A record-keeping approach to managing IoT-data for government agencies

Thomas Soedring[1], Petter Reinholdtsen[2], and David Massey[1]

[1] Oslo Metropolitan University, Oslo, Norway
tsodring@oslomet.no, davm@oslomet.no
[2] Norwegian Unix User Group, Oslo, Norway
pere@hungry.com

Abstract.
Purpose – Particular attention to the issue of information management will be required to meet the expected growth in IoT-devices, and the data they generate. As government agencies start collecting and using such information, they must also deal with the issue of privacy, in order to comply with laws and regulations. The approach discussed here shows that record-keeping principles may form part of a solution to the issue of managing IoT-data for government agencies.
Design/methodology/approach – The study uses the generally accepted record-keeping principles (GARP) as a basis for a high-level discussion on how IoT-data can be managed. This is followed by a presentation and discussion on how one record-keeping standard, the Norwegian Noark standard, can be extended to highlight practical issues.
Research limitations/implications – The main limitations are that the discussion cannot cover all types of IoT-devices, nor can all issues be captured with a limited choice of examples. The results should be seen within the context of the types of devices discussed and limited to the chosen use-cases. However, the level of abstraction used means the results may be applicable to similar scenarios.
Originality/value – The approach shows that record-keeping principles may be used as an approach to manage IoT-data. This discussion is useful when compared with other information science approaches, e.g., big-data or semantic web approaches. The practicalities of a record-keeping approach are also discussed and show relevant strengths and weaknesses.
Keywords Internet of Things, IoT, Record-keeping, Standards
Paper type Research paper

LaTeX-edition of article accepted for publication 2020-02-25 in Records Management Journal (DOI 10.1108/RMJ-09-2019-0050).

1 Introduction

Over recent years the Internet of Things (IoT) has received much attention as devices connected to the Internet play a larger role in both the gathering

and reporting of data. The amount of IoT-data to be generated in the future is expected to see explosive growth (Evans, 2011), in particular, when 5G networks arrive (Akpakwu et al., 2017). Likely, this growth will also see an increase in the amount of IoT-data government agencies collect, in order to provide deeper insights and better input to the governmental decision-making process.

As the amount of collected data grows, so too will the need to manage such data cost-effectively while also remaining in compliance with applicable regulations. In many cases, IoT-data may not directly infringe on a citizen's right to privacy. In other cases, citizen's privacy rights will be challenged, and there will likely be boundary cases. Government agencies may soon find themselves struggling to deal with this issue.

Government agencies will typically have a mandate in law to collect information about citizens. However, such mandates may have been defined at a time where technology limited the amount of information it was possible to collect and perhaps drafted without foresight into what possibilities lie in future technology advancements. The introduction and uptake of cloud computing, along with the increasing development and use of IoT-type devices, give government agencies new tools to carry out their duties. An example can be found in the increased use of Automatic Numberplate Recognition (ANPR) systems that allow government agencies to check many millions of vehicles per year. Previously such capabilities were constrained by manual processes. The scale at which such technology can be used can raise alarm bells when considering privacy issues. This poses the following question: what role can record-keeping play in the management of IoT-data?

The approach presented here is to demonstrate how record-keeping principles are relevant for the management of IoT-data, especially when information that potentially infringes privacy rights is collected. First, the theoretical framework defining the approach for the work is presented. Then related research is discussed. This is followed by a discussion on two use-cases that can give insight into the strengths and weaknesses of a record-keeping approach. The Generally Accepted Record-keeping Principles (GARP) (ARMA, 2014) are used to explore issues relating to the management of IoT-data, before a discussion on the practicalities of how a record-keeping standard can be extended to manage IoT-data. A general discussion on the findings is presented before concluding that record-keeping principles may play an important role in the management of IoT-data.

2 Theoretical framework

The theoretical framework for the work presented here bases itself on the eight generally accepted record-keeping principles (GARP) (ARMA, 2014), as well as a national record-keeping standard Noark 5 (Arkivverket, 2018).

2.1 GARP

The eight GARP principles are Accountability, Transparency, Integrity, Protection, Compliance, Availability, Retention, and Disposition.

Accountability requires oversight at a senior management level to ensure that policies and procedures are in place and to ensure that it is possible to audit the chosen record-keeping approach. Transparency requires that processes are openly and verifiably documented and that such documentation is available to relevant parties. Integrity is a requirement to ensure that records are verifiable as authentic. The protection requirement should ensure that records that require protection have adequate safeguards. For a government agency, this includes any information that is deemed private, confidential, or otherwise classified as a secret. In Norway, records are initially deemed to be publicly available, unless a government agency has a source of law to prevent the publication of information in response to a Freedom of Information (FoI) request. Compliance requires that an organisation adheres to relevant laws and regulations as well as any internal policies. Availability should ensure that it is possible to retrieve information in a timely and efficient manner. Retention ensures that records that require retention, actually are retained, while disposal requires the active deletion of records at appropriate times. A retention example is that a government agency may be required to retain financial records for a certain number of years, while a disposition example can be that a record detailing a DNA sample is deleted once a person is no longer a suspect in a criminal investigation.

Basing a discussion on the GARP principles provides a commonly accepted set of principles that may provide relevant insight into the issue of IoT-data management. However, such a discussion will provide limited insight as the actual implementation of a record-keeping system for IoT-data needs to take into account practical limitations that may arise. As such, further inquiry of record-keeping from a practical point of view is desirable.

2.2 Noark

Norway has a national record-keeping standard called Noark that manages records from the time of their creation to an eventual disposition. Noark stands for **No**rsk **ark**ivstandard, which roughly translates to Norwegian record-keeping and preservation standard and government agencies are mandated by law to undertake record-keeping per the standard. The standard was first developed in the early eighties and has seen multiple updates in an attempt to stay relevant with changes in the law, record-keeping practices, and information technology. It is currently at version 5.5 and consists of both conceptual as well as technical descriptions. An Application Programming Interface (API) is specified (Arkivverket, 2019) for the technical version of Noark 5 (Noark-API).

Noark is also *preservation ready*, in that it specifies standardised XML Schema descriptions that can validate an extraction of the record-keeping structure for preservation purposes. (Hagen Sataslåtten, 2014) notes that a formalised approach to the record-keeping process for government records is particularly useful from a freedom of information perspective as records are more easily identifiable.

The Noark standard proposes an implementation of a fonds. However, a fonds remains an elusive concept, as it may never be possible to identify all the records. While a Noark system may never capture all the records, it attempts to create a space for all potential records belonging to an organisation. (Yeo, 2012) discusses the differences and intersections between collections and fonds, and in some ways, it is possible to identify Noark as fitting both the definitions of a collection and a fonds.

The Noark standard lays forth a metadata framework that manages central metadata relevant both to record-keeping and preservation. The commonly used interpretation is shown in Figure 1. At the highest level, the *Fonds* and *Series* entities typically define an organisational context for records, while the *ClassificationSystem* and *Class* entities typically define a functional context. The *File* and *Record* entities define a transactional context that includes a document context, defined by the DocumentDescription and DocumentObject entities. For documents, integrity and disposition functionality is required. Integrity is supported using checksums covering documents, while disposition can be assigned at multiple levels of aggregation in the record-keeping structure, from Series down to DocumentDescription.

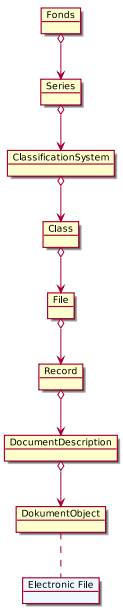


Fig. 1: Commonly used Noark record-keeping structure

2.3 Summary

The GARP principles and the Noark standard form the basis of the theoretical framework for the management of IoT-data as the former can provide a high-level birds-eye view, while the latter can give insight into the practical aspects of such an approach. Record-keeping theory represented with the GARP principles combined with a flexible record-keeping standard that bases itself on relevant international standards can provide a unique insight into how the issue of IoT-data can be managed.

3 Related research

The area of IoT-data management is a complex issue with a very comprehensive depth and breadth. In order to avoid a myriad of technical details, the IoT-data management issue is abstracted and looked at from the perspective of government agencies pondering information management approaches for such data.

The semantic web allows for a significant degree of flexibility and an extendable domain model. The Semantic Sensor Network (Compton et al., 2012) (SSN) ontology plays an important role here, building a foundation that can describe sensors in terms of capabilities, measurement processes, observations, and deployments.

A big-data approach may see the collection of large amounts of IoT-data that offer statistical insights into trends and patterns. An example is how a city may see swathes of people heading for the beach during days with good weather and can plan additional public transport requirements from the weather forecast. However, water quality may be an issue that keeps people from flocking to the beach, even in good weather, and an IoT-device that captures water quality properties may provide additional insight into how a city can predict public transport requirements. A big-data approach to handling large amounts of non-privacy infringing data should not pose problems. However, privacy can quickly become an issue when data about individuals is collected. (Perera et al., 2015) discuss issues relating to big-data for privacy-oriented data.

Perhaps the difference between big-data and linked-data approaches can be found in the quality requirements of the data. A big-data approach is more about statistics and may work, even in the face of poor data quality, while a linked-data approach will likely require a higher degree of quality to be useful. Recent efforts have attempted to merge big-data and linked-data, as a concept called big linked-data (Haque and Hacid, 2014) or blinked-data.

There is also an approach called fog-computing (Bonomi et al., 2012) that processes IoT-data closer to the sensor. Such an approach may allow for a quicker processing of data and subsequent decision regarding retention or disposition.

Within a smart city context (Zanella et al., 2014) discuss application areas and protocols applicable to a general architecture for IoT-data, with details about best-practice guidelines and technical solutions from a proof-of-concept IoT deployment in Padova, Italy. (Jin et al., 2014) present a solution that covers

an entire IoT-stack, from sensors up to the management of data in a cloud-based environment. Neither (Zanella et al., 2014) nor (Jin et al., 2014) discuss the issue of privacy in any detail. (Khan et al., 2014) looks at the issue from a security perspective and proposes end-to-end security for smart-cities applications that use open data and promote citizen participation.

There is no silver bullet to the complex issue of managing IoT-data, whether the approach is based on big-data, linked-data, big-linked-data, fog-computing, smart-cities, or record-keeping. The various approaches have their own strengths and weakness, and it is likely that an approach to manage IoT-data will be based on a coordinated effort taking best practices from multiple information science domains. Privacy is a major IoT research challenge (Sundmaeker et al., 2010), and this creates an argument to examine the issue of IoT-data management from the perspective of record-keeping as privacy is an inherent trait of record-keeping.

3.1 Relationship to international standards

The Noark 5 standard is an interpretation and application of multiple international record-keeping and preservation standards to the way government in Norway functions (Arkivverket, 2018, pages. 9-10). For the record-keeping side of things Noark 5 takes particular inspiration from (ISO-15489-1, 2016), (ISO-16175-2, 2011), (ISO-30300, 2011), (ISO-30301, 2011) and (ISO-23081-1, 2006). MoReq (Fresko and Waldron, 2001) was also an inspiration for the development of Noark 5 when it was initially published in 2008. These standards cover many aspects relating to record-keeping, including process descriptions and relevant metadata. Noark 5 also makes particular use of the information package description defined in the OAIS model (ISO-14721, 2012) and the PREMIS Data Dictionary for Preservation Metadata (PREMIS Working Group et al., 2005). Some of these ISO standards have been updated since, and Noark will likely make use of the updated versions in later versions.

The ISO 15489 definition of a record is based on a description that a record is made up of information that has been retained as evidence of an action. A document, on the other hand, is less formal and can be seen as information in a structured or unstructured format that is editable until the document becomes a record. As such, all records are documents, but not all documents are records.

Noark uses the terms record and document differently than ISO 15489. In the context of a correspondence archive, a record can be seen as a sort of envelope that contains multiple documents. A Noark document is more about a format description along with additional metadata, and its form is often something that reflects a word processing document or email. The record and document terms are not directly interchangeable between ISO 15489 and Noark, but a Noark document will, at some stage, become an ISO 15489 record.

In an international context, record-keeping concepts and approaches will naturally vary, as each country and region have traditions and perspectives of their own. Language and terminology may also play a role when trying to understand various approaches to record-keeping. A review of part of the landscape of international standards related to record-keeping is discussed in (Katuu, 2016).

(Stuart, 2017) provides an interesting insight into barriers to implementing successful records management programs in the context of Australian government agencies. (Wilhelm, 2009) discuss the role of various national European record-keeping standards to MoReq2. This work provides an interesting European perspective on standards like Noark.

3.2 Relationship to other management systems

The terms Electronic Records Management Systems (ERMS), Electronic Document Management Systems (EDMS), Records Management Systems (RMS), and Enterprise Content Management (ECM) have differing interpretations in an international context. (Katuu, 2016) discusses this issue in greater detail. Noark is also explored in that study, but mainly as a public records registry system that captures metadata about correspondence. Noark has much more to offer than merely being a registry of official correspondence. In particular, the API part of Noark 5 shows a high-level of flexibility in terms of extensibility and requiring the use of the OData Open Data Protocol (ISO-20802-1, 2016) to facilitate search capabilities.

Even though Noark is perceived to manage (word-processing type) documents, it is a records management system. The standard opens to capture and preserve records of multiple formats, including, e.g., instant messages. So the traditional document concept does not define what Noark can potentially manage.

The Noark standard and its flexible metadata model is used as a basis for studying the practical side of integrating record-keeping and IoT-data, in order to shed light on potential possibilities and any limitations that may arise.

4 Case studies

IoT-data is considered to be data generated by an internet-connected device controlled from a remote location. The definition is broad and adequate to cover the two chosen use-cases. Water quality readings from the open data portal belonging to the City of Chicago define the first use-case. The second use case is limited to images and associated metadata captured by an Automatic Number-plate Recognition (ANPR) system managed by the Norwegian Customs Agency that is used to identify cars crossing the national border. The former is an example of data that should be unproblematic to collect, while the latter has severe privacy implications that deserve particular attention. The use of two differing metadata domains exemplifies a range of issues when dealing with IoT-data and how record-keeping approaches can be used to manage IoT-data, by considering such data as records.

4.1 Open data case

The first use-case comes from the City of Chicago's open data portal (City of Chicago, 2020*b*), where the Chicago Park District publishes water quality

Attribute	*Value*
beach_name	*Ohio Street Beach*
wave_height	*0.115*
measurement_timestamp	*2018-08-27T10:00:00*
water_temperature	*22.7*
turbidity	*2.34*
measurement_id	*OhioStreetBeach201808271000*
wave_period	*4*
battery_life	*9.4*

Table 1: Example IoT-data describing water quality published by the City of Chicago

readings (City of Chicago, 2020*a*) collected from six beaches located along the lakefront of Lake Michigan. An example of one of the hourly readings is displayed in Table 1.

This data has no privacy concerns as it is related to observations of physical properties of water, bound to a given set of locations. Three categories of data are collected: observation, administration, and device status. The observed properties are wave height, wave period, water temperature, and turbidity (i.e. cloudiness). These observations are augmented with administrative information: location, measurement_timestamp, a unique identifier for the reading (based on location and measurement_timestamp), and a value indicating the battery status of the IoT-device.

4.2 ANPR case

The second use-case is one that has known privacy issues. The Norwegian Customs Agency was fined NOK 900,000 (over 100,000 USD) by the Norwegian Data Protection Authority in March 2019, for inadequately storing data about border crossings and infringing on citizens' right to privacy (Thon and Kaspersen, 2019). The data was collected using an ANPR-system from both fixed and mobile cameras. The Norwegian Data Protection Authority noted that the Norwegian Customs Agency had:

- inadequate access control to a shared database with the Norwegian Public Roads Administration
- monitored 80 million border crossings affecting an estimated 7-8 million people
- a lack of internal control and no documentation about information security concerning the ANPR-system

The Norwegian Data Protection Authority noted that this was a significant privacy infringement that affected a large number of people over an extended

Attribute	*Value*
number plate	*EL09164*
location_fixed	*Svinesund*
location_mobile	*59.0990°N,11.2696°W*
timestamp	*2019-08-27T22:03:24+0100*

Table 2: Example ANPR-data

period and that some of the data that was stored by the Norwegian Customs Agency for six months, was done so, without the necessary authority. The Norwegian Customs Agency had access to a database managed by the Norwegian Public Roads Administration and tested their system using observations from the ANPR-system belonging to the Norwegian Public Roads Administration. The Norwegian Data Protection Authority argued that the Norwegian Customs Agency did not have the necessary authority to store information from the other ANPR system.

Unfortunately, it was not possible to retrieve an official metadata list covering the ANPR-system from the Norwegian Customs Agency, but an informal discussion with the Norwegian Public Roads Administration that has a similar system revealed the following description of their ANPR-system.

The ANPR-system observes a car by taking two pictures. The first is an overview picture showing the car; the second is an infrared-picture of the number plate that is subsequently processed and interpreted using an optical character recognition (OCR) algorithm. The images, OCR recognised text, the date and time information, location, and an identifier of the camera that took the picture are then stored. Stationary cameras record the name of the location, while mobile cameras record the location as GPS coordinates. The observation is deleted within an hour if the car is not subject to an inspection. Table 2 shows the attributes captured during an observation.

The Norwegian Data Protection Authority noted that the occupants of the car were visible in the image, and their faces were not censored. The identification of a vehicle also resulted in the collection of details about the registered owner of the vehicle. As such, the Norwegian Customs Agency collected a significant amount of private information.

5 Applying the GARP principles to IoT data

Applying the GARP principles to the use-cases discussed earlier can serve as a basis to explore the relevancy of record-keeping principles when managing IoT-data.

5.1 Accountability principle

The open data portal does follow the accountability principles, as the portal was required to be established under an Executive Order by the mayor of Chicago in 2012, where the role of chief data officer reporting to a chief information officer is also defined. When it comes to the ANPR-case, there may be issues regarding accountability, based on the fact that the Norwegian Data Protection Authority stated that the Norwegian Customs Agency had inadequate access control to a shared database, there was a lack of internal control, and they had no documentation about information security about the ANPR-system.

5.2 Transparency principle

The open data case also appears to meet the transparency principle as data are managed within an open portal and made available to anybody visiting the portal. Descriptions of data sets are available, along with information describing how data are collected and how data sets are anonymised to avoid potential privacy issues.

The Norwegian Customs Agency publish limited information about the ANPR-system, but this may likely be to avoid giving away too much operational information so that criminals can try and get the upper hand over the border patrols when trying to cross the border undetected. The Norwegian Customs Agency disagree with part of the description from the Norwegian Data Protection authority and state that they have an ongoing project that covers privacy issues and information security. However, it is not possible to determine publicly, what level of transparency the ANPR-system follows.

5.3 Integrity principle

Dealing with the integrity of IoT-data may be a complex issue from a record-keeping perspective. The definition of an IoT-device is broad and can be anything from a cheap throw-away device to an expensive camera. IoT-devices can be owned by a government agency, private citizens, or even organisations, and this may challenge the perceived quality of the data stored as records. An example of poor-quality records could be a relatively cheap air sensor hosted by a private citizen on their building, reporting incorrect air quality values as the device is malfunctioning or not correctly configured. Another example might be the injection of false sensor data, or manipulation of the sensor itself or its location by people interested in manipulating the measurements. Problems with individual sensors may not necessarily be a problem as an overall statistical trend from a collection of sensors may be the factor of interest.

Traditionally, the record-keeping process is driven by humans and subjected to a form of quality control, which is something that helps define a certain level of authenticity for records. IoT-devices change the nature of the type of records that are collected, and a record-keeping system will need to take this into account.

The Noark standard treats the integrity issue from two perspectives. First, it requires that the protection principle is adhered to (see next section), but also requires the use of a checksum covering uploaded documents. The open data case has a limited need for integrity due to the open nature of the collected data. The ANPR-case, however, needs to ensure integrity mechanisms are in place as the data that is collected may be used in criminal proceedings at a later date.

5.4 Protection principle

The protection principle is an issue that is very broad in an IoT-context. Protection may include physically securing each IoT-device to ensure it has not been tampered with or is leaking data, but also to ensure the data transmitted over the network is kept private.

Protection, in a Noark perspective, is handled by a clear definition of users, groups, and roles in the record-keeping system and that an access control mechanism is in place. Further, there are requirements to ensure only users with the correct permissions can retrieve information, and there is an ability to grade material at various national security levels.

For the open data case, the protection principle is not relevant, as the collected data has no privacy or confidentiality implications. On the other hand, the ANPR-case had uncertainty regarding clear definitions of who had access to data, and there was a lack of documentation regarding security.

There may be a need to ensure that IoT-devices have limited access to the record-keeping system. An eventual API may need to ensure that IoT-devices can only push a defined set of information to the record-keeping system at predefined intervals. It is also possible to poll the IoT-devices for information at regular intervals. From an information security perspective, IoT-devices may quickly be hacked or swapped out with a device that tries to break into the record-keeping system, and as such, have limited trust.

The Noark standard has no particular mechanism concerning this issue. Instead, this is an issue that needs further analysis.

5.5 Availability

When considering availability in an IoT context, it can be viewed from two different angles. The first is that there may be a desire to combat the inherently temporal nature of IoT-data by capturing and managing the data, while the second is that IoT-data must be made available for consumers.

The water quality readings are temporal as, without a collection process that records the information every hour, the information will be lost. The open data portal makes water quality readings available to consumers in both machine-readable and human-readable formats.

The retrieval aspect is also vital for IoT-data as it does not make sense to capture IoT-data without processing it for a reason or making it available for data consumers. It is essential that a search interface for IoT-data is easy to

use and is readily adaptable to new metadata descriptions. The OData (ISO-20802-1, 2016) query specification is a protocol that can be adapted to support evolving metadata models. The open data portal already supports OData for retrieval, as does the Noark-API and, as such, it is possible to base retrieval for the ANPR-case on the OData protocol. However, the ANPR-case may come with additional geospatial data retrieval requirements, allowing for searches within a set of bounded regions to search based on GPS coordinates. The Noark-API has limited support for a region-bounded retrieval mechanism for GPS-coordinates using OData, and the issue needs further research.

5.6 Compliance

Lawmaking does not always manage to keep pace with advances in technology, and perhaps we first need to see the actual use of new technology before we are in a position to understand how new technolgy may infringe on our rights. Privacy in the modern cloud world is a sound example of such a change and how laws such as the GDPR (European Commission, 2016) play catch-up in order to maintain privacy expectations.

The ANPR-case illustrates this, as changes in technology afforded customs agencies the ability to undertake high volume observations of border crossings automatically. The collection of such data by the Norwegian Customs Agency was only recently authorised through changes in the law, and perhaps part of the misunderstanding in the ANPR-case can be attributed to a misinterpretation of the law.

Government agencies collecting IoT-data will have to ensure they comply with laws and regulations.

5.7 Retention

The retention of non-privacy IoT-data is likely a question of resources; the open data portal, for example, has water records going back to 2013. There is likely no formal requirement to store such records, but they are a new type of information that new tools and technology allow us to collect quickly. Perhaps, such data will play an important role in the future, but we may currently fail to see their long-term value.

The issue of retention is extensive, however. For citizens, retention can be vital as it can document various rights citizens have. Pension rights may be a challenging issue for a citizen if the records showing employment for the previous 40 years become missing. In particular, the transition from paper-based record-keeping to electronic record-keeping may have some surprises, where data may be lost.

Record-keeping, in non-formal record-keeping systems, can sometimes have a here-and-now approach, with little regard to the needs of future users. In Norway, for example, a general auditor report from 2010 (Kosmo, 2010) noted a severe loss of records from various electronic systems over previous decades.

5.8 Disposition

The disposition concept for records is an essential functionality a record-keeping system must support. IoT-data is a relatively new phenomenon, and the requirement for retention and deletion of such records may perhaps be a little unclear.

In Europe, with the GDPR legislation in force across the EU/EEA, there is a requirement to ensure that it is possible to locate, retrieve, and delete personal information upon request (Kelly et al., 2019). There may be other laws preventing the deletion of records, for example, a person subject to investigation for smuggling cannot expect that the information the customs agency has collected, while building a criminal case, can be deleted based on a GDPR request.

The ANPR-case clearly shows the need for retention and deletion functionality. The Norwegian Data Protection Authority states that the Norwegian Customs Agency failed to dispose of records in accordance with laws and regulations. A detailed description of the underlying problem is not publicly available, but it is clear from the fine (Thon and Kaspersen, 2019), that there is an information management problem.

6 Extending the Noark record-keeping standard for IoT

It may be possible to develop a centralised and standardised approach to the management of IoT-data. However, many a record-keeping practitioner is aware of the heterogeneous and elusive nature of records, and applying a one-size-fits-all approach to record-keeping has never really been successful. A record can, for example, be an email with associated information in a spreadsheet, along with a word processing document stored as a calendar entry in the organisation's calendar system. Aggregating such information to a single record can be a daunting task and may, at a minimum, require human intervention to ensure success. IoT-data may have similar properties with regards to heterogeneity and elusiveness, but it is worth arguing for standards-based approaches to manage IoT-data. Standards increase interoperability and can help avoid potential vendor lock-in situations (Simon, 2005). But standards are only [truly] open when they can be implemented without fear as free software in an open source community, as observed in (Phipps, 2007).

It is possible to base an approach for IoT-data on a record-keeping standard that implements a flexible API that can handle a high level of data heterogeneity. Any IoT-data processed by the API becomes available as records, retrievable via the search interface of the API. Such a standard would require the following:

- a flexible and extensible metadata model
- an extensible preservation metadata model
- a standardised search protocol
- retention and disposition support

To a large degree, the Noark 5 standard and, in particular, its API description, meet the above requirements. The decision to pursue a solution by extending the Noark-API is taken as it is then possible to see what kind of particular

issues arise. Even though the metadata model in the Noark 5 standard is flexible, there are limits to how far the standard can be extended without breaking compliance. In some instances, it is possible to extend entities (database tables), while in other instances already defined attributes (database columns) can be reused, e.g., the title attribute. The Noark-5 API also supports the use of custom metadata but, for the sake of brevity, it is not discussed here.

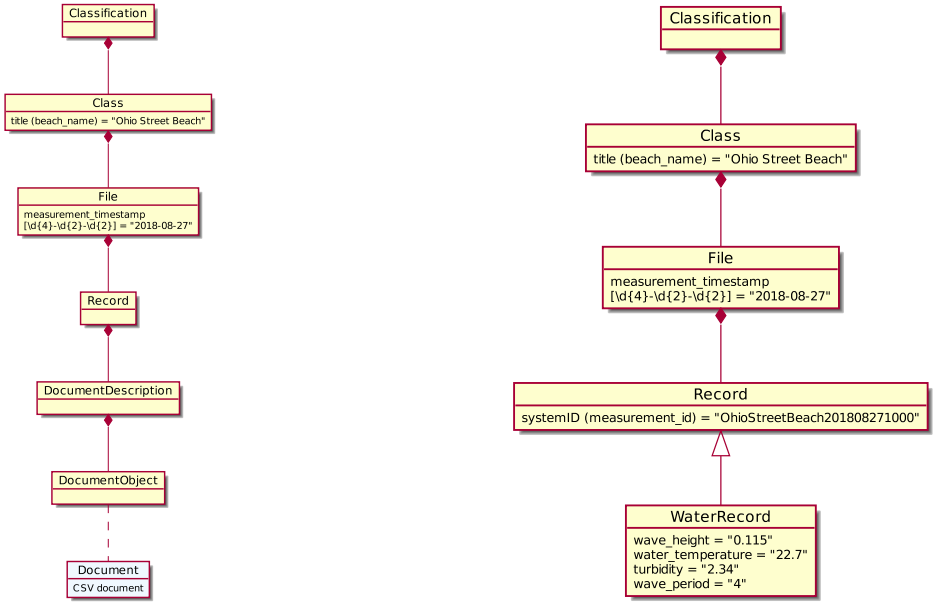
6.1 Extending for open data

Here, there are two approaches to extending Noark for water quality readings. The first is a *document-approach* where readings are ingested as documents (e.g., in CSV or JSON format). The second is a *record-approach* where the metadata model is adapted to extend entities and to add relevant attributes, recording the observed water quality values as records. Both approaches place water quality readings correctly within the record-keeping structure, something that is also important when migrating the data for preservation purposes. The record-approach has the added benefit of increased ease of retrieval, as data is directly searchable from the API, e.g., to search a range of water quality records based on observation timestamps. The Noark-API, however, does not describe search mechanisms for the content of documents, so there are additional limitations to the document-approach.

Figure 1 shows a typical implementation of the Noark metadata model and serves as a reference for the two approaches. The document-approach covering water readings is shown in Figure 2a. The beach name is assigned to the title attribute of the Class entity as the primary record aggregator for water quality readings. In this case, there are six classes, one for each beach. Records are further aggregated within the File entity, as shown in Figure 2a as a regular expression, aggregating readings daily. Readings are then stored as an electronic document using an appropriate file format, e.g., CSV.

Figure 2b shows a record-approach for water quality readings. The values of a reading become attributes across various entities. A similar approach to the document-approach is used here, aggregating records based on Class and File. In the record-approach, however, the generic Record entity is extended to be a *WaterRecord* with the addition of the water quality attributes *wave_height*, *water_temperature*, *turbidity*, and *wave_period*. This approach allows users to search for water quality records directly via the API. The following exemplifies a relevant OData query that retrieves all water records where the turbidity value is greater than 2, and the water temperature is greater than 22:

```
WaterRecord?\$filter=turbidity gt '2' and
  water\_temperature gt '22'
```



(a) Document-approach (b) Record-approach

Fig. 2: Two approaches to extending the Noark 5 metadata model for water records

6.2 Extending for ANPR-data

Figure 3 shows an extension of the Noark metadata model for the ANPR-case. The license plate number is assigned to the title attribute of the Class entity as the primary record aggregator for ANPR-data. Every time a car is observed crossing the border, the observation is stored using an extension of the record entity called *CrossingRecord*, which includes the *location* and *timestamp* attributes. The associated captured images are stored with relevant document metadata.

When aggregating records, the license plate number is the primary object of interest, and the proposed solution assigns the location information as attributes in the *CrossingRecord*. However, it is also possible to store such information as a secondary *Class*. Thus there are multiple ways to interpret and implement such a scenario in Noark. It is also possible to search this type of data using an OData query; for example the following retrieves every record of a car observed at the Svinesund fixed border crossing:

```
Class?\$expand=CrossingRecord(
\$filter= location\_fixed eq 'Svinesund')
```

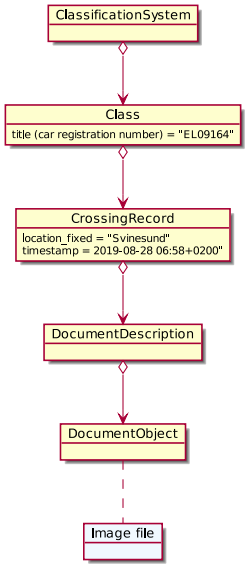


Fig. 3: Extending the Noark metadata model for the ANPR-case

The Norwegian Data Protection Authority noted that disposition and access control were two issues prevalent in the ANPR-case. Using a standardised approach, as proposed here, means that disposition rules could automatically be associated with records, ensuring the deletion of records at the appropriate time. If there are requirements to store some records for a more extended period, individual records can have the disposition rule removed.

An appropriate access control mechanism could further alleviate the criticism by the Norwegian Data Protection Authority, as information security is a requirement to achieve the protection and integrity principles.

7 Discussion

"The more connected a city becomes, the more it will generate a steady stream of data from and about its citizens" (Finch and Tene, 2018). The gathering of any steady stream of data should not be incidental; rather, it must be targeted and in compliance with relevant laws and regulations to ensure that the right to privacy is respected.

The Norwegian Customs Agency case details how the organisation uses technology to significantly increase border-crossing observations. This potentially infringes on the right to privacy and any data collected must be managed appropriately. The Norwegian Data Authority noted that the Norwegian Customs Agency was not in compliance with laws and regulations and infringed the right to privacy for a large number of individuals. This infringement came at a price and shows the need for a coordinated approach to managing IoT-data.

Including IoT-records as part of an organisation's formal record-keeping strategy may ensure an appropriate allocation of responsibility and descriptions regarding the treatment of IoT-data may be published. It is possible to pay particular attention to privacy issues once the handling of IoT-data comes under the responsibility of an information manager. A chief information officer is assigned responsibility in the open data portal, showing accountability principles. The ANPR-case illustrates that there are consequences when an organisation does not manage its records per laws and regulations. Treating IoT-data as records and managing them within a record-keeping system may enable a government agency to easier show that they meet the accountability principles. Transparency should lead to the availability of documentation appropriately describing the management of data. Both use-cases show a differing way that transparency manifests itself. The open data portal openly divulges information on the management of data, while the Customs Agency likely needs to keep such information confidential as it is operational information detailing how a law enforcement agency works.

Depending on the scenario, the perceived authenticity of IoT-data may be increased if it is stored in a record-keeping system as the IoT-data becomes associated with structural and contextual information linking captured records to the organisation and processes that use the data. Noark supports other integrity mechanisms as well. For example, checksums can be generated that cover uploaded documents, and any data derived from an IoT-device that becomes documents, including images, can also be covered by such fixity information.

The protection principle, as defined within the GARP (ARMA, 2014), is perhaps the most persuasive argument for a centralised approach for the management of IoT-data. Multiple systems, with multiple logins and access control models, may make the task of protecting information difficult for an information manager. A distributed approach may also increase the workload for an organisation, challenging the GARP principles of accountability and transparency. Interoperability and the need for government agencies to exchange data also becomes a challenging issue. It is likely there will be more sharing of information in the future as government agencies increase the use of collecting IoT-data, where data is cross-referenced against data managed by other government agencies. Such sharing of data should come with retention/deletion requirements that ensure an automatic disposition of data at the correct time.

It is easier to prove compliance with laws and regulations if an organisation has documented information management procedures and personnel that specifically work with issues of compliance. Compliance that can be shown by following a standard or standardised practices may more easily show organisational compliance with laws and regulations. IoT-data may represent a new class of data and records within a record-keeping system. IoT-devices are likely to have varying properties and requirements, resulting in potentially large volumes of temporal and non-temporal data from devices that perhaps have intermittent communication ability. If the volume of similar records is too large, the process of searching may be impeded, affecting the principle of availability and retrieval.

The open data portal manages multiple heterogeneous data sources in a way that does not impede retrieval. Metadata modelling, as used in record-keeping approaches, for example the use of the entities *Series*, *Class*, *File*, and *Record*, can offer similar approaches, with a high degree of automation for the collection and management of IoT-data. However, this will require an initial effort to define the metadata model.

Availability will require that the IoT record-keeping system is implemented using a modern REST-based API approach. The system will need to be able to handle high volumes of data with timely retrieval of records, as well as supporting a flexible approach to search that can quickly evolve as various domains of IoT-data come under its management. OData (ISO-20802-1, 2016) is an excellent candidate to handle the search mechanism but other protocols exist, e.g., GraphQL (Wittern et al., 2018), or vendors may implement a proprietary domain-specific language for querying. A preferred approach is to reuse existing standards as far as possible.

The City of Chicago open data portal has some properties that point to a best practice approach for managing IoT-data. Data are well managed with excellent search capabilities, and the portal supports the ability to publish data in multiple formats (e.g., CSV, XML, JSON).

The expansive growth of IoT-records will require clearly defined disposition functionality. Some data may have little value as individual records, as the observed trend in a data series is perhaps a more important factor, while other data might have no historical value. Temporal data with little value may be deleted unless there are other reasons to retain them. Data governed by privacy laws may require automatic disposition, and the record-keeping system must actively seek out and dispose of records at the correct intervals.

The case with the Norwegian Customs Agency shows the other extreme of managing IoT-data, as they were fined for not disposing data at required intervals. Even though the intricacies of the ANPR-case are not publicly known, it can be argued that the ANPR-system was not based on, or perhaps not following, sound record-keeping principles. Nor was it likely that the organisational accountability and transparency mechanisms were in place to ensure compliance with laws and regulations.

Disposition is a core record-keeping functionality and is a clearly defined functionality within the Noark standard. However, the functionality is limited to documents, rather than records (database rows). Disposition will ensure the deletion of a document occurs, but it will also preserve information about the document's existence and the time the deletion occurred. Whereas records, within the record-keeping structure, that makes up the structural and contextual bindings between information are deleted, without precise information about the existence of the record. It is possible to log a deletion event, but ultimately the deletion of (database) records in Noark is a different concept when compared to the disposition of documents.

It is not clear from the open data portal, how long records are retained before they are discarded, but modern storage requirement costs are, for all intents and purposes, negligible so records can easily be retained.

Retention ensures that records are available for the appropriate amount of time, according to various obligations. Retention will then also cover the preservation aspect and can enable a life cycle approach to record-keeping. The open data portal provides records in preservation friendly formats, and even allows users the ability to download entire data sets. There is also additional provenance metadata available as well, for example data owner.

The ANPR-case gives the impression that most records are deleted after a short time. Perhaps number plates that belong to cars that are known objects-of-interest, and that should be subject to control, need to be retained. It is unclear if such records should be migrated for long term preservation reasons at some stage.

The Noark standard defines a preservation metadata model with associated XSD schema descriptions that are extensible and can cover the two IoT-data cases here.

A classic decision a government agency may face when dealing with IoT-systems is whether or not to force the collection and management of such data to a centralised portal or to use multiple vendors' infrastructure. In today's cloud-based world it is attractive to let a third-party vendor manage the data within the vendors' infrastructure, but that might not be the best approach in light of the GARP principles. If the anticipated explosive growth in IoT-devices occurs, it may quickly become a nightmare for government agencies to deal with a myriad of software systems and vendors. The accountability and transparency principles of GARP may quickly be challenged. More importantly, each system that collects data about a person will require proof that it is capable of meeting disposition requirements and that it is possible to locate and retrieve records about citizens, in order to comply with the GDPR. There may be a significant cost implication, as each system will need to implement record-keeping functionality.

Building an IoT-portal based on sound record-keeping principles to deal with the explosive growth of IoT-data may help government agencies to develop a best practice approach to managing IoT-data. Further, developing an approach on top of a record-keeping standard may help avoid vendor lock-in and potentially see increased competition by vendors.

8 Conclusion

The amount of IoT-data collected by government agencies is likely to see phenomenal growth over the coming years and, as such, there will be a requirement to seriously consider the approach to managing such data or face the consequences when found to be in non-compliance with laws and regulations.

Two use-cases were chosen as a basis to explore the issue of managing IoT-data. One with publicly available data and one that had real privacy issues. The former use-case followed information governance practices and, to a certain

degree, the open data portal follows the GARP principles. It is unclear, in the latter use-case, to what degree a best practice was in place. It was not possible to get relevant information from the agency, nor do they publicly share internal operational information in the same way the open data portal does. It is clear that there are issues regarding the management of information, and the Norwegian Customs Agency found themselves in an unfortunate position where they collected privacy information beyond the scope of relevant laws.

The ANPR-data use-case shows that there is a need to apply record-keeping principles to the management of IoT-data. An organisation will likely achieve better control of their IoT-data by centralising the management of such data in a system that follows some record-keeping principles. An ad hoc approach to managing IoT-based data may quickly result in the imposition of sanctions by relevant regulatory agencies.

The GARP principles can serve as a basis for a discussion to explore the issue of IoT-data from a high-level. The principles are a useful tool and can quickly help identify potential data management issues. This is particularly relevant to the ANPR-case.

It is also important to explore the issue from a practical point of view. Treating IoT-data as records, collected and managed within a flexible record-keeping system, may allow for better information governance. A standards approach is desirable to avoid potential vendor lock-in situations and to ensure interoperability.

The record-keeping profession has relevant tools and techniques applicable to the management of IoT-data. However, a record-keeping approach certainly is not the only solution. Approaches to handling IoT-data will likely benefit from a discussion of how multiple information science domains can contribute to ensuring government agencies can use IoT-technology, without infringing on a citizen's right to privacy.

Bibliography

Akpakwu, G. A., Silva, B. J., Hancke, G. P. and Abu-Mahfouz, A. M. (2017), 'A survey on 5g networks for the internet of things: Communication technologies and challenges', *IEEE Access* **6**, 3619–3647. https://doi.org/10.1109/ACCESS.2017.2779844.

Arkivverket (2018), Noark 5 norsk arkivsystem versjon 5.0, Standard, The National Archives of Norway, Oslo, NO.

Arkivverket (2019), Spesifikasjon for noark 5 tjenestegrensesnitt, Standard, The National Archives of Norway, Oslo, NO.

ARMA (2014), *General Accepted Recordkeeping Principles® GARP*, ARMA. https://www.arma.org/page/principles.

Bonomi, F., Milito, R., Zhu, J. and Addepalli, S. (2012), Fog computing and its role in the internet of things, *in* 'Proceedings of the first edition of the MCC workshop on Mobile cloud computing', ACM, pp. 13–16. https://doi.org/10.1145/2342509.2342513.

City of Chicago (2020*a*), 'Beach water quality - automated sensors', https://data.cityofchicago.org/Parks-Recreation/Beach-Water-Quality-Automated-Sensors/qmqz-2xku. Accessed 2020-02-16.

City of Chicago (2020*b*), 'Chicago data portal', https://data.cityofchicago.org/. Accessed 2020-02-16.

Compton, M., Barnaghi, P., Bermudez, L., GarcíA-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A. et al. (2012), 'The ssn ontology of the w3c semantic sensor network incubator group', *Web semantics: science, services and agents on the World Wide Web* **17**, 25–32. https://doi.org/10.1016/j.websem.2012.05.003.

European Commission (2016), 'Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation)', https://eur-lex.europa.eu/eli/reg/2016/679/2016-05-04.

Evans, D. (2011), 'The internet of things: How the next evolution of the internet is changing everything', *CISCO white paper* **1**(2011), 1–11. http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411\protect\discretionary{\char\hyphenchar\font}{}{}FINAL.pdf.

Finch, K. and Tene, O. (2018), *Smart Cities: Privacy, Transparency, and Community*, Cambridge Law Handbooks, Cambridge University Press, p. 125–148. https://doi.org/10.1017/9781316831960.007.

Fresko, M. and Waldron, M. (2001), 'Model requirements for the management of electronic records (moreq)', *Cornwell Affiliates plc* . https://ec.europa.eu/idabc/en/document/2631/5585.html.

Hagen Sataslåtten, O. (2014), 'The norwegian noark model requirements for edrms in the context of open government and access to governmental informa-

tion', *Records Management Journal* **24**(3), 189–204. https://doi.org/10.1108/RMJ-09-2014-0041.

Haque, R. and Hacid, M.-S. (2014), Blinked data: Concepts, characteristics, and challenge, *in* '2014 IEEE World Congress on Services', IEEE, pp. 426–433. http://doi.org/10.5334/dsj-2015-002.

ISO-14721 (2012), ISO 14721:2012 Space data and information transfer systems – open archival information system (oais) – reference model, Standard, International Organization for Standardization, Geneva, CH.

ISO-15489-1 (2016), ISO 15489-1:2016 Information and documentation – records management – part 1: Concepts and principles, Standard, International Organization for Standardization, Geneva, CH.

ISO-16175-2 (2011), ISO 16175-2: 2011 Guidelines and functional requirements for records in electronic office environments, Standard, International Organization for Standardization, Geneva, CH.

ISO-20802-1 (2016), ISO/IEC 20802-1:2016 Information technology – open data protocol (odata) v4.0 – part 1: Core, Standard, International Organization for Standardization, Geneva, CH.

ISO-23081-1 (2006), ISO 23081-1:2006 Information and documentation — records management processes — metadata for records — part 1: Principles, Standard, International Organization for Standardization, Geneva, CH.

ISO-30300 (2011), ISO 30300:2011 Information and documentation — management systems for records — fundamentals and vocabulary, Standard, International Organization for Standardization, Geneva, CH.

ISO-30301 (2011), ISO 30301:2011 Information and documentation — management systems for records — requirements, Standard, International Organization for Standardization, Geneva, CH.

Jin, J., Gubbi, J., Marusic, S. and Palaniswami, M. (2014), 'An information framework for creating a smart city through internet of things', *IEEE Internet of Things journal* **1**(2), 112–121. https://doi.org/10.1109/JIOT.2013.2296516.

Katuu, S. (2016), 'Managing digital records in a global environment: A review of the landscape of international standards and good practice guidelines', *The Electronic Library* **34**(5), 869–894. https://doi.org/10.1108/EL-04-2015-0064.

Kelly, M., Furey, E. and Blue, J. (2019), Gdpr article 17: Eradicating personal identifiable information and achieving compliance in a hybrid cloud, *in* '30th Irish Signals and Systems Conference (ISSC)'.

Khan, Z., Pervez, Z. and Ghafoor, A. (2014), Towards cloud based smart cities data security and privacy management, *in* '2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing', pp. 806–811. https://doi.org/10.1109/UCC.2014.131.

Kosmo, J. (2010), 'Riksrevisjonens undersøkelse av arbeidet med å sikre og tilgjengeliggjøre arkivene i kommunal sektor - dokument 3:13 (2009–2010)', https://www.stortinget.no/nn/Saker-og-publikasjonar/publikasjonar/Dokumentserien/2009-2010/dok3-200910/dok3-200910-013/.

Perera, C., Ranjan, R., Wang, L., Khan, S. U. and Zomaya, A. Y. (2015), 'Big data privacy in the internet of things era', *IT Professional* **17**(3), 32–39. https://doi.org/10.1109/MITP.2015.34.

Phipps, S. (2007), 'Roman canaries', https://sunmink.wordpress.com/2007/09/27/roman-canaries/.

PREMIS Working Group et al. (2005), 'Data dictionary for preservation metadata', *OCLC*. **22**, 2008.

Simon, K. D. (2005), 'The value of open standards and open-source software in government environments', *IBM Systems Journal* **44**(2), 227–238. https://doi.org/10.1147/sj.442.0227.

Stuart, K. (2017), 'Methods, methodology and madness: digital records management in the australian government', *Records Management Journal* **27**(2), 223–232. https://doi.org/10.1108/RMJ-05-2017-0012.

Sundmaeker, H., Guillemin, P., Friess, P. and Woelfflé, S. (2010), 'Vision and challenges for realising the internet of things', *Cluster of European Research Projects on the Internet of Things, European Commision* **3**(3), 34–36. https://doi.org/10.2759/26127.

Thon, B. E. and Kaspersen, K. (2019), 'Varsel om overtredelsesgebyr', https://www.datatilsynet.no/contentassets/a1654dc804774a6ab315efb\protect\discretionary{\char\hyphenchar\font}{}{}0ce53a5e3/varsel-om-otg---tolldirektoratet.pdf. Notice by the Norwegian Data protection agency about the issuance of a fine (Doc. 18/01144-2/KBK).

Wilhelm, P. (2009), 'An evaluation of moreq2 in the context of national edrms standard developments in the uk and europe', *Records Management Journal* **19**(2), 117–134. https://doi.org/10.1108/09565690910972075.

Wittern, E., Cha, A. and Laredo, J. A. (2018), Generating graphql-wrappers for rest (-like) apis, *in* 'International Conference on Web Engineering', Springer, pp. 65–83. https://doi.org/10.1007/978-3-319-91662-0_5.

Yeo, G. (2012), 'The conceptual fonds and the physical collection', *Archivaria* **73**, 43–80. https://archivaria.ca/index.php/archivaria/article/view/13384.

Zanella, A., Bui, N., Castellani, A., Vangelista, L. and Zorzi, M. (2014), 'Internet of things for smart cities', *IEEE Internet of Things journal* **1**(1), 22–32. https://doi.org/10.1109/JIOT.2014.2306328.