tools with no training on your data and access limited to the team
Personal and patient	Trial participant data, safety case narratives, patient support records, employee records	Only tools approved for this class, within the limits of consent, privacy law, and contracts; de-identify where you can
GxP records	Batch records, deviations, lab data, trial master file, safety database entries	Stay in validated systems; AI works on them only through a classified and approved use

If a source mixes classes, it takes the strictest one. A deviation report that names a patient is both a GxP record and personal data.

Unpublished study results split three ways. Aggregated results with no participant data or partner restrictions are Confidential. Participant-level data is Personal and patient. Results under a partner or licensing confidentiality clause, and topline results not yet publicly disclosed, are also Confidential, with an extra handling rule: they may go only into named tools, or not into AI tools at all. For a listed company, undisclosed topline results can be market-sensitive.

A data map for each AI idea

Fill one in for each of your top AI ideas. If you can't answer a line, that's the gap to close first.

TEMPLATE · COPY AND ADAPT

Data Map: [AI idea]

Idea owner: [ ] | Date: [ ] | GxP class of the use: [Non-GxP / GxP supporting / GxP direct]

Data source	System	Data owner	Class	Quality today	Access route	Limits on use
[e.g. deviation reports, 2023–2026]	[QMS name]	[name, role]	[GxP record]	[complete? consistent? coded?]	[export, API, read-only view]	[consent, contract, privacy, validated-system only]
[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

Open questions: [ ] Decision: [ready / ready with conditions / not yet] Signed: [idea owner] [data owner] [Quality, required if GxP supporting or GxP direct]

The data owner is the person who can say yes to a new use of the data and is accountable for its quality, usually the business lead for the system, not IT. Name a person, not a department.

Your first 30 days

1. **Week 1: Pick your top three AI ideas.** Use ideas your team already wants, not a long list. Three is enough to find the patterns.
2. **Week 1: List the data each one needs.** For each idea, name the two or three data sources it depends on and the system each lives in.
3. **Week 2: Find the owners.** For every source, name the person who can approve a new use. Where nobody owns it, that's your first escalation to leadership.
4. **Week 3: Check class, quality, and limits.** With each owner, classify the data, look at a real sample for gaps and inconsistencies, and confirm any consent, contract, or privacy limits. Bring Privacy or Legal in for anything in the personal and patient class.
5. **Week 4: Agree an access route.** For each source, agree how an approved AI use will reach it: an extract, a read-only connection, or staying inside the validated system. Record it on the data map.

By day 30 you know which ideas are ready, which need data work first, and who to call. That covers the data question in the Quick Check at the Operationalizing level.

Data integrity when AI is involved

The data integrity principles your department already follows (often summarized as ALCOA+) still apply when AI touches GxP data. Here's what each one asks of an AI use.

Principle	What it means when AI is involved
Attributable	You can tell who ran the AI, on what input, and who accepted the output
Legible	AI output kept as a record is readable and unambiguous, so a reviewer can understand it
Contemporaneous	Review and acceptance of AI output are recorded when they happen
Original	The source record stays the record. An AI summary never replaces the original deviation, case, or batch record
Accurate	A person checks AI output against the source before it's used in GxP work
Complete	Everything needed to recreate how an AI output was produced is kept: the full input set, the request, and the reviewer's changes
Consistent	The input set is versioned and the AI use follows a defined process, so the same question asked again can be traced to the same data
Enduring	AI output kept as a record, and the inputs behind it, exist for the full retention period, not only in the AI tool's history
Available	AI output kept as a record can be retrieved in a readable form for review, investigations, audits, and inspections throughout the retention period

For GxP supporting uses, most of this is covered by a documented review step. For GxP direct uses, it belongs in the validation and monitoring plan.

Common mistakes

- **Starting a big data cleanup first.** A department-wide data program takes years. Map only the data behind your top three ideas and fix what those ideas need.
- **Making IT the owner of everything.** IT runs the systems; the business owns the data. Without a business owner, nobody can approve a new use.
- **Assuming trial data can be reused freely.** Informed consent, data protection law, and partner contracts can limit secondary use. Check before the pilot, not after.
- **Exporting GxP data to make AI easier.** Copying records out of a validated system to feed a tool can create an uncontrolled copy. Confirm how your data integrity SOP treats copies and extracts before you build the input set. Agree the access route with Quality first.
- **Testing on clean samples only.** Pilots run on hand-picked data look great, then fail on the real thing. Test on a representative sample, gaps and all.

Regulatory touchpoints

Status is as of October 2026.

Source	Status	What it means for AI and your data
FDA: Data Integrity and Compliance With Drug CGMP, Q&A	Final, December 2018	The integrity expectations for CGMP data apply whatever tool touches it
PIC/S PI 041-1: Good Practices for Data Management and Integrity	Final, July 2021	Data governance and ownership expectations used by many inspectorates
21 CFR Part 11	In force	Electronic records and signatures in GxP processes need audit trails and access controls
EU GMP Annex 22 (AI)	Draft published for consultation in July 2025; not yet final (status)	Expected to set expectations for the data used to build and test AI models in critical GMP uses
Privacy law and consent (for example HIPAA in the US, GDPR in the EU)	In force; depends on your role and markets	Limits on using personal and patient data, including secondary use of trial data

Trial sites are usually HIPAA covered entities; sponsors (pharma, biotech, and device companies) generally are not (SACHRP, an advisory committee to the HHS Secretary, 2021; see also HHS on covered entities). Once trial data reaches the sponsor, what it may do with the data is governed mainly by the participant's informed consent (and any HIPAA authorization signed at the site), the trial agreements, and privacy laws such as the GDPR.

This guide is general information, not legal or regulatory advice. Your Quality, Privacy, and Regulatory leads should confirm how these apply to your data and markets.

How OmniTek can help

The full AI Maturity Report Card measures data quality and access alongside the other 12 maturity dimensions, so you can see whether data is your real bottleneck or a symptom of something else. It comes with the **use-case scoring kit**, which helps you pick the ideas worth mapping first.

To talk through your data or your Quick Check results, book time with OmniTek.

Book time with OmniTek

SOURCES

- FDA: Data Integrity and Compliance With Drug CGMP, Questions and Answers (December 2018)
- PIC/S PI 041-1: Good Practices for Data Management and Integrity in Regulated GMP/GDP Environments (July 2021)
- eCFR: 21 CFR Part 11
- European Commission: Stakeholders' consultation on EudraLex Volume 4, Chapter 4, Annex 11 and new Annex 22 (July–October 2025)
- HHS: Covered Entities and Business Associates
- SACHRP (advisory committee to the HHS Secretary) recommendations: New Challenges in Interactions among Sponsors, Clinical Trial Sites, and Study Subjects (March 2021)

This guide is general information for leaders in regulated life sciences functions, not legal or regulatory advice. Confirm requirements with your Quality, Regulatory, and Legal teams. © 2026 OmniTek Consulting.