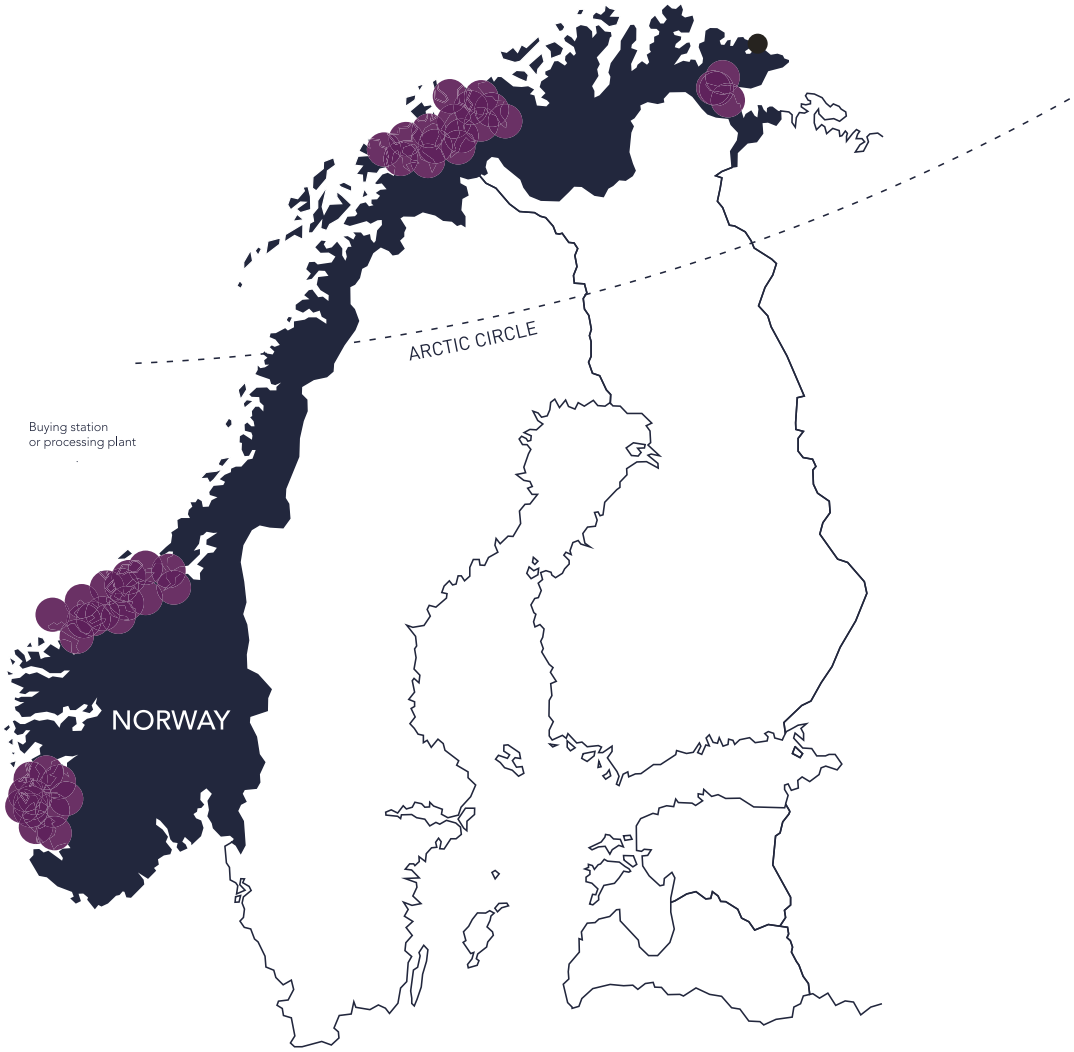
NORWEGIAN SEA FARMERS SENSORS STANDARDIZATION PROJECT

OPEN IOT HUB – FARMER IOT CLOUD

Authors: Tomas Finnøy & Sondre Slathia

CONTENTS



BACKGROUND AND PURPOSE

In today's sea farming and wild fish catching industry, the need and expectations on reporting is rising. As the figure above visualises, actors are spread across a wide geographical area and require information to fundament the right decisions. These reports are most often completed manually by operatives in the field, which can take up precious work hours. With every report requested by the organization, the cost of reporting rises, taking work hours from operatives which could have been used to perform their main, more value-adding work tasks instead.

Another aspect of the business focus today is optimising business processes through standardization and centralisation. This entails integrating systems from multiple suppliers and having them work seamlessly together. Core principles here are "one single source of truth" and "enter data only once, at the source".

These principles shall ensure that multiple versions of the same data in different systems should not exist, and at the same time the data is usable in all relevant systems. Ignoring these principles usually leads to deviations in reports from different systems, which again leads to extra workload to correct the issues, but also potentially incorrect choices being made by operatives and managers based on wrong data.

The situation on the sea farms today is that the data being entered into the systems is often read by operatives from one screen, and manually input into reporting systems. This is completed at a low frequency level, possibly only once per day. Entering data more often would not be cost-efficient, and in many cases not needed for the report levels that the organization is reliant on.

However, the data entered comes from sensors that have far higher capabilities. Data can be sampled down as low as milliseconds in some cases. The amount of data that is possible to register into reporting systems far exceeds the current level, and the possibilities to use that data for other purposes than daily reporting

This document attempts to describe the current status of the architecture deployed to gather sensor data, it's limitations and issues arising from it, and a possible solution.

The value of this sensor data includes, but is not limited to, decrease time for operatives to do reporting through automation, richer data sets for analysis by research projects as well as input for data-driven decision making through dashboards and reporting, enabling alert systems for industry wide cooperation by anomaly detection on current challenges like fighting lice and centralization through remote monitoring of sea farms.

SCOPE

Acquire a system platform that:

- Gathers data from all current and future sensors installed in our sea farms
- Expose data, structured and standardized, to internal and external actors
- Enable system providers to build solutions on top of the platform
- Enable data-driven decision making
- Move LSG towards the data driven management paradigm

VISION

To create a sensor data gathering and distribution platform built on open standards

NORWEGIAN SEA FARMERS SENSORS STANDARDIZATION PROJECT

GOALS

- To build an open IoT platform for sensor technology for use in the sea farming industry
  - Built on open communications standards
  - Support any sensor type and supplier in the market
  - Semantic information models based on OPC UA
  - Completely scalable, both in horizontal and vertical space
  - On-site data gathering and distribution layer – or Data HUB – with remote administration interfaces
  - Cloud solution for storing and distributing data through services – with administration interfaces
  - Enable industry wide cooperation by sharing sensor data across company boundaries
  - Enable machine learning systems suppliers to utilize detailed sensor data to create highly optimized analysis models for use in the industry
  - Establish an industry standard for Sensors and IoT

THE ISSUES

As stated the goal of this project is to build an open IoT platform, based on open standards, for sensor technology in the sea farming industry.

The reasoning for this is as follows.

Sensor technology deployed in the sea farms today are commonly delivered by suppliers to be used to support specific use cases defined by the supplier. This can be an oxygen sensor at certain depth level, connected to a dashboard system placed in the feeding platform to show real time level of oxygen at that depth. Alternatively, a live measurement of the amount feed the feeding system is currently pumping into the cages to be able for an operative to monitor and control by using a camera, which is also a sensor, to watch how much of the feed the fish is eating.

The possibilities of use cases for sensors are endless, so the number of sensors being deployed into a fish cage is growing.

The sensor data is rarely used for anything else than for the specific purpose the supplier itself has use for it, be it in a dashboard they supply with the oxygen sensor, and others bundled with it, or for an operative to monitor the feeding. The data itself is often manually registered into other reporting systems for use by other actors in the organization. This on an aggregate level, mostly daily.

The historic raw sensor data or details are usually stored on-site, if at all, and is not made available for use by other parties than the supplier itself. In some cases, cloud solutions provided by the supplier is offered as a solution.

Below is a description of issues due to the current situation visualised in figure 1 above.

SECURITY AND OWNERSHIP OF DATA

Since the data is stored in non-documented places with no clear security procedures or documented authentication or authorization processes, at least not clear to the author, this data cannot be considered secure. It would be trivial for an untrusted actor with some technical ability and know-how, to extract this data on a visit to a sea farm. Considering most servers and workstations on sites are set up with low security and remote-control solutions like Teamviewer, the user administration is out of the farmers view and control.

The data being sent to and from a supplier's cloud solution are also not documented as being secured or encrypted in any way, at least not to the authors knowledge. In the event of data being sent out of a farmer's infrastructure and on to the open internet and into a supplier's cloud solution, it can easily be extracted from the open data stream by an untrusted actor with technical ability and know-how.

Ownership of data is an important aspect of any data storage system. For the farmer, ownership of data is important to secure. Without it, the potential for data being misused by untrusted actors is considered high. Without data ownership, the potential for its use by the farmer is diminished and at the discretion of the supplier of the system. Today, due the inaccessible nature of the data and its lack of mention in any procurement contracts, the farmer does not have not complete ownership of data.

A security standard for gathering, distributing, securing and giving access to sensor data, where the ownership of data is both contractually and in practice upheld, would solve this issue.

SHARING

The same data which is easy for an untrusted actor to retrieve for nefarious purposes, is equally hard to get hold of for a trusted actor, of with low technical ability, to use for good purposes, such as giving value to the data by exposing it to applications, systems and users that could have benefited from it. The suppliers of these systems currently in use have rarely added good integration solutions to them, and if they do, they are not to the authors knowledge based on official standards.

They are neither implemented using a semantic information model for easy understanding or incorporated into any enterprise or organizational hierarchy which makes it possible to govern the access and extraction of data in a secure and efficient manner.

This leads to an increase in tasks for the operatives to perform, to report, and for moving data from one system to another. The data granularity and quality are severely obstructed, as an operative

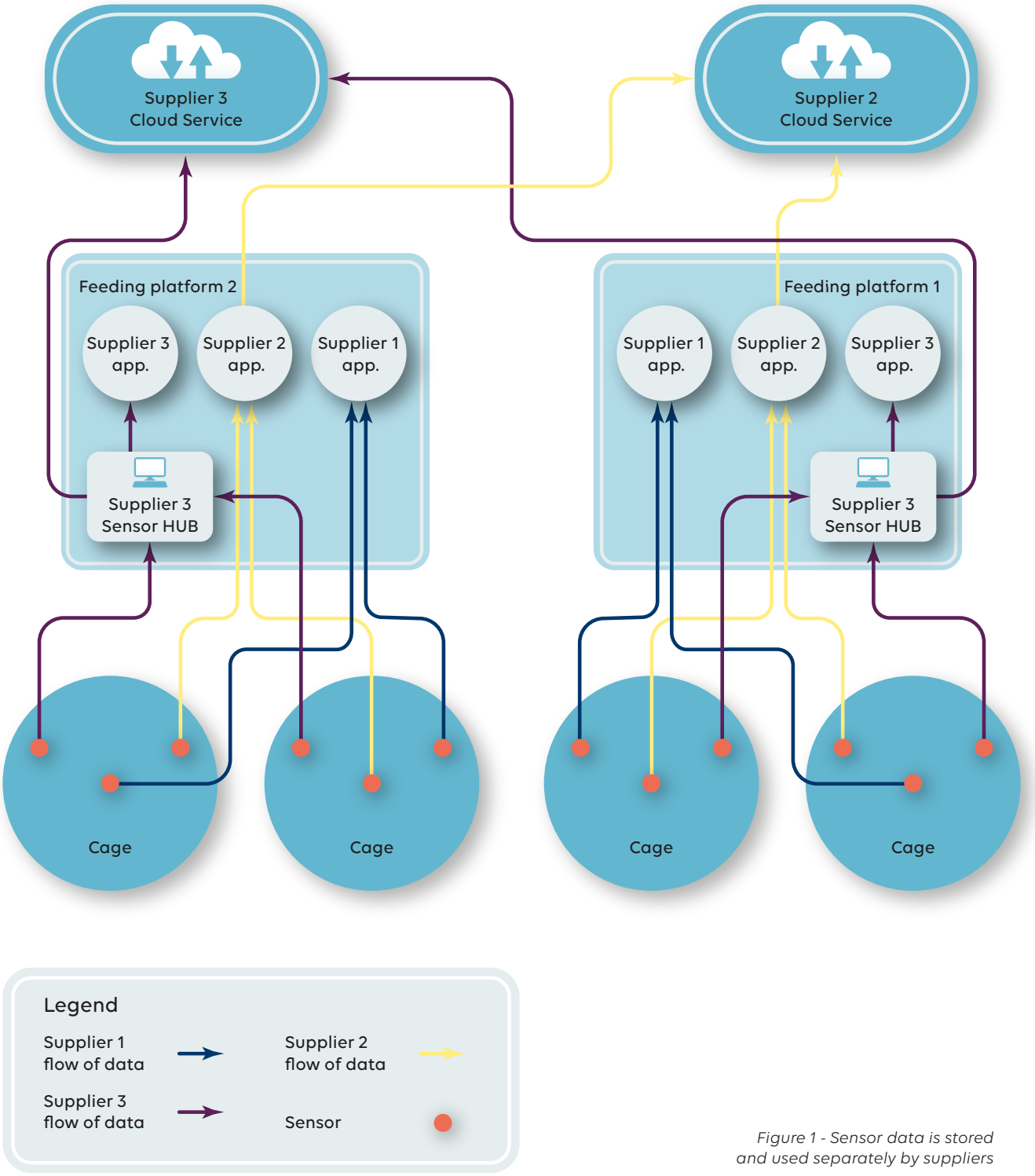


Figure 1 - Sensor data is stored and used separately by suppliers

is at most capable of inputting numbers once a day, whilst the data can be down to a detail level within millisecond precision. Thus, data loss is extensive. The quality of the data is also at risk, as the process needed to perform reporting today, is hindered by introducing the possibility of human error when moving data. To the authors knowledge, there exists little warning to an operative that a sensor has started malfunctioning and giving faulty data, so the actual data quality stored by the supplier's system is unknown.

The lack of any complete central storage solution for the sensor data also hinders analysis using, for example, a data warehouse or machine learning based systems. These systems require well defined and high-quality data to be able to add value to the data. Low granularity data with low data quality, stored in a non-structured way, makes it costly and time-consuming to create any good analysis, which in turn has led to few analysis projects being completed with any measurable success.

A system for sharing data based on open standards, defined by an easily understandable semantic information model and incorporated into the LSGs organizational hierarchy and infrastructure will solve this issue.

VALUE

The actual value of sensor data cannot be measured by its presence and quality only, but by the extent of its utility and adoption by multiple actors. To increase utility the data must to be easily accessible to all trusted actors, modelled in a sensible way that trusted actors natively understand, securely stored so that untrusted actors do not get access, and always available to the trusted actor wherever he is located.

None of these statements are true today, which means that the actual value of the sensor data today, no matter how much there is spread across a farmer's infrastructure, compared to its potential value, is low.

In addition, the potential amount of data this type of architecture can gather and store, and the scalability of the system are also important to secure. The architecture should not be affected by the total throughput of data, and the nodes of the architecture shall be able to handle its subscribers in an efficient way without being affected by the activities of other nodes in the system. In other words, the system should be both scalable horizontally, and vertically. Horizontally by being able to add computational resources to a system pool that distributes the work between nodes in an efficient manner, such as a http-load balancer does for large web sites, and vertically, scaling up the computational power of a single node, or several nodes, and thereby increasing its capacity.

A storage solution that is optimized for time series data is also required to secure the value of the data. A non-optimized storage solution will be encumbered by the sheer volume of data created by the platform and will be delivering data slower and slower over time, as each query will take more and more time to process, decreasing its actual value over time.

A system for extending utility and adoption of sensor data, securely stored, but easily accessible to trusted actors, in a highly scalable system with storage solutions designed to handle time series data will solve this issue.

FISH HEALTH AND WORK SAFETY

Because of the rising need of using sensor data for supporting operative processes, reporting and analysis, many sensors are added to cages where the fish reside to cover specific use cases by the farmers. As there is no, or minimal, cooperation between supplier's systems, the number of sensors rise with each use case implemented. An example is a cage that has a camera set up for feeding purposes, will not, and cannot, be used for other purposes than what the supplier sees fit. If one wanted to use the camera for taking a snapshot of the dead fish in the cage and automatically count them using some automatic image analysis system, the system would either need to be supplied by the fish feeding camera supplier, or a second camera for this purpose would need to be installed in the same cage.

This is, of course, not cost-effective, due to material and maintenance costs. It also carries a large amount of hidden costs with it. These costs can be increased complexity in operative tasks, inputting data from one supplier's system to another and having several systems to operate in, which takes precious time from operatives to do their work, which in turn creates the need for more human resources to perform these tasks.

The risk with working with equipment in the cages increases with the amount of hardware that is installed in it. The risk of damage to equipment and personnel increases with the number of difficult tasks performed by personnel or tasks performed with heavy machinery.

It is also known that the fish in the cages are affected by the number of objects put in cages by inducing stress. Stress is negatively linked to fish health. A system that enables us to minimize the number of sensors put into cages, but at the same time gather the same amount of data needed for operative tasks, reporting and analysis and to secure the fish health and safety of human personnel in LSG will solve this issue.

COOPERATION

The need for industry wide cooperation is present as operations performed at farms in one company can have benefits if they are performed in cooperation with other companies' farms such as monitoring and/or removal of lice. Today the industry is working manually by establishing communication channels through mail and phone calls when they discover anomalies to act on in data in their reports.

The sensor data are rarely used, if ever, automatically between companies, due to the above-mentioned status of the sensor infrastructure implemented today.

One can imagine systems built upon easily shared lice data flowing from cooperating companies to an analysis system that sends out alerts to regions where lice has a growing trend which often leads to the need to remove lice An alert could be sent automatically to operatives in the field of all companies in a region, so they can quickly act on the issue at the same time, without the need for cumbersome manual communication between multiple companies based on mails, telephone or other general communication devices.

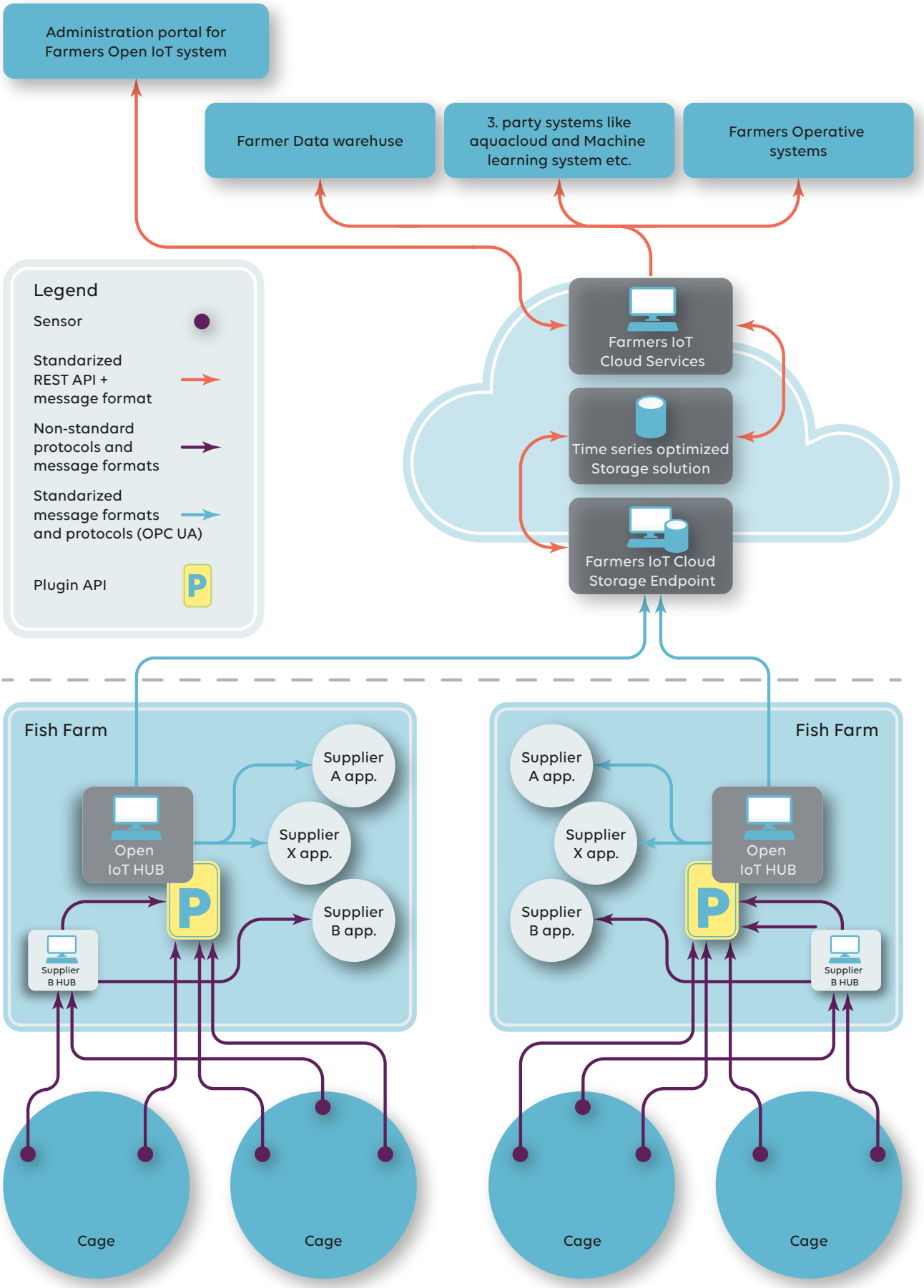
The cooperation could also benefit the entire industry if sensor data from all companies could potentially flow into machine learning systems, at a lower cost and shorter time frame from concept to solution than today. This enables companies that are experts in the machine learning market, and research institutions, to offer industry wide services, instead of having to implement company specific data gathering, data cleaning and data governance to perform the same.

Since the sensor data today is stored in different type of systems, with very different data models, to merge data from several sea farming companies is a monumental task. In most cases its unfeasible to set up a routine for collecting this data, cleaning of errors and transforming them into a usable data model for later analysis. A semantic data model that describes the relation between all data gathering points in the farmer's infrastructure, and the purpose of each data gathering point, be it collecting temperature, oxygen, feeding rate or any other type, is important to secure understanding of the data stored.

To achieve this one must rely on open standards that all companies that want to join in on this type of cooperation, can choose to implement.

A system that enables industry wide cooperation based on open and semantically defined information models and communication standards will solve this issue.

PROPOSED SOLUTION



CLARIFICATION

The proposed solution is to be viewed as what we believe is the optimal solution for the system architecture, design principle and philosophy to support the requirements to the solution described above at the time of writing. It is not to be viewed as the final solution to the problem. The architecture and design might change over time to facilitate new demands or unknown aspects of the problem.

OPEN IOT HUB

The author sees the main data gathering component in the system as publish/subscribe server that passively receives data from, or actively polls the sensors on the sea farms network, stores this data temporarily, and redistributes the data to any subscribers that are active.

The subscribers can be any applications or systems installed on-site, or any online applications or systems when the farm is online.

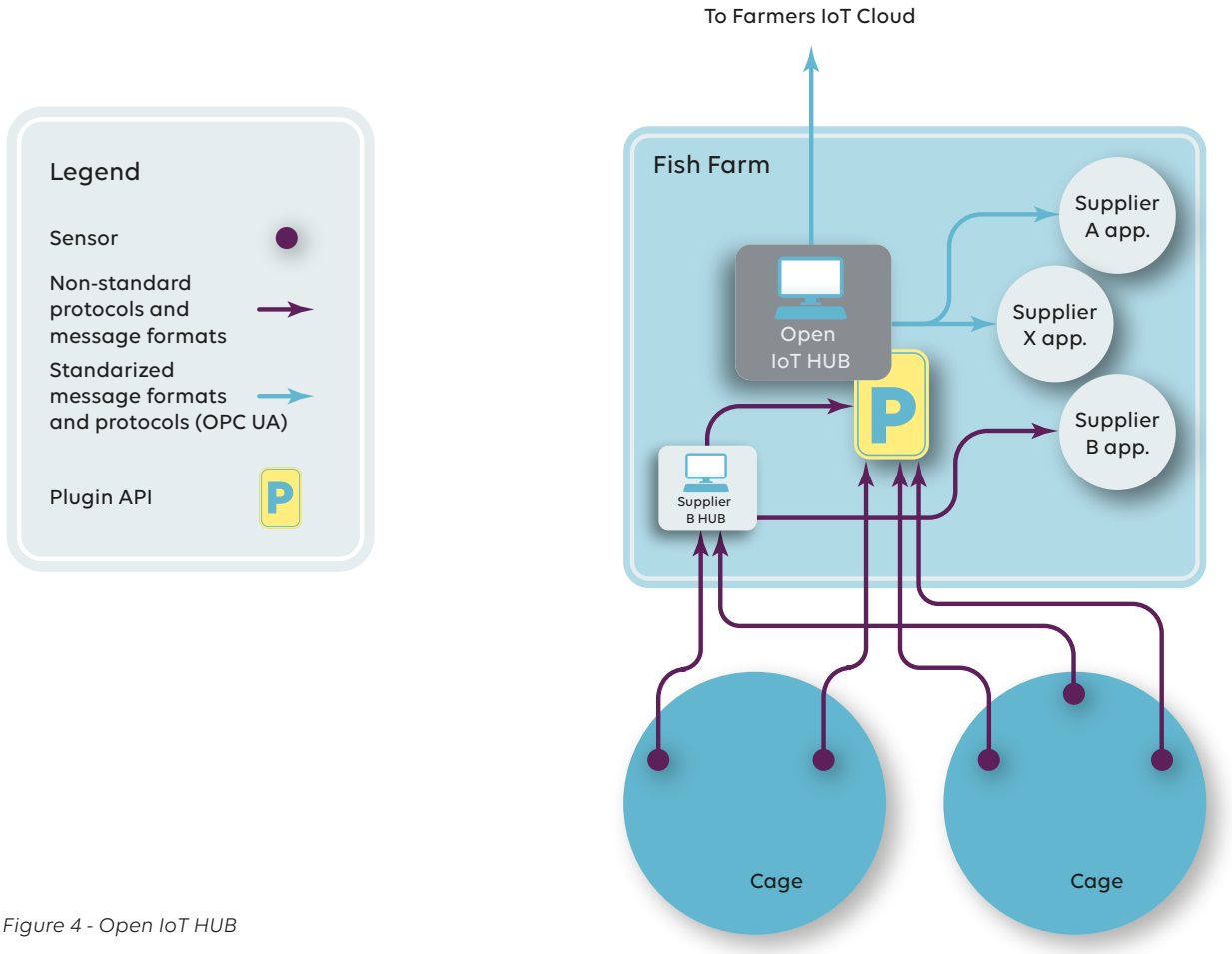


Figure 4 - Open IoT HUB

See figure 4 for a graphical description of the role of Open IoT HUB.

The communication from the Open IoT HUB and to the supplier applications and the farmers cloud solution, and any other system the farmer decides will receive this data, shall be defined by a semantically described data model using OPC UA.

see: https://opcfoundation.org/about/opc-technologies/opc-ua/

The OPC UA standard is in the authors view the most suited communications standard to be used for this type of platform due to its semantic modelling capability, publish/subscribe distribution method and support for the most modern and efficient communications protocols like MQTT.

The data itself shall be stored in a local storage solution as-is. The storage time for the local data shall be at least enough to secure its further storage in an online and more permanent storage solution.

When online the data shall be sent to the next component in the architecture, the Farmer IoT Cloud, complete and continuous, always sent in the same order as was received by the Open IoT HUB.

If at any point the data is lagging, because of bandwidth issues or similar, the data shall still be sent in a time orderly fashion. The layer shall always know how much lag is present in each layer. The transfer of data shall be as efficient as possible, and any solution that has documented efforts to decrease bandwidth usage will be appreciated.

The Open IoT HUB's web services shall be secured using the farmers organizational authentication and authorization system, for example Azure AD. The applications given access shall be able to subscribe to the data messages defined by the applications' needs.

Open IoT HUB shall also support a microservice-based ingress architecture. The task of this is to support a lean way of supporting new equipment by defining a common API interface so that new simple translations of sensors can be easily developed and incorporated by the developer of the Open IoT HUB developers.

For the protocols and message formats not automatically recognized by the common API interface, and not easily developed by the Open IoT HUB developers, a plugin API hosted in the Open IoT HUB that sensor and device suppliers can write plugins for their sensor or devices, shall also be part of the architecture. This would enable a light-weight solution for connecting sensors to the IoT infra-structure and avoid the need to connect through a Supplier HUB. The plugin API architecture shall allow for a wide range of standard protocols and formats, to support the widest range of sensors and suppliers out there. The plugin API must be open and possible to extend by any 3rd party, but every plugin shall be signed with a key from the rights holders to the Open IoT HUB.

The system on-site shall be able to function without constant online status but be able to gather and distribute data even if no internet connection is available for at least 1 month.

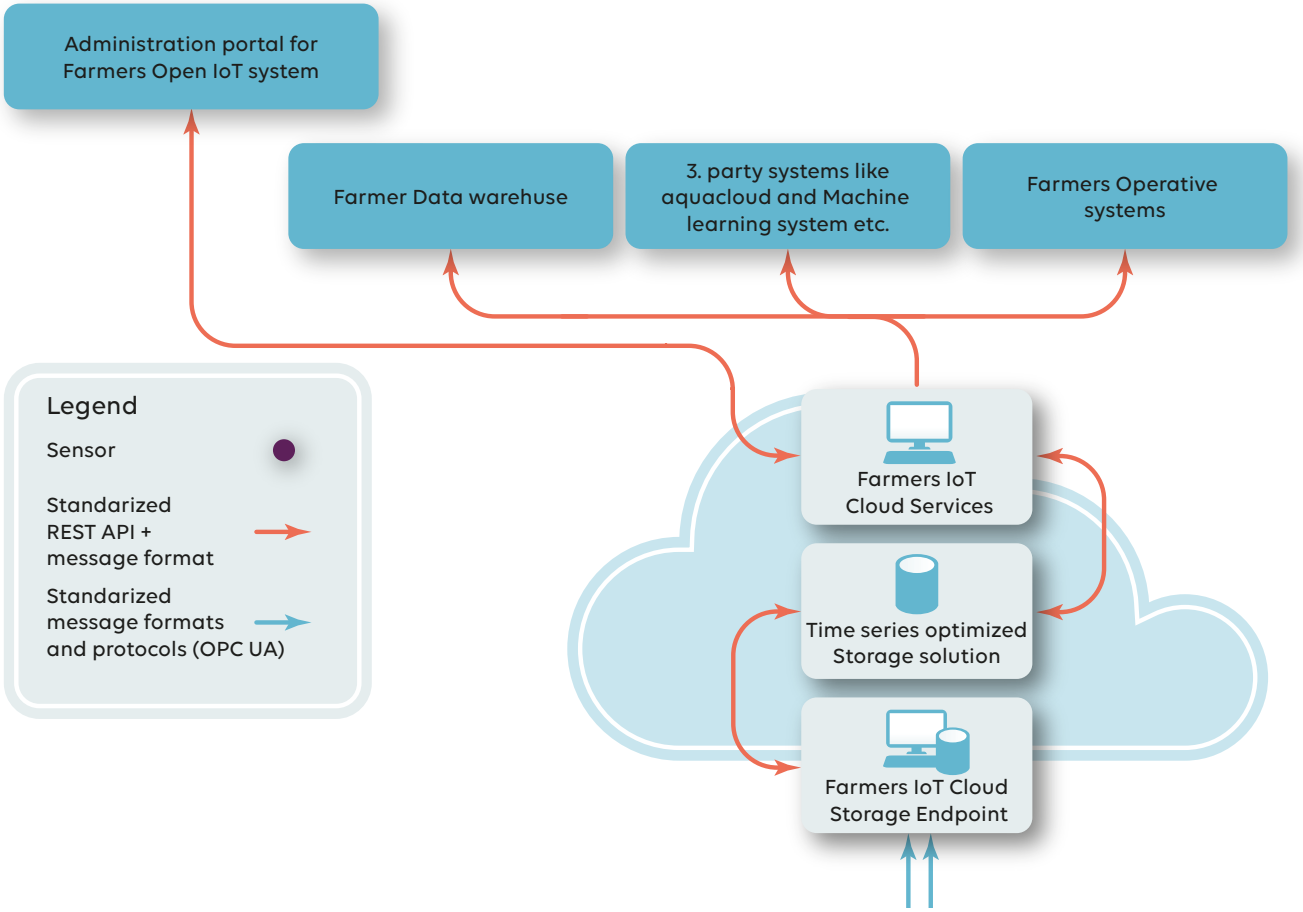
An administration interface to configure the on-site Open IoT HUB is needed in case of service work done offline (no internet).

Main points:

- Receives and distributes live sensor data in a time linear fashion
- Communicates with messages based on a semantic information model using (OPC UA)
- Registry of all sensors on-site
- Passive receiver of data from Supplier HUBs
- Passive receiver of data from sensors
- Active polling agent for data from sensors
- Local temporary storage of local data
- Publish/subscribe infrastructure for distributing data
- Query-able web services for local data with minimal query time
- Distributes data on-site, even if offline
- Sends all sensor data to the Farmer IoT Cloud solution when online
- Plugin API
- Administration interface for setup and maintenance

FARMER IOT CLOUD

The last component in the proposed solution is a cloud storage and distribution solution that the Open IoT HUB continuously streams data to when online. Each sea farming company can have its own cloud solution for its own data, maintained either by the sea farming company itself or a 3rd party chosen by the sea farming company. If the sea farming company already has a cloud solution, the components can be added to that.



The data is stored permanently in a well-structured model based on the same semantic information model defined for the messages flowing between the HUBs and farmers cloud solution.

The cloud solution shall work in many ways like the Open IoT HUB on-site, the difference being it having all the historic sensor data available for its subscribers. A possible way to handle incoming data in Farmer IoT Cloud, could indeed be multiple instances of the Open IoT HUB running in the cloud, writing data to the same storage solution.

Farmer IoT Cloud shall be the point where all parties interested in sensor data from not only one location, but multiple and all locations, shall connect to, and gather data from. This avoids the need for any party having to connect to all the individual sites to gather data but have a single point of contact to get the total input of information.

The cloud solution shall have a register of all the Open IoT HUBs, Supplier HUBs and sensors in the entire architecture.

The cloud solution shall have a monitoring function that keeps track of the timings of the data packets sent from each HUB, comparing sent to received times and if any anomaly occurs and maintenance operator shall be notified.

The cloud solution shall store all historical data by using the semantically defined informational model defined by the project and expose this data with a set of web services which shall be identical to the web services in the Open IoT HUB when it comes to structure, but with any added function-

ality that such a set of services requires. Like with the web services in Open IoT HUB, any actor that has been given access to the data, shall be able to run queries against the services and to set up subscriptions which then gets streamed to the subscriber.

The cloud solution shall be site redundant, globally scalable and be able to deliver as close as possible real time ingress and egress of data. It shall be designed as a multi-tenant architecture with a common data layer allowing data accessibility to be permission based, and not entity based.

STORAGE SOLUTION FOR IOT

The storage solution for such a system will most likely have to be custom-tailored for the IoT data which is defined as Time Series Data. Considering that live streaming data from such a wide range of data types, such as simple number values, images and possibly video, delivered often with extremely small-time intervals, scalable to include all imaginable sensors positioned anywhere in the Farmers system infrastructure; It will not be satisfactory with a proposed solution that does not show any Proof of Concept of its scalability capabilities or can point to other systems with the same or similar requirements that are in production today, either by the supplier itself, or by others.

The storage solution for the cloud shall be time series optimized, highly scalable, storage space-efficient, secure and have fast writes and reads.

Main points:

- Scalable cloud-based storage and distribution component for sensor data
- Communicates with messages based on a semantic information model
- Same message and communication protocol as Open IoT HUB
- Passive receiver of data from Open IoT HUB
- Publish/Subscribe model for live data with minimal delay
- Query-able web services for historic data with minimal query time

SENSOR DATA TYPES

The sensor specific data shall act as a standard for each specific sensor type. This will allow us to define a uniform data structure that can be used by all sensors of same type. The data model must also allow for metadata for the sensors.

- **Oxygen**
  - %
  - mg/l
  - depth
- **Temperature**
  - Celsius
  - Depth
- **Salinity**
  - ppt
  - depth
- **Sea current**
  - Depth
  - Direction
    - Degrees
  - speed
    - cm/s
- **Turbidity**
  - NTU
  - Depth
- **pH**
  - pH
  - Depth
- **Light**
  - Lux
  - Depth
- **Feed**
  - Kg/min

DATA QUALITY

The data quality in the system is critical. The system shall have ways of detecting anomalies in the data, like discovering potential drift, a faulty connection, sensors being installed at wrong depth compared to its metadata and similar situations.

The system shall be able to alert operatives of what sensor has an anomaly, how long it has been present and what the anomaly consists of.

Further scenarios that the system can handle when it comes to data quality, and how the system helps operatives to mitigate difficulties will have to be defined in planning stages of the project.

Any data that has been collected during a period the system thinks is of poor data quality shall be tagged as such, with an indication of how “bad” the data is.

ADMINISTRATION AND INTEROPERABILITY

AUTOMATION

The administration of this system shall be as efficient and automatic as possible. The focus on the project shall be to keep the amount of manual configuration to a minimum. The solution shall be able to handle adding new sources of data in the system in a "Plug & Play" fashion, alert users automatically on anomalies in the network, such as lag, low data quality, faulty configuration, missing sensors, moved sensors and further, will be prioritised.

AN EXAMPLE OF AUTOMATION:

A discovery solution for adding new sensors and HUBs to the system. If a new sensor is added to a sea farm and connected to the farm's computer network. The Open IoT HUB shall discover this new sensor and add it to its registries and then send a signal to the Farmer IoT Cloud that this new sensor is available. An operative will receive a notification that the sensor is available for use and request any needed metadata for it to operate. If the sensor is of an unknown type the system shall notify the maintenance operatives of an unknown sensor type being added to the system.

DATA QUALITY

Anomaly detection on the data flowing in from sensors would also greatly increase the effectiveness of monitoring and administering the system architecture. An automatic detection of sensors behaving badly, such as faulty data, high latency, no data etc., would significantly decrease the amount of manual operations needed.

The system shall contain information on changes to sensor data transactions, similar to what is completed in financial/accounting systems. The possibility of data to be entered into the system that is wrong because of faulty sensor or human errors such as sensor being installed incorrectly, and sensors not being calibrated are known to be likely events. The data shall be able to be modified with correct data. An example would be that it is discovered that a temperature sensor has registered a completely wrong temperature in a cage, and an operative need to change this to a correct value. He can then go into the administration interface and correct a series of registrations he knows is wrong, with a value he knows is right, or close to right, like the data from a temperature sensor from a neighbouring cage. The change shall be handled as a change log, always with a comment for the reason the change has been completed. The original value shall also be available.

METADATA

Besides sensor systems, there will be a need to enrich the data in sensor cloud storage system, and possibly the Open IoT Hub, with data from other systems. This data can be registered maintenance on sensor devices and movement of fish populations between cages and farms. This means that the cloud solution must offer an extendable web service architecture that supports a possibility to add this data to the already existing data over time. And even add this data to historical data later to facilitate deeper analysis.

SCALABILITY

Scalability and responsiveness have high focus. The whole end-to-end architecture shall scale into the future securing the ability to handle hundreds of sites with thousands of sensors per site and satisfy data types of today as well as the future. The solution shall allow for global scale with site redundancy without downtime.

Main points:

- Focus on ease-of use, automation and scalability
- Authentication and authorization access through LSGs own Azure AD/On-site AD platform
- Extendable web service interface for external data
- Change log
- Automatic Component Discovery

SECURITY

The data in such a system will over time represent a substantial value. The risk of this data getting out of hand needs to be mitigated using best practice methods to transfer and store the data in a secure way.

Encryption of data messages, communication channels and the storage solution itself, shall be

considered and if this does not impact the total cost of the system in an extreme way, chosen as a standard way of handling the data.

With the possibility of several clients in the same system, a separation of data between clients is necessary, both for security, but also for easier extraction of a client's total data collection, without needing an engineer to write specific queries to the system.

The focus on security in the system shall be to avoid data coming into an untrusted actor's hands, and the supplier shall deliver a description of its solution on how the system secures the data, both in communication and in storage.

Another aspect of security in the system is to deny any tampering of the data from untrusted actors. With a distributed model detailed in this document, each component carries a risk of this, and the system needs to be protected from this type of attack on every level. The communication between components is paramount here.

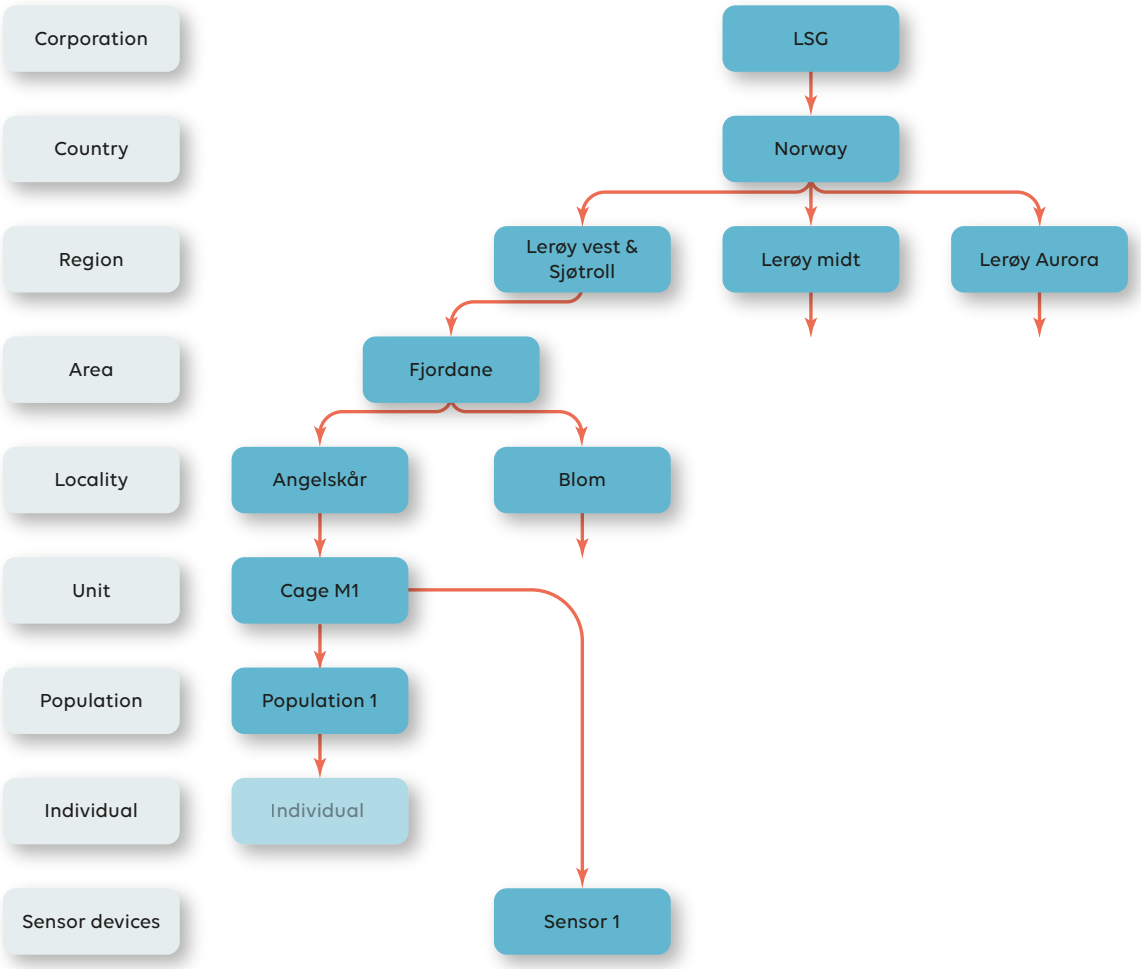
SEMANTIC INFORMATION MODEL AND RELATIONSHIPS BETWEEN SENSOR DATA

The semantic information model which defines how data shall be structured in the messages between HUBs and the storage model is to be defined with OPC UA. The model shall not be considered a final version, as the model will evolve in time as the requirements to the model will also evolve. A natural extension to the model would be to connect it to another model that will be defined for the farmers processing factories or hatcheries for example.

The focus for the OPC UA model mentioned in this document is to give semantic meaning to the sensor data for analysis and business processes. There can be other use cases for OPC UA models which is more aimed at securing good maintenance routines or measuring network limitations and the like.

The types of sensors supported by the system will be defined by the semantic data model. The model also describes how each sensor data type is related to whatever other sensor data that exists in the cage. For example, the data from an oxygen sensor in a cage has less value if the data does not automatically connect this reading with sensor indications of where the fish was placed at the same time. The oxygen level at 5 meters is less relevant to an analyst if the fish were positioned at 50 meters at the time of sampling. The semantic model will handle these cases.

The model will also contain information on where the sensor are placed in relation to the organization map and geographical positon.



The model shows the levels in the hierarchy needed to be modelled into the OPC UA model. It does not show the whole hierarchy, but examples taken from each level that a sea farmer's organisation often consist of. The supplier will have to supply the OPC UA semantic model as part of the delivery in the offer. Each level must be represented. The only level which is not expected to be implemented in the first version of the system, is the "Individual" level.

The model, as mentioned, will need to be extended in the future with additional nodes as the needs of the sea farming industry grows and changes.

OPEN STANDARDS AND ADOPTION

This project aims first and foremost to create a standard that describes how to gather and store sensor data for use by the sea farmers and its partners, as explained above. To secure adoption of the standard by facilitating its use, recommend contractual obligations to parties involved with sensor data and spread information in the industry about the sensor standard and its advantages over proprietary solutions.