

The Sedona Conference WG1 Draft Commentary on Discovery Implications of the Internet of Things

Drafting Team Members

Kevin Clark	Sean Cotulla
Warren Kruse	Sara Lockman
Kyle Pozan	Christopher Suarez
Hon. Juan Villaseñor	Josh Zylbershlag

Team Leaders

David Gaston	Steven Teppler
--------------	----------------

Steering Committee Liaisons

Jennifer Coleman	Ross Gotler
Claire Hass	Tessa Jacob
Greg Kohn	

Copyright 2022, The Sedona Conference. All rights reserved.



This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

DRAFT: The Sedona Primer on the eDiscovery Implications of the Internet of Things (IoT) The Sedona Conference Working Group 1 – March 2022

I. INTRODUCTION

A. What is the Internet of Things and what are IoT devices?

In its Commentary on ESI Evidence & Admissibility, The Sedona Conference defines the Internet of Things (“IoT”) as “...a network of computing devices and sensors embedded in everyday objects that create, collect, and share data through the internet.” IoT devices are objects that embed a communication technology such as WiFi or Bluetooth, that allows them to communicate with the Internet and other networks, such as those in homes and businesses. IoT devices have become commonplace in the consumer, commercial, and industrial contexts. Some examples of IoT devices include wearables such as smart watches and fitness and sleep trackers, smart thermostats that adjust to our habits, home control devices like garage door openers and home cameras that we can control from anywhere, smart speakers that play content on command, and devices that automatically track and report on attendance and productivity in the business setting. The data that these devices create is often stored in structured databases and may be stored in multiple locations including on the devices themselves and in the Cloud.

The universe of discoverable electronic information has been significantly expanded by the emergence of billions of IoT devices as data sources. These devices, their ubiquity, and the potentially discoverable information generated, collected, transmitted and stored in association with them pose unique discovery challenges which may not fit neatly into the current discovery regime of the FRCP and associated decisional authority. In this primer, we offer information and guidance to parties, the bench, and the bar on the intersection of eDiscovery law and the IoT.

II. CATEGORIES OF IoT ESI

As a starting point for discussing the eDiscovery implications of the Internet of Things, it may be helpful to set forth some different categories of electronically stored information (“ESI”) relating to IoT devices – “IoT ESI.”

A. Data generated by the IoT device itself

IoT devices generate ESI, sometimes a voluminous amount of it. Some of the ESI may be intended for easy review by the end-user while other ESI may only be intended for rare instances when a technician or engineer needs to review the performance of the device but not otherwise meant for end-users to ever access. Examples of the former include video recorded by a Ring doorbell for review by the homeowner or workout metrics captured by a Fitbit device which are meant for the wearer to review on the device or on an associated app. Examples of the latter include logs maintained by a car’s onboard computer that records engine performance, or an Apple Watch maintaining a record of app crashes.

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

As regulatory and consumer attention has focused on privacy concerns related to the information created by IoT devices, manufacturers of such devices have begun to provide consumers with greater insight and control over ESI maintained by the devices. For example, some car manufacturers now provide consumers with granular control over how data from the onboard computers systems is used. Consumers can choose whether driving metrics are shared with insurance companies.

B. User communication devices used to remotely activate or interact with the IoT device (“User Activation Device”)

Many IoT devices are designed to be accessed, controlled, or used through communication with another device. Examples include a wide-variety of home automation devices that are controlled from applications installed on mobile devices, like smart phones. These can include lights, heating and air-conditioning systems, lawn sprinklers, and many others. Some IoT devices are designed to only be controlled via another device. For example, Apple’s AirTag doesn’t include any buttons; consumers configure and use the device via their phones. Other devices can be operated without the use of another communication device but connecting them to a phone or tablet provides access to functionality that cannot be accessed directly from the IoT device. As an illustration of this point, many modern air-conditioners can connect with smart phones and then provide consumers with details around usage and energy consumption that cannot be accessed directly from the air-conditioner.

C. Web-based management platform (WMP) for a device or devices.

Web-based management systems often store ESI generated by single or multiple IoT devices. A foreperson monitoring the work of a host of factory floor robots can use a web interface to assess the operation of the robots and provide control input from such an interface. A consumer can access their broadband provider’s website and then review and control a variety of devices connected to their home network.

D. Repository (local on the device itself or cloud based) used to store data.

The vast quantities of ESI generated by IoT devices are often gathered and stored on local or cloud-based storage repositories. A typical Apple Watch now contains 32GB of storage space, which can include both data intentionally placed there by the user (e.g., music, notes, photos) or maintained by the device as part of its ordinary operation. However, the same Apple Watch is usually integrated with the owner’s iCloud account and historical information can often reside there for years. Another example is the Amazon cloud where users of Alexa devices can find (and delete) audio recordings created every time they activated the Alexa voice assistant on their devices (both intentionally and unintentionally).

E. Third party applications, systems integrations, or ‘downstream’ locations that connect to the WMP or storage repository.

Much of the success and power of IoT devices relies on their inter-connectivity with a host of other devices, networks, and platforms. Fitness trackers for instance, not only

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

provide consumers with workout and health metrics, but can also connect to wellness platforms maintained by insurance companies, which then reward the consumers based on their activity levels. Such downstream repositories can contain ESI from millions of devices with differing levels of anonymization and accessibility.

III. IoT ESI AND DISCOVERY SCOPE, LIMITS, AND PLANNING

A. Federal Rule of Civil Procedure Rule 26(b)

Under Federal Rule of Civil Procedure 26(b), “the scope of discovery is as follows: Parties may obtain discovery regarding any nonprivileged matter that is relevant to any party's claim or defense and proportional to the needs of the case, considering the importance of the issues at stake in the action, the amount in controversy, the parties’ relative access to relevant information, the parties’ resources, the importance of the discovery in resolving the issues, and whether the burden or expense of the proposed discovery outweighs its likely benefit.”

To the extent IoT ESI is probative, proportional, and not privileged, it may be in scope. Key proportionality and burden considerations directed to IoT ESI discovery include:

- Potentially higher cost of collection activities.
- The availability—or lack thereof—of tools (i.e., hardware, software, processes) available for identification, collection, processing, review and production of IoT ESI.
- Endpoint firmware considerations – is the IoT ESI still able to be view and understood if accessed through the current version of the IoT device or interface?
- Potentially massive data volumes.
- Can data be retained and, if so, how?
- Managing both structured and unstructured data elements.
- Voluminous, broad metadata.
- Encryption and other security protections built into the IoT ESI.
- Data location and category of IoT ESI
- Non-parties which may have control – potentially exclusive – of the IoT ESI.
- Privacy interests themselves, and additionally the burden involved in reasonably balancing the need to provide discovery with the requirements of applicable data privacy, data secrecy, or employment laws (including, for example, the EU’s General Data Protection Regulation).¹

B. Federal Rule of Civil Procedure Rule 26(f)

Where IoT ESI is implicated as discoverable in litigation, it is important to understand the “what,” or type and format of IoT ESI involved, as well as the “where,” or the custodial

¹ Privacy considerations, protective orders, and associated special handling for IoT ESI that may be worth considering during the course of discovery are out of scope for this paper but covered in other publications including The Sedona Conference Data Privacy Primer and *The Sedona Conference Commentary on Privacy and Information Security*.

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

repository of such ESI, prior to making a determination of relevance and proportionality. Such disclosures may take place during Rule 26(f) meet and confer sessions, where parties can make informed, good faith efforts to reach initial agreement on discovery scope. Such initial agreements (to the extent possible) should then be reflected in an ESI protocol that permits some flexibility to both requesting and producing parties for changing circumstances.

IoT is a broad term and the types of ESI implicated in litigation can vary greatly depending on how the technology was designed, how it functions, and how it interacts with other systems. With reference to and reliance on existing Sedona papers, the following concerns, and best practices related to meet and confer discussions², and ESI protocols, should be taken into consideration at the earliest possible time to assist in determining (and agreeing to) the proper scope of discovery pursuant to FRCP 26(b) and 26(f).

Parties will benefit from an appropriate level of planning and preparation as part of effectively conducting meet and confer activities relating to IoT ESI. In many circumstances, it could be helpful to engage technical experts who are familiar with the specific IoT ecosystems at issue, as well as the business owner who uses the information to inform, report, or otherwise perform their job function. Examples of key questions to ask business and technical experts about a target IoT system:

- What is the architecture of the IoT ecosystem and what IoT categories are in play?
- Has any of this ESI been produced for litigation in the past?
- What are the origins of the system – is it proprietary or externally licensed?
- Who has the power, possession, or control over the data?
- What data elements are collected and stored?
- What decisions does the system make or support?
- Is there regular reporting that is generated?
- How is the data stored and who can access it?
- What options exist for searching or exporting the data?
- Are there any retention rules or auto-purging of the data?
- What encryption or access barriers exist?
- Does the system collect protected or sensitive data?

These technical resources may provide information on what search and collection capabilities are available across the IoT categories of ESI. In many situations, more standard ESI search and retrieval methodologies may not be usable or applicable to IoT ESI. Custom approaches may be necessary, especially when an IoT system is newer technology, under the primary control of a non-party, or never designed with search and retrieval in mind.³ Absent built-in search capabilities, data indices, or similar filtering functionality, the burden of identifying potentially relevant data within the raw data stores may increase. In these instances, attorneys may look to engage IT software development resources, third party tools, or

² Some states, including Florida, do not have an analog to the Rule 26(f) meet and confer requirement.

³ Consider as a parallel example email archiving systems designed to comply only with archiving requirements that did not have usable search and retrieval capabilities, thus severely complicating discovery and production.

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

consultants to navigate the search and retrieval of ESI within complex IoT systems. These processes can be very time consuming and bring with them potential issues including alteration of the data or system itself. All of these custom approaches may increase the burden involved and should be weighed against the value and proportionality of the ESI.

As part of a meet and confer process, understanding the manner in which a party intends to use IoT data may be helpful, including as part of evaluating the burden v. benefit portion of proportionality considerations. By way of example, see TSC Database Principles (2014 Edition), Comment 6.A.:

Comment 6.A. Discussing the Intended Reasonable and Legitimate Uses of Database Information Can Result in a More Useful Production Format

While a requesting party is not required to divulge its counsel's work product or its litigation strategy, it may be impossible for a responding party to take appropriate steps to provide database information in a reasonably useful format if it has no idea of how the requesting party intends to use it. A requesting party's failure or refusal to identify the intended use of database information, especially upon request, may limit the responding party's ability to accommodate the format request, particularly where the responding party's preferred format is less expensive and appears ex ante reasonable. To maximize the value of the database information it will receive, a requesting party should provide detail sufficient to describe the tools or broad evidentiary use that it intends to make of this material.

Depending on the specific claims or defenses, the requestor may be seeking something more complex than just IoT data, for example decisions made by or facilitated by artificial intelligence (AI) or machine learning (ML) processes embedded within the IoT ecosystem. In this example, if the intent is to establish a specific fact (e.g. a temperature captured by a sensor at a certain point in time) the collection approach will be different than if they are seeking to understand a decision made by the system (e.g. to lower the temperature). For the latter, the logic underpinning the decision may be relevant and require review of the software itself, like source code, rather than the raw data stored within the IoT system's repository. Therefore, if requesting party can gain sufficient detail as to the intended use of the IoT ESI, they can more effectively scope and respond to the request for discovery.

C. Discovery Plans/ESI Protocols

IoT data plays a ubiquitous role in our lives and continues to become more complex and far-reaching with time. This complicates discussions around the preservation, collection, review, and production of IoT ESI. As such, when included in the allowable scope of discovery, parties should address issues around IoT ESI proactively as part of discovery plans and ESI protocols.

The volume and scope of the data captured by a target IoT ecosystem can be extremely large and untenable to produce in its entirety. Leveraging the technical and business expertise

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

available to answer key questions upfront is critical to navigating a 26(f) conference effectively. Those resources can help devise a practical and proportional approach to IoT ESI production. For example, in some situations, it may be helpful for the responding party to come prepared to educate the requesting party on specific IoT data sources, the available methods of collection, and any associated barriers (e.g., time, cost, volume) that may complicate, burden, or prolong the discovery process.

Such efforts, especially in conjunction with details provided by the requesting party on why they are requesting the IoT ESI and how they intend to use it, as discussed above, may facilitate negotiation of the ESI protocol and reduce the effort required to scope and agree on an approach between parties.

IV. IoT ESI AND PRESERVATION, COLLECTION, REVIEW, AND PRODUCTION

D. Preservation

1. Preservation obligations

As is often the case, a party may have a pre-litigation obligation to preserve ESI. The parameters of that obligation, however, may be difficult to establish when addressing discovery of IoT ESI. The ubiquity of devices and types of data, some undoubtedly proprietary, will complicate determinations of possession, custody, control, and ultimately, reasonableness and proportionality.

Possession, custody, or control are discussed in further detail below, but the “smart” thermostat (i.e., one that can be controlled remote with a phone, tablet, smart speaker or other IoT device) is illustrative of the issue. An individual may own and install a smart thermostat but while the device is considered to be in the possession of the homeowner, the possession and control of some, or all of the IoT ESI from that thermostat data may be stored, processed, and reprocessed at any number of cloud based data “locations”, with no one location comprising the repository of all of the device’s IoT ESI. Some of the data may reside on the device itself, such as information about the furnace and air conditioning unit, as well as pre-programmed schedules. Location data for geofencing features and local weather data may reside on the phone application permitting remote control of the device, while the activity history and the majority of the remaining data may be stored in cloud and unavailable to the individual through normal channels (e.g., by contacting customer support). Acquiring “all” the thermostat’s IoT ESI may require the service of several Rule 45 subpoenas to non-parties as well as from direct party-to-party discovery pursuant to Rule 34.

Even if all of the data generated by the IoT device is in an individual’s or entity’s possession, custody, or control, obtaining the data can be complex, difficult, or beyond what may be considered “reasonably accessible because of undue burden or cost” under Rule 26(b)(2)(B), or proportional to the needs of the case under Rule 26(b)(1).

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

In such circumstances, courts may determine that it is the burden of the device manager or party with possession, custody, or control to demonstrate why native preservation is not reasonable or proportional, including by submitting specific information, often by declaration, explaining the burden. In such cases, if preservation is even feasible, alternate means of preservation may be available, and include the use of separate media or off-device storage. Additional burden may be demonstrated if the majority of the data is stored in the cloud in a fully encrypted format, in which case it may not be possible, even with a valid subpoena, to obtain that IoT ESI from the device manufacturer without substantial time and cost, rendering it not reasonably accessible. This determination will hinge largely on the purpose for seeking such IoT ESI. If, for instance, the purpose is to obtain the thermostat's IoT ESI to determine whether an individual was present in the same location on a given day, and that is the only source of relevant evidence, a burden objection might not be sustained. If however, the requesting party already has video footage of the location, the burden to obtain the smart thermostat IoT ESI—even if relevant—may not be proportional to the needs of the case. It is important to keep in mind that while this burden test is already embodied in the rules, ascertainment of IoT ESI burden itself may be burdensome.

IoT ESI preservation obligations also depend on the type of data involved and the determination of who, or what is the custodian, as discussed further below. While retention policies should be in place for those that administer and maintain IoT ESI, questions remain: Who, or what has IoT ESI possession, custody, and control, and who, or what has an obligation to preserve IoT ESI? IoT ESI data controllers (the data “owners”) and processors (those who are engaged to process data on behalf of data controllers) may need to be have, and be transparent about their retention policies, is it even fair to assume that IoT ESI device component manufacturers (of which there may be many in any single device) must anticipate litigation for what may be millions of devices in use? For example, a third party software developer whose code comprises only a small portion of the software intended for a single operation (e.g., smart auto braking systems) might not be deemed to “reasonably anticipate” litigation (and trigger a duty to preserve). And what obligations would a party have in such a situation? The complexity of these systems makes that determination under federal (and many states’) common law challenging, and underscores the need for guidelines for both the bench and the bar.

2. The preservation process

Preservation challenges will be very similar to the ones described below relating to collection. A potentially additional complicating factor will be the configuration of the data storage and associated retention within the IoT system. In the context of the original system's purpose, it may be reasonable to purge or overwrite data once it no longer provides value to the end-user, but that may not be sufficient for purposes of litigation. Understanding what specific IoT data categories need to be preserved and the specific time frame of relevance prior to discussing with technical system owners will aid in determining an approach for preservation. This approach

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

may require collecting to preserve, rather than preserving in place, depending on the storage limitations or configuration of the system itself.

A central issue relating to IoT ESI preservation is how a responding party should preserve IoT ESI that may, and potentially in short course, become unreadable or inaccessible due to changes in technology for IoT devices or management platforms. What activities short of those that are neither reasonable nor proportional, should be expected by a party to maintain accessibility and usability of the IoT ESI?

An early conversation about the objective of requests for IoT data can help raise issues relating to the need for and discoverability of the data in question. Additionally, it can help shape the understanding of what may comprise preservation efforts that are reasonable and proportional. Early discussions can be held with custodians or those responsible for the management of the IoT devices to identify relevant IoT data, and to discuss with them and applicable technology resources a strategy for preservation of the data.

3. Is IoT ESI ephemeral?

While not inherently ephemeral, IoT data can be ephemeral or may persist per the needs of the user, the function of the device, or the relevant records management policy. However, ESI from an IoT device may be functionally ephemeral depending on the design and function of the device. IoT data may also be ephemeral on the device itself where the primary retention of the device data is intended to be maintained elsewhere, such as on paper (the short term memory of a copy machine) or in a centralized server (a fitness step counting device).

If information is functionally ephemeral, parties may be well-served to discuss (including being raised by the producing party when appropriate) and determine means of early collection, remediation, or alternative data sets that capture the discoverable information in question. As an example, a preserving data from a program that captures step data daily from a device could be used as a substitute to preserving the data on the actual device.

E. Collection

IoT ESI may be highly proprietary and difficult to collect and process using standard e-discovery tools and workflows. IoT device acquisition, end-use application data, repository or WMP exports, and other IoT ancillary device data collections may not be routine discovery activities for the requesting party. Depending on the target of collection, it has the potential to be expensive, complex, and time consuming. Whether IoT ESI of any category is reasonably accessible is highly dependent on the source itself along with the facts and case at hand. Below are considerations for each of the IoT ESI categories identified in this commentary.

1. Understanding the IoT ESI

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

An understanding of the technology and nature of IoT ESI may be helpful as a starting point for collection activities. Some initial questions that can be asked of an expert in the IoT ESI to help assess its nature and accessibility are:

- Is the native format of IoT ESI reasonably usable or in a proprietary format?
- What collection or export tools are available?
- Does business intelligence (BI) reporting or similar aggregations of the IoT ESI exist that are more reasonably accessible and meaningful than raw data exports (e.g. logs, streams, sensor data)?

2. Physical collection of IoT devices and user activation devices

Devices within a target IoT ecosystem can take on many forms and although some consumer IoT devices may be standalone and physically accessible, like a smart thermostat or smart watch, those utilized in industrial applications can be more challenging to acquire. Highly engineered specialty IoT devices may be physically difficult to gain access to, like weather sensors positioned on a commercial building rooftop, or embedded into larger machines, appliances, or ecosystems making them problematic to isolate and retrieve. Particularly for IoT devices supporting industrial applications like transportation, safety, or healthcare, factors like the number of devices or the potential disruption of service that could result from disconnection should be considered prior to collection.

Even once an IoT device is in hand, the components or operating system may evade traditional forensic collection tools and software available to digital forensic teams today. It is important to understand if the physical collection of the IoT device is necessary given the circumstances of your case and the issues involved. Similar to mobile device ESI collections, if the IoT device does not store data locally or otherwise transmits all information downstream to a data repository or ancillary system then direct IoT device acquisition may not be necessary or preferred. This is because the ESI is likely duplicative and retrievable through a more reasonable, less-burdensome method. The same applies to user activation device collections as well.

3. Web-based management platform (WMP) and data repositories

The WMP application or data repository categories of IoT ESI may offer better options for accessing and collecting potentially relevant data through custom queries of structured data or existing reports.

Native production of an entire data repository may be overly broad and unreadable by the receiving party without appropriate hardware or software licenses. Every data element (or field) available could be assessed for relevance prior to exporting any ESI. This process of querying raw tabulated data within a repository could allow for the targeting of specific data elements, assemble them in a readable format, and produce a “document” to support a discovery response that meets the needs of the requesting party.

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

Accessing backend database structures and extracting data from them can be complicated and require elevated access permissions to view, so it is important to engage a knowledgeable technical resource to validate the accuracy and completeness of the query itself before relying on the resultant export. Further compounding this challenge is the dynamic nature of many IoT data fields, which may make historical snapshots or static pulls problematic. Different cases will require different types of information and will make use of database information in different ways – so a custom collection approach will be common.

If there are existing aggregations of IoT ESI available, such as business intelligence reporting or data visualizations (sometimes referred to as dashboards), within the WMP interface or the data repository utilized by the end-users or custodians, those may serve to be more meaningful than exports of raw data tables. Standardized reports are often available in commonly accessible export formats (e.g., Excel, PDF, .csv), which are less burdensome to retrieve, and could provide useful context for how a person (or machine) viewed or used the ESI to make a key decision at a specific point in time.

4. Cloud-based service provider repositories

IoT ESI may be stored in Cloud-based repositories with service providers, potentially complicating the collection process. Parties should aim to understand the complexities of such repositories, including:

- What are the terms of service associated with each IoT ESI location?
- Which repositories need to be searched?
- What parameters can be used to define a “document” or a “record” within the IoT ESI data set?

5. Downstream ancillary systems

Like any database, the data generated can be fed directly and directly to a wide variety of systems. These are considered “downstream” systems. They can vary considerably in retention, organization, and accessibility. When these systems are known, they may be candidates for production sources, or as alternative production sources.

6. Collection practices

With consideration given to the results of the meet and confer process, the responding party should determine an effective search methodology that takes into account both relevance and proportionality. The available supporting resources (e.g., internal IT, system users, third party discovery consultants) may be useful sources to provide estimates as to the time and cost to address the challenges that may come into play when searching for and retrieving IoT ESI. As with ESI generally, where IoT ESI volumes are large, consider sampling or limiting the timeframe of the search so

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

the case team can review and vet (or vet with the opposing party) the resultant data export(s).

As ESI searches and exports are generated by technicians or system owners, the results should be reviewed to confirm they are readable and can be produced in a usable format, and capture the intended scope of ESI that is potentially relevant to the discovery request. For IoT systems that have not previously been in scope for litigation, expect the search and collection testing process to take several iterations, including detailed feedback discussions between the IT and legal resources before a final search and collection method can be decided upon. Performing collection validation steps can help to ensure the accuracy and repeatability of the IoT ESI collection process.

As collection is completed, documenting the methodology used to perform the collection may assist in defending the approach and facilitate supplemental collections as necessary in the future. For example, the specific data sources, the collection process, performing technician, quality control steps taken, and the date and time of the ESI acquisition. Additionally, any reference information used to support the collection process, like a data dictionary or process flow diagram, can be documented to demonstrate the state of the system at the time of collection.

Nonetheless, collections may also be complicated by or dependent on other technical factors that could impact the ESI being collected even on a granular data field level. Some examples of these types of collection considerations include:

- Storage format, which may be proprietary in nature and require specialty software to access or otherwise render the content readable.
 - The state of a specific data element, or field, which may be dynamic or static in nature.
 - Custom data visualizations, reports, or similar individualized settings that could alter how the data is displayed, read, or interpreted.
 - Permission-based user accounts that may limit the view of IoT data within an application.
 - Data elements or fields generated by machine learning or AI models (rather than collected by sensors or inputted by end-users).
- Upstream or downstream sources of data that may integrate with other data within the IoT system.

F. Review and Production

The resultant ESI from the collection process will determine the best approach for an appropriate and effective route for review. The export format(s) of data may limit, augment, or evade traditional ESI document review processes and production labeling (bates, redactions, endorsements).

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

If the IoT ESI is human-readable in a standalone file format like Microsoft Office product files or Adobe PDF files, then there will be more options to support review and production. These types of files are conducive to nearly all commercially available eDiscovery software platforms that ingest, process, and render the content viewable and provide work product tagging functionality.

Where large volumes of structured data are involved, it would be beneficial to agree upon parameters that can define a “document” or “record” for purposes of managing the review, work product, and productions. A common approach is to leverage metadata that capture dates/times to time-bound document definition (for example, every 24 hours of data will represent a document). However, other criteria may be more relevant to the case at hand, like segregating data sets by a specific geographic area, sensor type, or similar. Without setting these document guidelines, it may become cumbersome to open the data set without taxing a computer system, effectively tag information for relevance, perform necessary redactions (like for PII/PHI), or otherwise label ESI for reference across parties (e.g. bates stamping, endorsements). Once structured data sets are delimited and/or document parameters defined, they will be much easier to manage and often ingestible into eDiscovery software platforms for reviewers to access. Alternatively, depending on the circumstances, it may be less burdensome to review the data in the native database format for reviewers to use, query, and/or tag existing tables, line items, or records as originally defined within the system.

Where IoT ESI requires a specific interface to display or contextualize the information (e.g. software, visualization, or other graphical interface) then review becomes more complicated and, potentially, more costly. Similarly, should source code be in scope, it will require specialized knowledge to understand and review, if not specialized development software licenses to access.

Potential conflicts – or at least complexities – may exist between the requirement under Federal Rule of Civil Procedure 34(b)(2)(E)(ii) that “If a request does not specify a form for producing electronically stored information, a party must produce it in a form or forms in which it is ordinarily maintained or in a reasonably usable form or forms” and the reality with IoT ESI that the format in which it is ordinarily maintained may not be reasonably usable. While this suggests a possible need to address more emerging forms of potential evidence in future rules, it also highlights the practicality of parties reaching agreement on production format of IoT ESI so that it may effectively be reviewed and managed as part of discovery.

V. IoT ESI AND POSSESSION, CUSTODY, OR CONTROL AND NON-PARTY CONSIDERATIONS

Federal Rule of Civil Procedure 34 allows parties to serve on other parties a request for documents within their “possession, custody, or control.” While this topic is discussed at length in *The Sedona Conference Commentary on Rule 34 and Rule 45 “Possession, Custody, or Control”* and *“The Sedona Conference Commentary on Rule 45 Subpoenas to Non-Parties, Second Edition,”* IoT ESI presents some distinct considerations.

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

A. Who is the owner or custodian of the IoT ESI?

Determining who is the owner or custodian of IoT ESI for purposes of determining possession, custody, or control can be challenging and will necessarily vary by IoT ESI categories. The “owner” may often be a combination of the service provider and the user (individual or organization). An understanding of the relevant service terms and conditions may be a helpful part of this analysis.

Also helpful is an understanding of where the relevant data is generated. For example, the IoT device, the user activation Device, WMP, or some combination thereof. This may help demonstrate practical control over the IoT ESI. Parties, and courts, may look to local law, decisional authority, or terms of service.

B. Possession, custody, or control

There has been an explosion of distributed or “cloud” computing, remote ESI hosting, Software as a Service, mobile data platforms, wearables data, smart home devices, industrial IoT ESI, infrastructure IoT ESI, and other Internet-connected devices. Ascertaining possession, custody or control of such IoT ESI will require a case-by-case determination, and will necessarily vary with the category of IoT ESI, and a concurrent determination of accessibility to a producing party. The principal issues to consider are:

- The physical location of the device and the data it creates.
- The location of the data created by the device.
- The reasonable access a party has to the device, or to a device component or sub-component, etc. and the data it generates.
- The retention policy of the manager of the device and the data it creates.
- Whether data physically held on the device and the data it creates is reasonably accessible.
- Whether data on local network-attached storage is reasonably accessible.
- Whether data stored offsite/in the cloud is reasonably accessible.
- Who possesses the data stored offsite/in the cloud?
- Who created the data, user or device based on default settings?
- The function within the ordinary course of business (in the context of Rule 34) of the device and the data it creates.
- Any resources that aggregate or otherwise log and/or process information from the device.
- Any resources that duplicates data created by the device.
- Accessibility (or demonstrable lack thereof).

These considerations not only should apply to any IoT ESI category, but should be used as guidelines for proportionality determinations, and ultimately determining whether a party must produce documents sought by a requesting party. The weight accorded to these factors will also vary depending on which test—either the legal right test or the practical ability

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

test—the court employs to define possession, custody, or control for the discovery sought by the requesting party.

C. Differing preservation burdens for parties and non-parties

As mentioned above, non-party considerations in the discovery process are covered in detail in “The Sedona Conference Commentary on Rule 45 Subpoenas to Non-Parties, Second Edition.” Potentially relevant IoT ESI may be controlled – possibly exclusively – by a non-party, perhaps even a massive corporation with little interest in responding to non-party discovery requests. While a party to litigation may be held to a broad spectrum preservation trigger, it is suggested that, in the case of IoT ESI, a non-party respondent to an IoT ESI discovery request has a narrow spectrum retention trigger, with no duty to retain IoT ESI absent service of a subpoena or express obligation to preserve by way of contract, some special relationship with one of the parties, or a stake in the litigation.

VI. IoT ESI AND ADMISSIBILITY

A. Introduction⁴

As discussed above, the IoT consists of billions of networked computing devices. IoT ESI already plays a significant role in cases. For example, in one murder case, data indicating movement from a wife’s fitness wearable convinced the police that her husband killed her. In another, prosecutors used Fitbit data to show that a victim falsely accused a man of raping her. With the rapid evolution of technology, the use of IoT ESI as evidence in criminal and civil litigation is becoming more sophisticated. IoT ESI generation encompasses home-security camera systems, like Nest (from Google), or Blink (from Amazon). Nest, for example, can recognize persons, animals, vehicles, or familiar faces thanks to face-recognition technology. Such systems record video and sound to a cloud and may be stored for future use. In one case, for example, the sound recorded on a Nest camera was “married” with video from a police surveillance camera to provide (e.g., through a new method of corroboration) a more complete depiction of a crime scene in which a fight and a shooting occurred between multiple individuals.

This section addresses whether the relevant IoT ESI may or may not be admissible in a particular civil or criminal case assuming there is an already established duty to preserve and produce.

As with any evidence, there is a risk that IoT ESI data might be corrupted, or manipulated at some point after it is collected or generated by the IoT device, or stored in the cloud. But the risk that IoT ESI data could be manipulated should not bar this evidence entirely. In the best-case scenario, the wearer, owner or controller of an IoT device can testify to authenticate the device and its data (and metadata) as a witness with personal knowledge under Rule 901(b)(1). Any analysis of the data would need to undergo a separate process to authenticate the data produced and its accuracy using 901(b)(3) (expert testimony), 901(b)(4) (distinctive characteristics, including circumstantial evidence), 901(b)(9) (system or process capable of proving a reliable and dependable result), 902(13) (certified records

⁴ Some of the below content is taken directly from or derived from The Sedona Conference Commentary on ESI Evidence & Admissibility, Second Edition (2021).

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

generated by an electronic process or system), or 902(14) (certified data copied from an electronic device, storage medium, or file).

Whether IoT ESI is or is not admissible in a particular case will depend on the facts and circumstances of that case, and the authenticating witnesses that might be needed will vary depending on the case. In some circumstances, it is possible that a single witness may be sufficient to (1) authenticate the existence of an IoT device; and (2) authenticate the existence of data that might be stored directly on the device or on a person's phone. In other circumstances, IoT ESI might be integrated into IT systems that would require testimony from a non-party witness. The bottom line is that practitioners should always consider IoT ESI discussed throughout this paper might be relevant to their clients' claims and defenses, and assess the provenance of such data. Being familiar with the "chain of custody" of IoT ESI and how it flows from IoT devices to particular storage media will be vital to practitioners. Given the proliferation in the volume of ESI and the changing nature of such "documents," actors in the legal system have and will continue to turn to technology, and experts, for assistance in identifying, analyzing, and ultimately authenticating ESI for use as evidence in both civil and criminal cases. Such technology may also be important in establishing the closely related chain of custody. While deficiencies in the chain of custody do not destroy the admissibility of the proffered evidence, they go to the weight that the jury may give to the evidence. In light of the interplay between Rule 104(a) and (b), however, deficiencies in either authentication or chain of custody may produce the same result.

B. IoT ESI Authentication Techniques and Hurdles

There are several techniques for authenticating IoT ESI evidence and data that should be discussed in additional detail. In the paragraphs that follow, we address some of the techniques, hurdles, and opportunities that come with authenticating IoT-related evidence.

1. Authentication under Fed. R. Evid. R. 901.

Federal Rule of Evidence 901 requires the proponent of evidence to "produce evidence sufficient to support a finding that the item is what the proponent claims it is." Fed. R. Evid. 901(a). This can be achieved using a non-exhaustive list of examples recited in Rule 901(b), which include testimony of a knowledgeable witnesses, distinctive characteristics, and comparisons made by expert witnesses. Fed. R. Evid. 901(b)(1), 901(b)(3), 901(b)(4).

(a) Knowledgeable Fact Witnesses

As noted above, authentication might require the procurement of knowledgeable fact witnesses. In some instances, this may involve a party witness. But in many instances, it may require non-party subpoenas of entities that process, collect and store IoT ESI, such as the technology companies that process or store the data in the cloud or elsewhere. The facts associated with the IoT ESI are not necessarily facts that judges are accustomed to, and this may create additional hurdles for authentication that are not present with more "typical" documents such as emails or run-of-the-mill corporate records. Therefore, adverse counsel may not be willing to undertake stipulations as to authentication that may occur in the more typical course.

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

(b) Distinctive Characteristics

Aside from fact witnesses, it may be possible to authenticate IoT ESI based on the distinctive characteristics of the data. IoT ESI could have unique or distinct metadata (as confirmed by a witness), or it could be characterized by unique security features or fingerprints. Section III.B of the *Sedona Conference Commentary on ESI Evidence & Admissibility* has an extensive and detailed review of how concepts such as hashing, encryption, metadata, computer forensics, and blockchain can be used to authenticate ESI data generally, and may assist in the authentication of IoT ESI. While the use of such technologies can be used to *confirm* the distinctive nature of a particular set of data, there is also some risk that the technologies can be exploited. The relatively recent emergence of blockchain, for example, is of questionable use if a weak initialization vector is generated, and also where the veracity of IoT ESI associated with that blockchain is not capable of prior, independent authentication. Therefore, the security of the data will always remain paramount.

(c) Technology Experts

Given the complexity of IoT ESI and how it is generated, distributed and stored, the sheer reliance of fact witnesses or distinctive characteristics may not be sufficient for its authentication. Over time, and with the help of experts, we will better understand the nuances of IoT ESI, how and where it is stored on servers, and what signatures, metadata or other unique attributed will be associated with SUCH data. Technology experts who are familiar with the “signature” artifact or attribute of a particular type of IoT ESI could provide additional insights into authenticity, in akin to how a handwriting expert is able to authenticate the veracity of a traditional signature. A cottage industry of IoT ESI forensic experts is already developing that will likely be useful for such purposes.

2. Self-authentication of IoT ESI under Fed. R. Evid. R. 902(13).

Federal Rule of Evidence 902(13) states that “[a] record generated by an electronic process or system that produces an accurate result, as shown by a certification of a qualified person that complies with the certification requirements of Rule (902(11) or (12)).” This rule provides an opportunity to authenticate electronic information “other than through the testimony of a foundation witness,” as the comments to the rules confirm. The purpose of the rule is to avoid the unnecessary expense of procuring foundation witnesses when parties are likely to stipulate to the authenticity of the evidence anyway.

In light of the rule, one can envision scenarios where a certification could be procured that verifies the provenance of IoT ESI without the need for live testimony. The certification can point to various IoT ESI characteristics, confirm its chain of custody, and certify that the data is authentic. Such certification may become desirable, especially in contexts where subpoenaed non- parties would prefer to avoid an authentication deposition, but would be comfortable providing a certification. An example of an entity that provides certifications of electronic data is the Internet Archive, which certifies the veracity of various copies of websites that are archived on its website. One could imagine that such transparent and ubiquitous sources of IoT ESI might undertake similar procedures. As trusted sources of IoT ESI proliferate, certifications may be accepted for authentication of IoT ESI as a matter of course, or certification authorities could emerge that could adopt procedures for purposes of providing certifications under Fed. R. Evid. 902(13). Just as such procedures could vitiate

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

the need for authentication depositions or live testimony at trial, they might also vitiate the need for Rule 104(a).

Aside from the Internet Archive example noted above, there is at least one empirical example of self-authenticating evidence that could be used in the IoT context. Pursuant to 12 V.S.A. § 1913, Vermont has a law on the books that states that “[a] digital record electronically registered in a blockchain shall be self-authenticating pursuant to Vermont Rule of Evidence 902, if it is accompanied by a written declaration of a qualified person” that meets particular requirements. This law parallels the certification requirement of Rule 902(13) of the Federal Rules, and indicates that courts may be receptive to such certifications in the context of emerging technologies, including the IoT and blockchain. It is important to note that authentication under federal and most state evidence rules presents a low bar, and as such it is equally important to note that authentication does not create a presumption of credibility for the finder of fact; rather, it merely allows the finder of fact to judge its credibility.

Hearsay and other objections to the *content* of IoT ESI might also persist, regardless of any authentication that might be achieved through a certification process. As the advisory committee notes to Rule 902(13) make clear, “[t]he opponent remains free to object to admissibility of the proffered item on other grounds—including hearsay, relevance, or in criminal cases the right to confrontation.”

3. Rule 104(a) hearing topics

Given the uncertainties pertaining to the authentication of IoT ESI, practitioners should be prepared for the possibility of a Rule 104(a) hearing to resolve the question of admissibility of IoT-related evidence. Of course, such a hearing could include evidence bearing the forms of proof discussed above, and the question that will need to be asked is whether the evidence has sufficient probative value to sustain a rational jury finding that the evidence is what the proponent claims it to be. For purposes of such hearings, a wide range of topics might bear on the answer to that question, including questions of the provenance and storage of the data that might be answered by party or non-party witnesses (such as the non-parties that sell IoT devices and store their data), and questions pertaining to the policies and procedure for collecting and retaining such data. While Rule 104(a) hearings are far from common, IoT-related evidence may be an area where such hearings occur, particularly when the use of such data is less frequent in litigation.

Quite apart from whether the IoT ESI itself is authentic, a related subject of a 104(a) hearing might be the question of whether the IoT ESI is sufficiently linked to the facts of the case. For example, a Fitbit device may have authentic data, but it may not properly be established as a Fitbit that is associated with the relevant people and facts of the case. Similarly, a non-party witness might authenticate data that is uploaded to the cloud from an IoT-enabled factory, but the relevant facts of the case may relate to the operations of a different factory.

4. Potential IoT Hearsay Issues

The two examples above illustrate classic hearsay problems that arise in the context of the IoT. IoT ESI is only as relevant as it is to bearing on the relevant actors and speakers in a case, and

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

there is a fundamental distinction between the “point of control” of IoT ESI (such as on an IT server) and the “point of use” of that data by a person or a company. Indeed, there are multiple levels of complexity from the IoT sensor itself, to the database on the server where the sensor’s data is stored, to systems that might aggregate and process such data. Therefore, the use of IoT ESI might be riddled with hearsay problems where it cannot adequately link the data to the relevant actors in the case. While beyond the scope of this paper, there are also risks of manipulation of IoT ESI or cybersecurity breaches that could occur along the chain of the IoT ESI lifecycle that occurs from the moment IoT is generated, through to the time it is collected, stored, processed and disposed or retained.

In effect, IoT devices are listening and actuating devices. Amazon Alexa literally hears information and stores the audio, and sends actuating information to other resources in the IoT universe. The other resources (devices or software) detect the fact that you are moving, your temperature, and take other measurements. All of this information can then be processed and the output can be considered another type of IoT ESI, and offered as evidence in criminal or civil litigation, just as someone who overhears a conversation might testify about an overheard conversation at trial. There will be legitimate questions as to whether such “overheard” information is sufficiently reliable to overcome hearsay or other evidentiary objections. Evidentiary tools such as the business records exception might be used to overcome these objections, but this may in turn beg further inquiry into whether non- parties maintain such data in the ordinary course of business, what their procedures are for maintaining such data, and whether a party opposing the admission of IoT ESI asserts that the source of IoT ESI sought to be admitted under that exception, or the method or circumstances of preparation indicate a lack of trustworthiness under Fed.R.Evid. 803(6)(E) .

5. *Daubert/Frye* Applicability and Experts

As noted above, experts will likely become useful for IoT ESI authentication and admissibility. These experts might be able to testify on proper methodologies of IoT ESI analysis and authentication, proper retention and deletion policies for IoT ESI, and on particular IoT technologies, such as blockchain, network infrastructure and architecture, and the mechanics of data collection using sensors. Moreover, such experts might be experts on particular applications of the IoT, such as the supply chain, connected cars, and cryptocurrency. Expertise in some such industrial and other applications may bear on the authenticity or relevance of particular types of IoT ESI, as industry practices might inform the question of whether such data is authentic.

Because such expertise is emerging, there may be risks of *Daubert* challenges to such experts. Therefore, as with all expert witnesses, it will be important to identify qualified experts with meaningful expertise that connects to the facts of any given case. And of course, the credentials of such experts should be well-established and incorporated into the record, and in any expert reports.

6. Forensic Acquisitions and Consulting

Additional forensic considerations might be worthwhile in this context. Aside from individual experts, proponents of IoT ESI admissibility might seek to use a forensic consultant or expert to help procure and authenticate IoT ESI. Such companies might be used for the purposes of on-site acquisition or collection of IoT ESI, much like how companies can assist with the collection and retrieval of traditional emails and electronic documents. There is an additional level of complexity of

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

collecting IoT ESI that might not be fully understood by opposing counsel, and relying on such an entity may be worthwhile, particularly in high-value cases. Incorporating procedures for the storage, collection, retrieval, and use of such data might also be a worthwhile component of an ESI protocol, or might be worth its own protocol, much like how review and collection of source code sometimes has its own review and production protocols.

DRAFT

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

VII. APPENDIX

A. IoT ESI ecosystem

To help provide more background on how and where potentially discoverable ESI exists within an IoT ecosystem we turn to the “The IoT Reference Model,”⁵ which describes four layers in such an ecosystem :

Device layer: These are the physical components that directly capture information via sensors or via input from people. Some devices can communicate directly to the Internet (e.g., a wi-fi enabled air-conditioner) while others require the use of a gateway that can aggregate data from multiple devices and potentially pre-process the data before transmitting it to a cloud platform (e.g., an office building with networked motion-sensing lighting and climate controls).

Networking layer: Provides network connectivity and communication capabilities for IoT devices. May include Bluetooth, wi-fi, cellular, or other networking technologies.

Service support and application support layer: an intermediary layer in the IoT ecosystem that provides technology support such as storage or computing to the IoT ecosystem.

Application layer: the software system used to control IoT devices, leverage the data gathered by the devices, or otherwise derive value from the IoT system.

Each of these layers can create and store ESI with greatly varying degrees of complexity for accessing or exporting the data.

B. Example(s) of an IoT ESI ecosystem

1. Home automation and control:

- Smart Locks
- Thermostat
- Lights
- Security camera
- Garage door opener
- Lawn sprinkler controller
- Speaker / Assistant

⁵ A. Bassi et al. (eds.), *Enabling Things to Talk*, 2013. Chapter 7 IoT Reference Model Martin Bauer, Nicola Bui, Jourik De Loof, Carsten Magerkurth, Andreas Nettsträter, Julinda Stefa, and Joachim W. Walewski.

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

- Mobile apps



2. Industrial Internet of Things

- Robotic arm
- Inventory tracking using RFID tags
- GPS based logistics
- Enterprise Resource Planning dashboards
- Facility security
- Augmented reality glasses



C. IoT Data

1. Where does it come from?

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

IoT devices often include one or more sensors. A relatively simple IoT device such as a motion-activated light would include the motion sensor alone while a complex IoT device such as a fitness tracker can include an accelerometer, gyroscope, ambient light sensor, microphone, and a heart rate sensor.

Such sensors collect a bevy of data. Some of the data is intended to inform the user or consumer, either in aggregate form or in detailed form, while other data points are only ever used for troubleshooting or system health checks by engineers. The full scale and power of the data generated by IoT devices is manifested when their independent data streams are aggregated in cloud platforms and are subjected to various analytics and machine learning processes.

2. Exports of IoT Data

A raw data export from an IoT device can include logs from sensors, file system information from on-board storage, and a variety of other data points regarding the functionality of the device. Given the immense diversity of IoT devices in the marketplace and the often-proprietary nature of their software and hardware, it can be exceedingly difficult to extract, collect, and use raw data exports.

Because IoT devices by their very nature include the ability to communicate with other devices, networks, and platforms, it is generally simpler to gather and review aggregate data sent from the device. For example, a report from a health and fitness dashboard containing aggregate data on workouts and locations is typically far more accessible than trying to gather the underlying data from the device and sensors that captured them. Similarly, trying to extract information from a smart lock on when a door was unlocked can be quite difficult. But reviewing the lock's history on an accompanying home automation app is typically much simpler.

3. Immense Data Sets

IoT generates a large amount of high-speed, varying and unstructured big data. The large amount of data that IoT generates needs to be processed before the information can be used. Because the data often comes from numerous devices or in different formats, there are several things that may be required before processing or applying any type of analytics to the data, such as:

- Standardize or transform the data to a uniform format, ensuring that that format is compatible with your application.
- Store or create a backup of the newly transformed format.
- Filter any repetitive, outdated, or unwanted data to help improve accuracy.
- Integrate additional structured (or unstructured) data from other sources to help enrich your current data set
- Store or create a backup of the newly transformed format.
- Filter any repetitive, outdated, or unwanted data to help improve accuracy.

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

- Integrate additional structured (or unstructured) data from other sources to help enrich your current data set.

D. Where does the data go? IoT data and connectivity considerations

1. Common transmission protocols

Before we look at "where the data goes", it is important to understand the protocols that are used to transmit the data. IoT devices and sensors create data. This information is then sent over the network back to the central application. Which standard is the data created in and how it will be sent over the network? For delivering this data back to the application, MQTT, HTTP and CoAP are the most common standard protocols used. Each of these 3 protocols support taking information or updates from the individual device and sending it over to a central location.

- HTTP provides a suitable method for providing data back and forth between devices and central systems. Originally developed for the client-server computing model, today it supports everyday web browsing through to more specialist services around Internet of Things devices too.
- MQTT was developed as a protocol for machine-to-machine and Internet of Things deployments. It is based on a publish / subscribe model for delivering messages out from the device back to a central system that acts as a broker, where they can then be delivered back out to all of the other systems that will consume them.
- CoAP is another standard developed for low-power, low-bandwidth environments. Rather than being designed for a broker system like MQTT, CoAP is more aimed at one-to-one connections. It is designed to meet the requirements of REST design by providing a way to interface with HTTP, but still meet the demands of low-power devices and environments.

2. Additional transmission protocols:

- AMQP - AMQP is an open standard publish/subscribe type protocol focused on the business and financial services sector. The specification describes such features as message orientation, queuing, routing, reliability and security.
- Bluetooth - Bluetooth is a wireless technology that allows the exchange of data between different devices. It uses wavelength to transmit information, it generally only works within a short distance for the devices to stay connected. Typically, devices that use Bluetooth have a maximum connectivity range of about 30 feet, and that distance is reduced when obstacles are in between the devices.
- DDS – DDS was designed by the Object Management Group (OMG), and is a protocol for real-time M2M communication which enables scalable, reliable,

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

high-performance and interoperable data exchange between connected devices independent of the hardware and the software platform. It is considered a middleware – a software layer that lies between the operating system and applications. It is not a typical IoT solution, but the DDS protocol still finds its application in the Industrial use of Internet of Things deployments.

- LwM2M – recognizing the need to manage and use remote sensors and devices in areas with intermittent connectivity and situated far from power connections, this protocol was designed to reduce power and data usage for low-power devices (resource-constrained devices). It builds on the Constrained Application Protocol (CoAP). Uses for this protocol include tracking shipper containers, and cargo railways, as well as in telecommunications, security devices, and within the automotive industry.
- WiFi – Wifi has been around for several years and is one of the most common methods of transmitting data due to the high mobile phone and laptop usage. Wi-Fi networks require devices that can send wireless signals such as mobile phones or computers. A router is used to transfer the internet connection to these devices. WiFi provides an Internet connection to nearby devices that are within a certain range. The range of a standard WiFi connection can be up to 100 meters, although, the most common range is usually 10-35 meters. WiFi uses radio waves that broadcast information on specific frequencies, such as 2.4 GHz or 5 GHz channels.
- XMPP – XMPP, or Extensible Messaging and Presence Protocol is an open XML technology for real-time communication. It powers a wide range of applications including instant messaging, presence and collaboration. It is being used in the context of the Internet of Things because it is an open community supported standard, XMPP IoT's strengths are addressing and scalability capabilities, which makes it perfect for consumer-oriented IoT deployments.
- ZigBee - networks are characterized by low power consumption, low throughputs (up to 250 kbps) and connectivity range of 100 meters between nodes. The ZigBee standard is a relatively simple, resistant to communication errors and unauthorized readings, packet data exchange protocol, which is often implemented in devices with relatively small requirements, such as microcontrollers, sensors etc. ZigBee was developed as a standard for self-configuring, short-range radio networks, intended for use in telemetry systems, for communication between various types of sensors, monitoring devices, as well as for wireless reading of measurement results from energy and heat meters, etc.

IoT devices create data that is sent to the main application to be sent on, consumed and used. This data can be sent in real time or in batches at specified times

This working draft document was created for discussion purposes only for the 2022 Midyear Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to dbl@sedonaconference.org.

or amounts. Time sensitive data has to be accurate and timely, by its nature, to fulfil the purpose of the application.

3. So, where does this data go?

After the initial creation of data, which takes place on the device, it is transferred over the Internet, typically using protocols described above, to a central system which collects, storing, organizes, and analyzes this data.

- (a) IoT data is mostly unstructured and is typically stored in databases in the cloud or on premises. A database is a collection of logically related information organized so that it can be easily accessible, managed, and updated. We usually think of a database on a computer or server that holds data.
- (b) Datacenters - a data center is a physical facility that is used to house critical applications and data. A data center's design is based on a network of computing and storage resources that enable the delivery of shared applications and data. The key components of a data center often include cybersecurity systems, firewalls, routers, servers, storage systems, and switches. Servers are a main aspect of a datacenter, and are high performance and use powerful processors. A server or a group of servers are utilized to collect, store, organize, and analyze the data