

KNOWLEDGE LIBRARY

What Is a Data Farm?

A Future Vision for an Industry That Learns Together



Author

Ender Öztürk

OETECH Solutions, Bursa, Türkiye

Published

18 October 2026

Version · Language

1.0 · English

ABSTRACT

Growing data together, harvesting knowledge together.

A problem solved in one factory resurfaces in another factory years later. An energy loss that one company uncovered through long trial and error keeps growing, unnoticed, on another company's bill. Experience gained with the same machine family, a similar material or the same production method usually stays within the walls of a single company.

What if companies could learn from each other's experience without revealing their trade secrets or compromising their people's privacy? What if an improvement validated in one factory could become the starting point for research in another factory operating under similar conditions? What if this body of knowledge, spanning production and quality, energy and training, could speed up how fast an entire country learns together?

Keywords: Data Farm · AI · Digital Twin · Energy · Industry 5.0

Suggested citation: Öztürk, E. (2026). What Is a Data Farm? A Future Vision for an Industry That Learns Together (Report No. 2026-05). OETECH Solutions. <https://oetechsolutions.com/nedir/veri-ciftligi-nedir>

Report No.	OETECH-TR-2026-05
Author	Ender Öztürk — Enterprise Solutions Account Manager, OETECH Solutions, Bursa, Türkiye
Published	18 October 2026
Version / Language	1.0 / English
Scope	27 sections, 9 figures, 22 Q&As
Online version	https://oetechsolutions.com/nedir/veri-ciftligi-nedir

This report is the print edition of the online guide in the OETECH Knowledge Library. Example calculations, budgets and timelines are explicitly hypothetical; they are not customer results, price quotes or performance commitments. Images are conceptual illustrations.

CONTENTS

— Introduction

1 Why the farm metaphor?

2 How is it different from a data center or a report pool?

3 What does it mean from an Industry 5.0 perspective?

4 What data should be grown?

5 What does code-based collection achieve, and what doesn't it?

6 Does all the data have to be moved to one place?

7 How do we learn from the collected data?

8 How is one factory's experience carried over to another?

9 What role does the digital twin play in this structure?

10 How should the dark factory goal be approached?

11 Why should energy efficiency be one of the main harvests?

12 Why is HVAC / climate control part of production?

13 How does the carbon footprint become visible and shrink?

14 How should employee multi-skilling (polyvalence) and training be assessed?

15 What can be done for OHS and accident prevention?

16 Why are work, breaks and rest a data topic?

17 How should food, protein and employees' energy needs be addressed?

18 How does production knowledge create marketing value?

19 How does it contribute to logistics, assembly and supply resilience?

20 Which areas might users not think of at first, but should be added?

21 What do contributing companies gain?

22 Who should govern this structure?

23 How should the rights to data and models be protected?

24 How can the budget and roadmap for this dream be approached?

25 How should we measure success?

26 What this dream means for OETECH

— Sources and vision note

A Frequently asked questions

Introduction

Data Farm is the name for this future that OETECH founder Ender Öztürk has envisioned for many years. It is the idea of an industrial ecosystem, at regional or national scale, that participating factories feed with selected data, and that turns this data into reliable knowledge, validated models and actionable solutions.

At the heart of the vision is a cycle: **Feed it with data. Let the knowledge mature. Put it to work for production and marketing. Feed the results back into learning.**

This article does not announce an established national platform or public funding that has been secured. It is a design for the future, one that requires major investment, long-term partnership and a serious framework of trust. But a big dream does not mean the first step has to be vague.

1. Why the farm metaphor?

Simply sowing seeds in a field is not enough to get a crop. Preparing the soil, using water wisely, tending, time and deciding how to use the harvest all matter. Data needs a similar process.

A signal from a machine or a record entered by an operator is the starting material. Cleaning records, linking them through shared definitions and placing them in the right context is the tending. Showing that an analysis works in an independent trial is the knowledge maturing. And when that knowledge improves a factory's decision or way of working, that is the harvest.

Mature data is not data that has simply sat around for a long time. It is knowledge whose source is known, whose quality has been assessed, whose usage rights are defined and whose benefit for a specific job has been validated.



Figure 1. The harvest is not a data table getting bigger; it is benefit that can be verified in the business and in people's lives.

In this metaphor, a contributing factory is not just a data supplier. It should get something back: better benchmarking, early warnings, validated methods, training and new business opportunities. Mutual benefit should be a founding principle of the system, not a feature bolted on later.

2. How is it different from a data center or a report pool?

A data center is computing and storage infrastructure. A data pool can bring records together. A Data Farm aims to build **shared meaning, learning, validation, rights management and benefit sharing** on top of these.

Table 1

Layer	Role	Value created
Factory data infrastructure	Reliably recording events and conditions	Traceable production reality
Shared data dictionary	Mapping codes, units and indicators	Comparable information
Authorized sharing space	Defining which information may be used by whom	Trust and controlled collaboration
Learning and research	Developing relationships, predictions and methods	Testable models and hypotheses
Validation	Measuring usefulness across different facilities and conditions	Solutions with known limits
Application and harvest	Bringing knowledge into agent, engineering or planning services	Measured business benefit

The shared model does not have to be a single AI that knows everything about every factory. Separate models for energy, quality, maintenance and planning; knowledge search for document access; and [AI agents](#) for explanation and task execution can all work together.

3. What does it mean from an Industry 5.0 perspective?

Reading Industry 5.0 as merely more sensors, more robots and fewer people misses the point. The European Commission's framework emphasizes human-centricity, sustainability and resilience. ¹

The Data Farm aims to bring these three goals to the same table. While a company produces more reliably with fewer resources, employees' skills, health and voice must also be protected. While flexibility against supply chain shocks is gained, environmental impacts must become visible.

Progress in national competitiveness does not come from cheaper production alone, either. Learning faster, solving quality problems earlier, developing skills, reducing energy dependence and giving customers reliable evidence all matter together. The Data Farm should be seen as infrastructure that can support these, not as a guarantee of automatic superiority over other countries.

4. What data should be grown?

The starting question should not be "What can we collect?" but "What data do we need to solve which problem?" Production events, resource conditions and outcomes become meaningful when they are linked.

Table 2

Area	Data that could be selected	Benefit that could be explored
Machine fleet	Machine family, technology, capacity range, age and maintenance context	Fair benchmarking for similar resources
Production	Operations, cycle distribution, quantity, changeovers and waiting	Bottleneck and method improvement
Downtime	Start/end, cause, intervention and recurrence	Reducing recurring losses
Quality	Defect type, inspection method, lot and rework	Early risk detection and solution transfer
Process	Temperature, pressure, speed and similar variables within the approved scope	Exploring condition–outcome relationships
Assembly	Station times, error types, work sequence and material supply	Line balancing and ergonomic review
Warehouse / shipping	Picking, preparation, waiting, loading and transport times	Flow and delivery reliability
Energy	Submeters, power profile, fuel, compressed air and idle consumption	Measured savings opportunities
Environment	Temperature, humidity, air quality, noise and zone measurements	Suitable working and process conditions
Maintenance	Failure, part replacement, work performed and verified outcome	Developing maintenance strategy
Competency	Task-based training and validity of authorizations	Training plans and safe task assignment
Work organization	Shifts, workload, adequate rest and team planning	Reducing overload and unplanned work
OHS	Risk assessment, near misses, hazards and corrective actions	Preventive learning
Resource cycle	Water, waste, reuse and material loss	Circularity and environmental improvement
Market capability	Processes, capacity class and certificates the company chooses to publish	Matching the right manufacturer with the right demand

This is not a list for unlimited collection. For each area, the purpose, data quality, sharing level and retention period must be defined. Personally and commercially sensitive areas should not be opened to shared model training by default.

5. What does code–based collection achieve, and what doesn't it?

Shared coding structures can be built for factories, machines, product families, operations and defect classes. This can reduce the disconnect that arises when two companies record the same concept using different words.

However, coding is not anonymization on its own. Writing an employee code instead of a person's name does not remove the personal-data status of a record if the person can be identified by matching it with other records. Likewise, a rare combination of machine and product can reveal the source even if the company name is hidden. ²

The three proposed sharing levels are as follows:

Table 3

Level	Where is it used?	Example
Internal detail	In the factory's own authorized system	Work orders, person-level authorizations, sensitive process details
Contractual limited sharing	Within a defined project or supply relationship	Quality/traceability records shared with an authorized buyer
Shared learning output	After appropriate protection and re-release review	Comparable group results that do not reveal the source

For small groups, methods such as report suppression, reduced detail, time delays and query limits can be considered in combination. None of them is an automatic security certificate for every data type. Technical testing and relevant legal review are required.

6. Does all the data have to be moved to one place?

No. Instead of a single central pile, a Data Farm can be built as a **network in which factories keep control of their data and are connected through shared standards.**



Figure 2. Shared learning does not require sharing every raw record with everyone.

The center could host a shared catalog, data definitions, rights registers, research services, model versions and outputs approved for release. Some raw records can stay at the factory. Some work can be carried out in a secure project space, and some with approved aggregated data.

For suitable tasks, federated learning allows a model to be trained on local data and the model updates to be combined. But this method is not a privacy guarantee in itself either. Information may leak from the updates or from the trained model. Depending on the threat model, additional methods such as secure aggregation, differential privacy and output controls should be considered, and the balance between utility and protection should be tested. ³

Industrial data ecosystems such as Catena-X show that the ideas of shared standards and control over data are also being worked on around the world. This example does not mean the Data Farm is ready or implements the same scope; it is a reference worth studying in the design. ⁴

7. How do we learn from the collected data?

No single model should be expected to solve every industrial problem. Different tasks require different methods:

- **Statistics and benchmarking:** Understanding cycle, energy or quality distributions under similar conditions.
- **Predictive models:** Estimating risk or the expected outcome under specific conditions.
- **Optimization:** Calculating options for capacity, routing, work sequence or resource allocation.
- **Knowledge retrieval:** Finding approved instructions and past solution documents in the relevant context.
- **AI agents:** Researching authorized sources and drafting evidence and proposed actions.
- **Digital twin and simulation:** Testing a change before it is applied physically.

Permission to use data for model training must be defined separately. Access granted to prepare a report should not automatically count as the right to train a model or to distribute that model to third parties.

Errors in the data, wrong labels and malicious contributions can also corrupt learning. That is why tracking data provenance, quality control, versioning and keeping the research environment separate are important.

8. How is one factory's experience carried over to another?

A similar machine name does not mean the same production conditions. Material, tooling, tolerances, operator methods, maintenance and environmental conditions can all differ. One of the most important requirements of shared learning is not to lose sight of these differences.

Suppose, for example, that a pattern increasing energy consumption has been found in a certain machine family. Rather than sending a new setting directly to a second facility, what should be shared is under which conditions the pattern occurs, which data is needed and how the recommendation should be validated. The local engineering team should then evaluate it under its own conditions.

The maturity level of a solution must be visible:

Table 4

Maturity	Meaning
Observation	A relationship or recurrence was seen in the data
Hypothesis	There is a possible explanation and a proposed test
Single-site validation	Shown to work under defined conditions
Multi-site validation	Its limits were examined in different environments
Version-controlled use	A service that is monitored, updated and rolled back when needed

Training and test data must not be mixed with repetitions of the same part, the same event or the same factory. Tests on new facilities, different time periods and unusual conditions should be carried out. Alongside good average results, performance for small businesses and different process groups should also be evaluated.

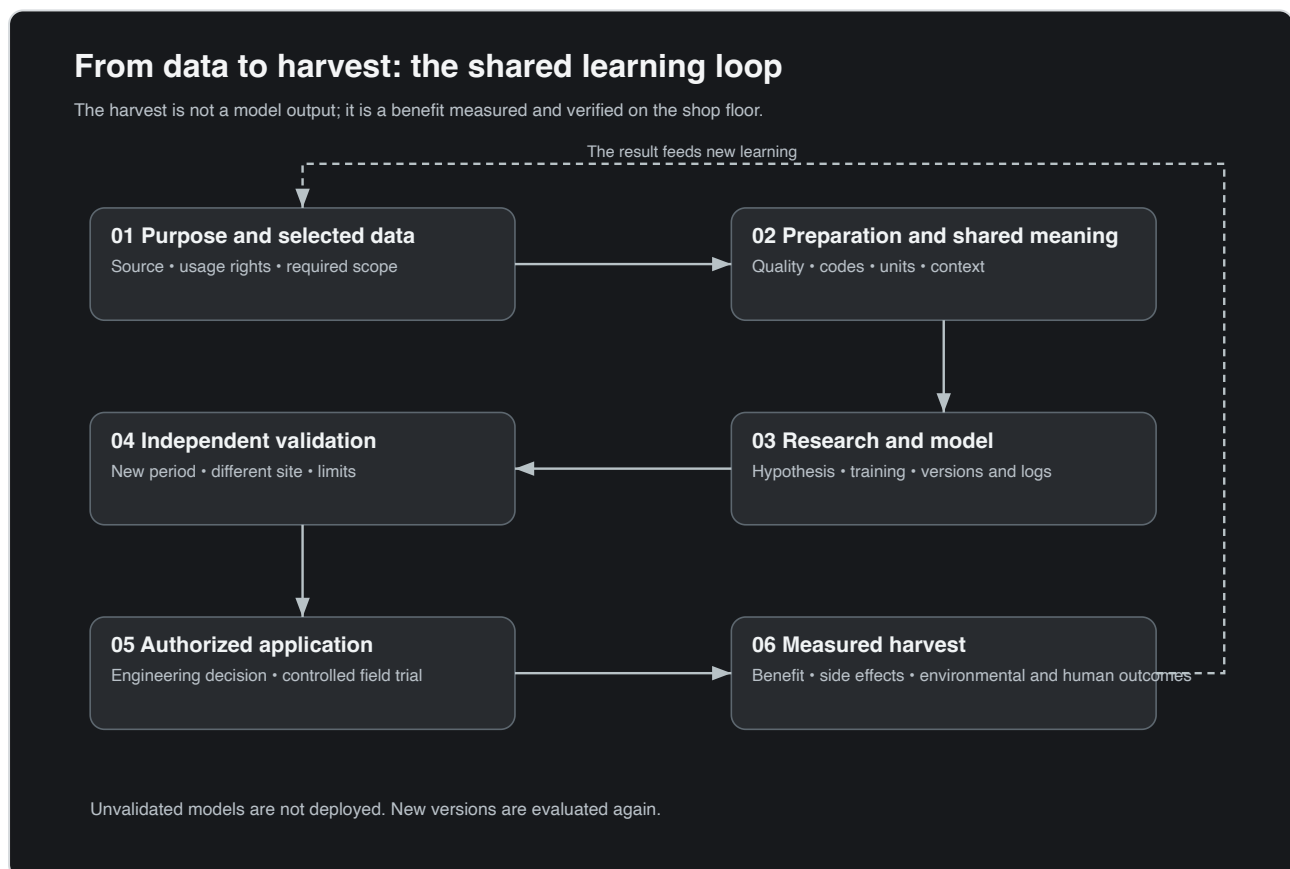


Figure 3. The harvest cycle: learning is not considered complete until the result of the application has been measured.

9. What role does the digital twin play in this structure?

A digital twin is not just an impressive 3D model of a factory. It is a model built for a specific purpose that has a data relationship with the physical system, represents its behavior and has validated results. It has to be reliable enough for the decision it will be used for. ⁵



Figure 4. Visuals make communication easier; decision reliability depends on how well the model represents the real system.

The Data Farm can support shared machine and process definitions, validated behavior models and scenario experience. Each factory's twin must still be calibrated to its own local conditions.

A company can first explore in a virtual environment the impact of a new order load, a different shift pattern, an additional machine or a change in material supply. Results should be presented with assumptions, uncertainty and validity limits rather than as a single definitive number.

Scenarios such as an alternative plan during machine downtime, spreading energy demand to other times or changing warehouse flow can be compared before any real investment. A solution that looks good in simulation must not be transferred automatically to the physical system without on-site validation.

10. How should the dark factory goal be approached?

A dark factory (lights-out factory) describes a high level of automation in which certain production operations can run continuously without people present. Not every factory needs to get to the same point; product variety, process stability and investment economics are the deciding factors.

The Data Farm can support automation decisions by making failure patterns, quality escapes, material flow and changeover requirements visible. But model success does not replace mechanical reliability, safety design, fire protection or maintenance organization.

From the perspective of a human-centric industry, the goal is not to devalue people; it is to reduce dangerous and repetitive work while developing technical expertise and the capacity for safe intervention. Even where there are lights-out production areas, human work continues in maintenance, development, validation and decision-making.

11. Why should energy efficiency be one of the main harvests?

Energy is not just the electricity a machine consumes on its own. Compressed air, process cooling, furnaces, steam, pumps, ventilation and idle running must be assessed together. Recovering waste heat under suitable conditions is another area worth exploring. ⁶

Using validated methods from facilities operating under similar conditions, the Data Farm can support research into questions such as:

- Which systems stay at high consumption when there is no production?
- Why does consumption for the same job increase over time?
- Where are compressed air leaks or inefficient cooling likely?
- Can the production plan spread energy demand more evenly without compromising quality?
- Is there a local or regional match for heat recovery?
- Did the intervention actually deliver net savings?

When comparing the consumption of similar machines, product, material, cycle, capacity utilization and environmental conditions must be matched. A facility that produces less and therefore uses less energy is not necessarily more efficient.

The platform's own server, network, storage and cooling consumption must also be included in the balance sheet. The environmental value of shared learning should not be calculated while ignoring the impact of the digital infrastructure.

12. Why is HVAC / climate control part of production?

A factory's temperature and humidity can affect both process conditions and employee exposure. In some types of production, the stability of temperature, humidity and process cooling is important for keeping products within specification. The requirement is determined by the product and the process. ⁷



Figure 5. Climate control is managed by designing working conditions, product requirements and the energy balance together.

An environment module within the Data Farm could examine zone temperature and humidity measurements, process load, outdoor conditions and maintenance status together. Heat maps can be used to make risks and conditions in the work area visible, rather than to continuously monitor people's bodies.

A single thermometer may not represent the whole hall. Radiant heat, air movement, the physical workload of the task and protective equipment also matter in heat stress assessment. The required measurement and assessment method should be determined with OHS and relevant technical experts. Heat stress control addresses engineering measures, an appropriate work-rest schedule and access to water together. ⁸

Cutting necessary ventilation or employees' rest conditions in the name of energy savings must not be counted as success. Inefficiencies should be eliminated while preserving air quality, process needs and exposure conditions. CO₂ measurement can contribute to assessing ventilation; it does not replace measuring all industrial contaminants.

13. How does the carbon footprint become visible and shrink?

The Data Farm can be used to collect activity data consistently and to track the impact of reduction actions. A carbon calculation, however, requires defined boundaries and appropriate emission factors; it should not be derived solely from production volume or an AI estimate.

Under the GHG Protocol framework, the general distinction is as follows: Scope 1 covers direct emissions from sources the company owns or controls; Scope 2 covers indirect emissions associated with purchased energy; and Scope 3 covers other indirect emissions in the value chain.⁹

For example, fuel, electricity, refrigerant losses, materials and transport data can each be handled in their own categories. A corporate inventory and a product carbon footprint are not the same calculation. Allocating a facility's emissions to products also requires an appropriate method and boundary.

In an entirely hypothetical calculation, suppose an annual electricity reduction of **120,000 kWh** is verified under comparable production conditions. Using a factor of **0.40 kg CO₂e/kWh**, chosen purely for illustration, the result is **48,000 kg**, or **48 tonnes CO₂e**. This factor is not presented as Türkiye's current official factor; a real report must use a factor appropriate to the relevant year, region, method and energy supply.

Trends should track both total emissions and suitable intensity indicators. Emissions per product can fall while total emissions rise because of higher output. In addition, one company's Scope 1 may appear in another company's Scope 3; double counting must be avoided when calculating network totals.¹⁰

14. How should employee multi-skilling (polyvalence) and training be assessed?

Multi-skilling (polyvalence) means an employee being competent in more than one task or station. What is valuable from a Data Farm perspective is not reducing a person to a single productivity score, but understanding which training and working conditions support the safe development of competence.



Figure 6. Competence is not just a training attendance record; it gains meaning through appropriate practice, assessment and a valid task authorization.

A training matrix can link training content, assessment, task authorization, renewal date and on-the-job practice. Critical competency gaps or expired authorizations can become visible at the time of task assignment. Mastery should not be inferred directly from training attendance.

Without exposing individual employee histories to the shared ecosystem, it is possible to explore which training methods lead to better learning for specific tasks, which skills are needed regionally and which training materials are useful.

Employees gaining skills should not be used only to cover that day's production gap. Development opportunities, a safe workload and fair task assignment must also be part of the design. If a personal competency record that can be carried to another company is considered, the employee's access and control over sharing must be designed separately.

15. What can be done for OHS and accident prevention?

When accidents, near misses, risk assessments and corrective actions are examined together, a space for shared learning can emerge. Recurring hazards on similar equipment, points that must be checked before starting work and effective engineering controls can be shared.

The goal is not to find more incident records and blame employees. It is to eliminate hazardous conditions, design work more safely and make reporting easier. An increase in near-miss reports can also initially reflect a better reporting culture; it should not automatically be concluded that safety has deteriorated.

AI vision inspection can support observations of human-machine proximity or equipment use in certain areas. However, a camera model does not replace a validated safety system or a risk assessment. Loss of view, false alarms and detection limits must be evaluated in the real environment.

Instead of sharing names, health details and unnecessary images, the shared platform can use information such as hazard class, conditions, measures applied and verified outcomes. No promise should be made to eliminate the risk of workplace accidents entirely.

16. Why are work, breaks and rest a data topic?

Workload, shift structure, adequate staffing, opportunities to rest and environmental conditions should be examined together. The goal is not to shorten breaks or continuously monitor employees; it is to see whether the organization is suited to people working safely.

Topics such as task rotation, excessively long uninterrupted work, the need for recovery in hot environments and insufficient backup coverage can be explored. Where needed, employees' views and the assessments of relevant experts should accompany the numerical records.

If output drops when someone takes a break, that does not mean the break is wrong. There may be a gap in team planning, buffer stock, work distribution or backup design. The solution may be to organize the flow to meet human needs rather than to eliminate rest.

Rest requirements related to heat and physical workload must be addressed through expert assessment. A data model must not propose an "optimum" that goes beyond the applicable work and rest rules. In hot conditions, adequate cooling and access to water are basic protective measures.⁸

17. How should food, protein and employees' energy needs be addressed?

Human energy cannot be managed through a single indicator, the way a machine's electricity consumption can. Adequate, balanced, varied and safe nutrition matters; needs vary with a person's characteristics, physical activity and health status. Protein alone does not represent overall nutritional adequacy.¹¹



Figure 7. Employees' nutrition, rest and well-being are valuable in their own right. Data should be used to improve these services.

To start with, the cafeteria service can be examined instead of tracking individual consumption: expert review of menus, service hours suited to shifts, access to water, food safety, food waste and voluntary employee feedback are all meaningful topics.

If personalized nutrition is needed, appropriate professionals such as the occupational physician and a dietitian must work within their own scope of authority and confidentiality. A person's diagnoses, blood test results or personal dietary history should not be transferred to the shared industrial model by default. It should not be assumed that all conditions for use and sharing are resolved simply by stating that employee consent was obtained.

A change in meals coinciding with a change in production performance does not prove that nutrition directly caused the difference. Factors such as workload, environment, equipment, shifts and learning must also be investigated. If health research is to be conducted, appropriate ethical and scientific processes are required; commercial performance pressure must not take precedence over employee well-being.

The harvest in this area should not be a narrow goal like "getting more output from employees"; it should be better nutrition, safe working conditions, accessible support and sustainable work organization.

18. How does production knowledge create marketing value?

What a company can reliably do is valuable from a marketing perspective. Process capability, workable material classes, size ranges, quality certificates and verifiable production capability can make it easier to match the right customer with the right manufacturer.



Figure 8. Shared knowledge can bring the capabilities a company chooses to publish to the right demand. Confidential prices and competitors' commercial plans are not brought onto a shared market screen.

The company should choose which information is visible in the capability catalog. Detailed quotation and order information can be shared between the parties under appropriate contracts and access. For regional training or investment needs, appropriately assessed group information that does not reveal the source can be used.

The Data Farm must not become a mechanism for joint price setting, exposing competitors' current costs or sharing customers. Exchanging competitively sensitive information can create risk even when it goes through an intermediary platform. Separate review is required, especially for current/future prices, wages, commercial plans and company-level sensitive capacity information. Aggregation does not automatically ensure compliance in every case either. ¹²

19. How does it contribute to logistics, assembly and supply resilience?

Shipping time is not just the time a vehicle spends on the road. Searching for products, picking, packing, paperwork, loading, waiting and customer delivery processes can each be tracked separately. In assembly, part supply, work sequence, station balance and quality repeats matter.

Shared learning can explore the impact of methods used in similar flows. Reduced waiting achieved by improving how materials are supplied in one facility may not be a result to copy directly at another facility, but rather a method that can be tried.

At regional scale, a map of alternative production capabilities could also be considered for disasters, power outages or critical supplier problems. This map does not have to be a screen that exposes everyone's confidential capacity and orders. In an authorized crisis scenario, access limited to appropriate participants and purposes can be designed.

20. Which areas might users not think of at first, but should be added?

Table 5

Area	Why does it matter?	Starting practice
Water and wastewater	Resource pressure and process continuity	Water balance, leak and reuse studies
Material circularity	Scrap and product lifetime should be considered together	Material passport and appropriate recovery records
Meter / sensor reliability	Bad measurements feed bad models	Tracking calibration, missing data and drift
Supplier data quality	Errors in the chain don't start at the factory gate	Approved supplier data mappings
Accessibility	Technology must be usable by different users	Language, visual clarity and accessible training
Employee participation	The cause of a problem isn't always visible in the records	Suggestion system and safe feedback
Cyber resilience	In connected facilities, the impact of a failure can spread	Network segmentation, controlled connectivity, recovery drills
Model security	Wrong/harmful contributions can corrupt shared outputs	Data provenance checks, testing and version approval
Environmental footprint of models	AI consumes resources too	Tracking energy per task and total benefit
SME access	Those with weak data readiness may be left behind	Shared connectors, consulting and training
Open standards	Single-vendor lock-in makes adoption harder	Portable formats, versioned APIs and an exit plan
Accountability for outcomes	A recommendation is different from a physical action	Authorized decisions, logging and post-incident review

Not all of these areas have to be covered in the first version. For each new data topic, it should be assessed whether the benefit it brings is worth the additional cost and risk.

21. What do contributing companies gain?

For mutual benefit, a "give us your data and you'll benefit later" approach is not enough. A participating company must know what it will have access to, how each contribution will be valued and under what conditions shared results will be used.

Proposed benefits in return could include improving the company's own data quality, fair group benchmarking, access to validated methods, limited shared model services, and training and research support. Contribution credits or discounts on commercial services could be considered; their financial sustainability must be calculated separately.

Contribution should not be measured in gigabytes alone. A small but correctly labeled defect set or a well-documented solution can be more valuable than a large volume of inconsistent records. Scoring rules must be transparent, and gaming the system with wrong or duplicate data must be prevented.

It is also important that SMEs are not left out because they cannot prepare their data. Shared infrastructure and support should reduce initial inequality; large participants' data volume should not automatically determine governance power.

22. Who should govern this structure?

The Data Farm's institutional architecture is as important as its technical architecture. A structure in which a single company can see all other companies without limits will struggle to build trust.

In the proposed governance model, participating factories, SME representatives, employee representatives, OIZs (organized industrial zones) and industry associations, universities and independent experts could take part with different responsibilities. With its experience in MES, automation and data integration, OETECH could take on a role in technical contribution and implementation. The division of roles should be decided by the partnership that is formed.

Collaboration with the Sanayi ve Teknoloji Bakanlığı (Ministry of Industry and Technology), relevant public institutions or other supporting bodies could be considered for this vision. This statement does not imply eligibility for any specific call, that support will be received, or that the project has been adopted by the public sector. Suitable programs and conditions must be researched separately before applying for funding.

The answers to the following questions must be clear in governance: Who processes the data and for what purpose? Who holds the right to use the trained model? What can researchers access? Who reviews outputs before they are published? Who intervenes in case of an error or breach? Under what conditions can a participant leave?

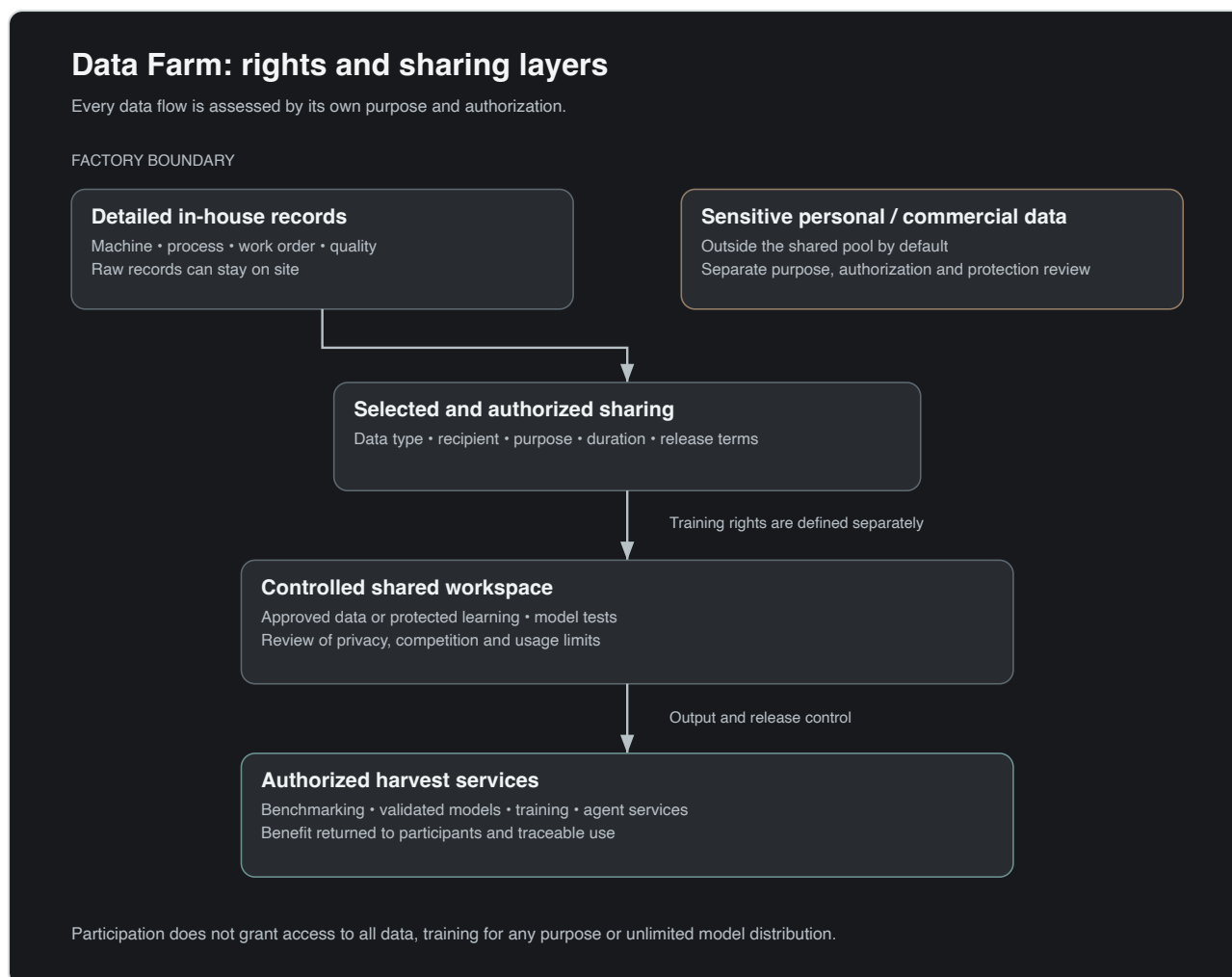


Figure 9. Separate rights for sharing and for model training; shared outputs and publication require their own review.

23. How should the rights to data and models be protected?

Control over data, trade secrets, personal data responsibilities and the intellectual property rights in a model may not be resolved by a single "ownership" sentence. The participation agreement and the technical access structure must be designed together.

A company should be able to track for which purposes its data is processed, to whom it is transferred, how long it is retained and which model versions it has contributed to. Research access, productization and offering a model to third parties can be treated as separate permissions.

Stopping access to raw data upon leaving the system and undoing its influence on a trained model are not the same operation. Fully removing a learned contribution may be technically difficult or, in some designs, infeasible. The policy for retention, retraining and possible removal from models must be clearly defined at the outset; no promise of instant forgetting should be made.

The rights of data subjects must also be protected separately where personal data is concerned. Joining the platform as a company does not in itself make every use of data about employees legitimate. The conditions required for coding and shared use must be assessed according to the type of data. ²

24. How can the budget and roadmap for this dream be approached?

A regional or national Data Farm is not just a project for buying GPUs. It requires data preparation, factory connections, measurement infrastructure, trust, legal work, research, on-site validation and a long-term operating budget.

To think concretely, let us assume **a 24-month regional development program with 12 factories**. The table below is only a budget sketch; it is not a market price, a feasibility result, an OETECH quote or a funding program budget:

Table 6

Hypothetical item	Example allocation
Site connections, measurement and local infrastructure	USD 600,000
Shared data platform and integration development	USD 700,000
Modeling, research and validation work	USD 400,000
Compute, storage and operations	USD 350,000
Security, rights management and independent assessment	USD 250,000
Training, change management and participant support	USD 250,000
Uncertainty / contingency budget	USD 450,000
Total hypothetical program	USD 3,000,000

The actual amount is determined by existing data infrastructure, the tasks selected, staff effort, hardware and procurement conditions. National expansion is not a linear multiplication of this example by the number of factories. Along with central capacity, the burden of local adaptation, representation, support and governance also changes.

The proposed development sequence is as follows:

- 1. Founding preparation:** The shared purpose, governance, rights and the first two problems are defined.
- 2. Narrow pilot:** Data quality and access are validated at a small number of volunteer facilities.
- 3. First harvest:** Measurable results are sought, for example for idle energy and recurring quality problems.
- 4. Regional network:** New facilities and tasks are added with independent assessment.
- 5. National scale:** Rollout expands as standards, service sustainability and participant trust mature.

Rather than waiting for a model that understands all of industry, the first real benefit needs to be shown early. A small pilot does not shrink a big dream; it supports the big investment decision with evidence.

25. How should we measure success?

Terabytes collected, the number of connected factories and model size are important operational indicators, but they are not the ultimate measure of the project's success. The real question is who benefits, and how.

Table 7

Dimension	Outcome that can be tracked
Production	Verified improvement in flow and delivery
Quality	Escapes, rework and recurrence measured against an independent reference
Energy / environment	Condition-adjusted net savings and appropriate emissions accounting
Employees	Access to training, safe work design, rest conditions and participation
Learning	Number of methods and solutions validated at another facility
Market	Participants' verifiable new business or capability visibility
Trust	Unauthorized access, resolution of complaints and exit processes
Inclusiveness	SME access and distribution of benefits
Sustainable operations	Service cost, revenue and long-term maintenance capacity

One factory's savings must not conceal another's losses, and shared results must not be cherry-picked from successful cases only. Models that did not work and recommendations that could not be implemented should also be recorded as part of learning.

26. What this dream means for OETECH

OETECH's work with production data, MES, automation and various business processes provides a foundation that shows how easily data stays fragmented, and how powerful the questions it can open up become once the right context is established.

The Data Farm vision turns this experience into a broader question: Can we go beyond digitalizing a single factory and build an industrial network whose members learn from one another?

In this dream, OETECH's role is not to have all the answers. It is to contribute by building technical solutions, defining real shop-floor problems, working with partners and making sure the knowledge learned returns to businesses as benefit.

The ripe fruit of data is not just more output. Safer work, better use of expertise, less resource waste, more resilient supply and a more livable future are also part of this harvest.

Sources and vision note

The Data Farm name, partnership model, budget and phases in this article are a vision proposal prepared for OETECH. The sources do not indicate that the project has been approved or successfully implemented; they support the related technical, environmental and human-centric principles. Numerical examples are illustrative assumptions.

1. [European Commission — Industry 5.0](#)
2. [KVKK \(Turkish Personal Data Protection Authority\) — Regulation on the Erasure, Destruction or Anonymization of Personal Data](#)
3. [NIST — Privacy Attacks in Federated Learning](#)
4. [Catena-X — About us: industrial data ecosystem](#)
5. [NIST — Digital Twins for Advanced Manufacturing](#)
6. [US Department of Energy — Industrial Efficiency Technologies](#)
7. [US Department of Energy — Process Cooling and HVAC](#)
8. [CDC/NIOSH — Heat Stress: Workplace Recommendations](#)
9. [GHG Protocol — Emissions scopes and categories](#)
10. [GHG Protocol — Corporate Value Chain \(Scope 3\) Standard](#)
11. [WHO — Healthy diet](#)
12. [Rekabet Kurumu \(Turkish Competition Authority\) — Guidelines on Competition Violations in Labor Markets, information exchange section](#)

Sources checked: October 18, 2026. Any real investment, personal data processing, health program, emissions reporting or inter-company information sharing must be assessed on a project-specific basis with experts in the relevant fields.

Related reading: [What Is MES?](#), [What Is OEE?](#), [What Is AI Vision Inspection?](#), [What Is an AI Agent?](#).

Appendix A — Frequently asked questions

A.1 Is the Data Farm a national system that exists today?

This article describes a future vision that Ender Öztürk wants to develop under the OETECH umbrella. There is no claim of an established national network, support received or a completed pilot.

A.2 Why is the project described as utopian?

Because it requires a broader consensus than technology alone: factories, employees, researchers and institutions must build a framework of shared benefit and trust. Even though some of the technical pieces exist, combining them in a sustainable way is a major undertaking.

A.3 Will our raw production data be exposed to competitors?

Not by default in the proposed design. Internal data, contractual sharing and shared output layers are kept separate. Every access is evaluated by its purpose and scope.

A.4 Is coding company and employee names enough?

No. Codes matched with other records can reveal identity or source. Anonymization and trade secret protection must be assessed separately.

A.5 If data stays at the factory, how does the shared model learn?

Some tasks can be carried out with aggregated results; others in a controlled project space or with suitable federated learning. Each method is chosen based on the task and its risks.

A.6 Does federated learning solve all privacy problems?

No. Model updates and the final model can also reveal information. Additional protections and evaluation against realistic attack scenarios are required.

A.7 Can a small company's contribution be valuable?

It can. High-quality labels, well-documented solutions and data from underrepresented processes matter. Value should not be measured by data volume alone.

A.8 Can the same model be used for every industry?

There may be shared components, but materials, processes and acceptance criteria differ. Industry-specific and local validation is required.

A.9 Is one company's successful setting automatically applied to another?

It should not be. Conditions and safety limits may differ. Local assessment and controlled validation come first.

A.10 Does a digital twin require a 3D drawing of the entire factory?

Not for every task. What matters is the behavior and data model relevant to the purpose. Visuals alone do not make a validated digital twin.

A.11 Is the goal to remove people from the factory?

The human-centric side of the vision is to reduce dangerous and repetitive work while developing expertise, training and safe working. Automation decisions are evaluated according to local needs and responsibilities.

A.12 Will employees' health data go into the shared pool?

That is not the default in this proposal. Health and private-life information must be protected separately. Human-centric improvement does not require opening individual health records to the shared model.

A.13 Doesn't tracking breaks and meals create more surveillance?

With the wrong design, it could. The proposed goal is not to track a person's every move, but to improve adequate rest and service conditions. Unnecessary personal monitoring should be kept out of scope.

A.14 If protein intake increases, will output definitely increase?

No such general conclusion can be drawn. Nutrition is assessed as a whole and according to individual needs. If the relationship with production is to be investigated, scientific methods and appropriate expertise are required.

A.15 Does carbon reduction come only from electricity consumption?

No. Emission boundaries, sources and factors must be defined. Reduced electricity use can be an important part, but it does not represent the entire footprint on its own.

A.16 Does reducing HVAC count as energy savings?

If the necessary air quality, process and working conditions are compromised, it should not be counted as success. The goal is to eliminate inefficiency while preserving those conditions.

A.17 Is sharing joint capacity and market information always appropriate?

No. Promoting capabilities is different from sharing competitors' sensitive commercial information. The type of data, how current it is, the participants and how it is used must be reviewed separately.

A.18 Can a participating company leave later?

The conditions for leaving, cutting off access, data retention and managing the contribution in the model must be defined at the outset. It should not be assumed that a trained model will forget instantly.

A.19 If there is public support, should participation be mandatory?

In this vision, the starting point that builds trust is voluntary participation under clear terms. Any future public regulation or program requirement would be determined separately; no claim of mandatory participation is made here.

A.20 What could the first concrete pilot be?

At a limited number of volunteer facilities, two measurable tasks could be selected, such as idle energy and recurring quality problems. Data, rights and acceptance conditions are clarified first.

A.21 When can we say it has succeeded?

When verified benefit returning to participants, trust, employee well-being and net environmental results have been demonstrated. Simply collecting a lot of data or training a large model is not enough.

A.22 Where should those who want to contribute to this dream start?

By jointly defining a shared problem that can be solved, the quality of the available data and the limits of sharing. The first partnership should be built around a real problem, with a measurable first harvest as its goal.



ABOUT THE AUTHOR

Ender Öztürk works on enterprise solutions at OETECH Solutions, focusing on bringing manufacturing execution systems, industrial data and AI applications to the shop floor.

ABOUT OETECH

Founded in 2013 and based in Bursa, OETECH Solutions develops software and hardware for production management, industrial data collection and automation. With IoE+ MES Cloud, HR+ Cloud PDKS and SoSo Inventory, it carries production data from the shop floor to decisions.

© 2026 OETECH Solutions. All rights reserved. Quotation permitted with attribution. ·
oetechsolutions.com