

The Sedona Conference WG1 Database Principles Brainstorming Group Recommendations Outline

Brainstorming Group Members

Jeffrey Bannon	Lilith Bat-Leah
Mathea Bulander	Courtney Fletcher
David Kessler	Shari Mauney
Gregg Parker	Anthony Pastore
Chuck Rothman	Jonathan Swerdloff

Team Leaders

Scott Clary	Laura Hunt
-------------	------------

Steering Committee Liaisons

Claudia Morgan	Eric Schwarz
Amy Sellars	

Copyright 2021, The Sedona Conference. All rights reserved.



This working draft document was created for discussion purposes only for the 2021 Annual 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.

October 6, 2021

Sedona Database Principles
Brainstorming Group Recommendations' Outline
For Consideration by WG1 Steering Committee

Overall Recommendation: Update the Database Principles

The brainstorming group agrees that the Database Principles need to be updated given the many changes in technology as well as the law that have taken place since the last time the commentary was updated in 2014. Below is a summary, per section, of the recommended changes. The Sedona Principles as well as Sedona Database Principles are included in appendices for reference.

Section-by-Section Recommendations

I. Introduction

The brainstorming group recommends an update to this section.

- a. Expand scope to address databases hosted by third parties, for example, a database underlying a SaaS or that hosts data generated by IoT devices
 - i. Impact of cloud-computing and location of data for data hosted by third parties
- b. Focus on defining and articulating what it means to produce responsive “information” rather than a “file” or “document”
- c. Consider clarifying the types of discovery issues the principles are intended to address rather than over-focusing on the definition of a “database”
- d. Address how changes in technology or even specific software affect whether the Database Principles are relevant to discovery of certain types of data
- e. Discuss scope to clarify the extent to which we are addressing edge cases or not
- f. Emphasize proportionality and nuances around assessing proportionality

II. Application of the Existing Sedona Principles to Databases and Database Information

The brainstorming group recommends updates to this section.

- a. Each of the principles already identified in the 2014 version should be retained and revisited to update references to more current technologies.
 - i. Updates to language should be considered for the sections that address Sedona Principles 3, 5, 6, and 12
- b. Recommended Updates to Current Principle Language
 - i. Principle 1: Clarifying that database preservation requirements (as discussed in section five) may be different than other data types but are no less subject to discovery requirements merely because the data resides in a database.

This working draft document was created for discussion purposes only for the 2021 Annual 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.

- ii. Principle 8: Make clear that database data is not presumptively inaccessible, and that understanding a client's databases is important to know what information within a database may indeed be inaccessible and what information may not be accessible due to the way that the database is configured or operates.
- iii. Principle 9: Databases operate differently than unstructured and semi-structured data. Reference should be made to the nature of databases, the possibility that data is not in fact stored but a stored procedure or view "creates" it on demand, or otherwise performs operations between the stored data and the end user. There may or may not be anything to preserve and what is preserved may not look or behave in an "expected" way. This is similar to, though obviously distinct from, ephemeral data. This may not be appropriate to fit under principle 9 which is focused on slack space and data fragments, but we put it here as it is logically similar. The space between how the application presents data and how the data is stored on the backend is (or can be) significant.
- iv. Principle 11: In light of section III.3 it may be duplicative to simply point out that test queries and pilot projects are both recommended, though tying the Database Principle to one of the core Principles may also be of value.

III. Principles for the Preservation & Production of Databases & Database Information

A. Section III. 1. Scope of Discovery - Absent a specific showing of need, a requesting party is entitled only to database fields that contain relevant information, and give context to such information, and not to the entire database in which the information resides or the underlying database application or database engine.

The brainstorming group recommends updates to this section.

- a. This section does not need a significant rewrite, but it might be worth it to make three clarification points:
 - i. We should consider addressing the tension between the general rule that only pre-existing documents are within the scope of discovery, and the fact that in order to discover information from databases the options are to a) run an existing report, b) create an *ad hoc* report (i.e. a new document), or c) produce an entire database. If option a does not provide all the necessary information for discovery, options b or c can pose serious challenges. We should address these challenges and look to case law and general principles that may help guide parties and the court to reach proportional discovery of databases.
 - ii. We should consider addressing the general rule that a responding party should not redact documents on the basis of relevance. The 2014 principles observe that only relevant data in a database is within the scope of discovery. But the same is not true for other types of documents. For example, if part of a memo is responsive to a discovery request, it is arguably improper to redact the portions of the memo that are irrelevant, absent some valid privilege claim. Producing only the relevant data out of a database might strike some as akin to redacting on the basis of relevance, so we should consider addressing that point. The way to do

This working draft document was created for discussion purposes only for the 2021 Annual 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.

that might be to observe that the database itself is not exactly a “document”—instead, the report created from the database is the document.

- i. There is a split in the eDiscovery community on the topic of redacting for relevance. If this is added to the commentary, recommend presenting both sides. With database discovery, given a database likely contains a combination of both relevant and non-relevant data, more emphasis may need to be placed on identifying relevant data through reporting rather than needless production of non-relevant data.
- iii. We suggest that the WG1 consider whether the Database Principles and/or case law illogically favor disclosure from the producing party regarding the structure/organization/content of a database that may contain relevant information without clearly articulating that the requesting party should also be required to be specific about their needs for the data and what it is they seek in the data. While the requesting parties’ needs may get into work product, it could be argued that this is no more work product than search terms or other facts regularly shared between parties in the course of discovery.

B. Section III. 2. Accessibility and Proportionality - Due to differences in the way that information is stored or programmed into a database, not all information in a database may be equally accessible, and parties should therefore apply proportionality to each component of a database to determine the marginal value of the information to the litigation and the marginal cost of collecting and producing it.

The brainstorming group recommends an addition to this section.

- a. This section should add a new comment containing practical guidance on how parties can address—and hopefully resolve—proportionality issues. This guidance could include:
 - i. In lieu of extracting relevant data from the database, the responding party could make the entire database available to the requesting party on a terminal at the responding party’s place of business. We would need to clarify whether we believe this completely discharges the responding party’s discovery obligations (or whether the responding party would need to provide the requesting party with assistance).
 - ii. If it is unclear to the requesting party how to extract relevant data, the responding party could explain how it plans to use the data. The responding party might hesitate to do so because it might claim that this information is its work product. But we view this information as akin to search terms/key words, which are not typically viewed as protectable work product.

C. Section III. 3. Use of test queries and pilot projects - Parties should use objective information, such as that generated from test queries, pilot projects, and interviews with persons with relevant knowledge to ascertain the burden and benefits to collect

This working draft document was created for discussion purposes only for the 2021 Annual 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.

and produce information stored in databases and to reach consensus on the scope of discovery.

The brainstorming group recommends updates to this section.

- a. This section should add a new comment encouraging requestors to specify the types of information they want from a database:
 - i. Instead of issuing a blanket request for a database schema or for all information contained within a database, requestors should be encouraged to request only that information necessary to address their matter.
 - ii. Criteria for determining the accessibility of information in a database should be based on a proportionality analysis rather than how difficult it is to access the database. Unless accessibility is impossible, proportionality is a more appropriate way of determining what information should be produced.

D. Section III. 4. Validation – A responding party should use reasonable measures to validate that its collection from the database is both reasonably complete and did not inadvertently modify the ESI.

The brainstorming group recommends including additional examples to this section.

- a. This section should add additional examples of the ways that data can be validated when extracting from a database:
 - i. Verify the data is actually the data you expect.
 - ii. Check for truncating of data when results are converted to a different format.
 - iii. Verify contents of fields – data within the field may be different than what the name implies.
 - iv. Time Zone conversions – validate time zones as they are not always UTC.

E. Section III. 5. Data Authenticity and Admissibility - The proper validation of collection from a database does not automatically make the substantive information stored in the database authentic, admissible or true. These are separate issues that need to be analyzed by the appropriate decision makers.

The brainstorming group recommends updates to this section.

- a. Most database discovery is a “record of regularly conducted activity” under 803(6), admissible does not equal true
- b. Production should be fielded, searchable with metadata (keys and codes)
- c. The producing party has the right to limit production by relevance and privilege, but the requesting party should not have to pay for the exercise of that right
- d. Additional Thoughts from the Team
 - i. Under the Federal Rules of Evidence, to be admitted into evidence, information from a database must be relevant, authentic, and admissible (principally as a

This working draft document was created for discussion purposes only for the 2021 Annual 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.

record of a “regularly conducted activity” under Fed.R.Evid. 803(6)). Meeting these standards is not a guarantee that the information is “true” only that it is admissible.

- ii. Almost all database information sought to be admitted today comes from a database maintained in the regular course of business: personnel and wages records, sales and financial transactions, production records, etc. When information is produced by a party under Rule 34 from a database maintained in the regular course of its business, neither authenticity of the data as produced nor hearsay should be valid issues. As Judge Grimm said, “... the inability to get evidence admitted because of a failure to authenticate it almost always is a self-inflicted injury which can be avoided by thoughtful advance preparation.” *Lorraine v. Markel Am. Ins. Co.*, 241 F.R.D. 534, 542 (D. Md. 2007). The same can be said for establishing that the information is a record of regularly conducted activity.

F. Section III. 6. Form of Production - The way in which a requesting party intends to use database information is an important factor in determining an appropriate format of production.

The brainstorming group recommends updates to this section.

- a. The default form of production for database information should be “a fielded and electronically searchable format preserving metadata values, keys and field relationships.” Ball-Lawyers Guide to Forms of Production_Ver.20140512_TX.pdf at 32.
- b. If the responding party ordinarily maintains the information it is producing in a way that makes it searchable by electronic means, the information should not be produced in a form that removes or significantly degrades this feature.
- c. Rule 34, Committee Notes on Rules—2006 Amendment
Dynamic data may need to be produced in native format, or in a modified format in which the integrity of the data can be maintained while the data can be manipulated for analysis. If raw data is produced, appropriate applications, file structures, manuals, and other tools necessary for the proper translation and use of the data must be provided.

G. Potential New Sections for Database Principles and Additional Thoughts

The brainstorming group recommends the following additions.

- a. Potential New Sections for Database Principles
 - i. Primer on Databases: In the brainstorming group we discussed that some of the audience for the Database Principles may not be as well-versed in the technical background of databases as others. The group recommends an appendix that has a primer that provides an overview of key concepts and key

This working draft document was created for discussion purposes only for the 2021 Annual 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.

terms of art related to databases so that all readers can have the benefit of this baseline knowledge.

- ii. Database Discovery JumpStart Outline for Requesting and Receiving Parties: While not to encourage discovery on discovery, the group recommends adding an appendix that contains key questions for consideration by both a requesting and receiving party when preparing for discussions about database discovery.

b. Additional Thoughts on Issues to Address

- i. If we were trying to do a collection from a filesystem and did not have methodologies for that already, a filesystem is a database and we would need to apply the same techniques and the same processes. Having said that, operating systems and the Microsoft suite are technically databases, they are just databases for which there are already controls and processes to appropriately collect and produce and which would cause confusion among readers. Getting a file tree listing, you are querying the file system database. In the database definition, revisiting the distinction between data which is covered by these principles and which isn't would be of value.
- ii. Hierarchy of desired production formats - also should make clear that "native file" is generally the wrong framework to use for database data.
 - i. Native
 - ii. JSON
 - iii. CSV with agreed upon delimiters
 - iv. Reports in a csv/tsv or excel format
 - v. PDF
 - vi. Excel workbook
- iii. There are a number of existing controls for filesystem based productions to ensure consistency, clarity, and usability such as bates numbers, page counts, and MD5 hashes. Database data does not always lend itself to these solutions and as such, proposing solutions for the analogous issues could be valuable. It's currently covered to an extent in sections II. (C)(5) and II. (C)(6) but more clarity might be valuable. Additionally – hashing of records and transmission of files.
- iv. "Creating" evidence for the convenience of the requesting party or the court – joining tables for production, developing reports, if needed.
- v. Counterpoint: Slack, Twitter, etc takeouts – you have no idea what data you are NOT getting – you don't know what the data in the database is.
- vi. Add the "talk nerdy to me" part and why not to use a DBA. ACCESS databases aside, anything else needs someone specific who focuses on this stuff.
- vii. Ethical responsibility of lawyers to be technically competent and if they are not, to meet it the need to bring in an expert, and the attendant supervisory responsibilities.
- viii. FRE 902(13) and (14) application, especially re: the original section "Data Authenticity and Admissibility"

This working draft document was created for discussion purposes only for the 2021 Annual 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.

Appendix 1

The Sedona Conference Database Principles Addressing the Preservation and Production of Databases and Database Information in Civil Litigation (September 2014)

1. Scope of Discovery

Absent a specific showing of need, a requesting party is entitled only to database fields that contain relevant information, and give context to such information, and not to the entire database in which the information resides or the underlying database application or database engine.

2. Accessibility and Proportionality

Due to differences in the way that information is stored or programmed into a database, not all information in a database may be equally accessible, and parties should therefore apply proportionality to each component of a database to determine the marginal value of the information to the litigation and the marginal cost of collecting and producing it.

3. Use of Test Queries and Pilots

Parties should use objective information, such as that generated from test queries, pilot projects, and interviews with persons with relevant knowledge to ascertain the burden and benefits to collect and produce information stored in databases and to reach consensus on the scope of discovery.

4. Validation

A responding party should use reasonable measures to validate that its collection from the database is both reasonably complete and did not inadvertently modify the ESI.

This working draft document was created for discussion purposes only for the 2021 Annual 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.

5. Data Authenticity and Admissibility

The proper validation of collection from a database does not automatically make the substantive information stored in the database authentic, admissible or true. These are separate issues that need to be analyzed by the appropriate decision makers.

6. Form of Production

The way in which a requesting party intends to use database information is an important factor in determining an appropriate format of production.

DRAFT

This working draft document was created for discussion purposes only for the 2021 Annual 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.

Appendix 2

The Sedona Principles, Third Edition: Best Practices, Recommendations & Principles for Addressing Electronic Document Production (October 2017)

Principle 1 Electronically stored information is generally subject to the same preservation and discovery requirements as other relevant information.

Principle 2 When balancing the cost, burden, and need for electronically stored information, courts and parties should apply the proportionality standard embodied in Fed. R. Civ. P. 26(b)(1) and its state equivalents, which requires consideration of 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.

Principle 3 As soon as practicable, parties should confer and seek to reach agreement regarding the preservation and production of electronically stored information.

Principle 4 Discovery requests for electronically stored information should be as specific as possible; responses and objections to discovery should disclose the scope and limits of the production.

Principle 5 The obligation to preserve electronically stored information requires reasonable and good faith efforts to retain information that is expected to be relevant to claims or defenses in reasonably anticipated or pending litigation. However, it is unreasonable to expect parties to take every conceivable step or disproportionate steps to preserve each instance of relevant electronically stored information.

Principle 6 Responding parties are best situated to evaluate the procedures, methodologies, and technologies appropriate for preserving and producing their own electronically stored information.

Principle 7 The requesting party has the burden on a motion to compel to show that the responding party's steps to preserve and produce relevant electronically stored information were inadequate.

Principle 8 The primary sources of electronically stored information to be preserved and produced should be those readily accessible in the ordinary course. Only when electronically stored information is not available through such primary sources should parties move down a continuum of less accessible sources until the information requested to be preserved or produced is no longer proportional.

Principle 9 Absent a showing of special need and relevance, a responding party should not be required to preserve, review, or produce deleted, shadowed, fragmented, or residual electronically stored information.

This working draft document was created for discussion purposes only for the 2021 Annual 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.

Principle 10 Parties should take reasonable steps to safeguard electronically stored information, the disclosure or dissemination of which is subject to privileges, work product protections, privacy obligations, or other legally enforceable restrictions.

Principle 11 A responding party may satisfy its good faith obligations to preserve and produce relevant electronically stored information by using technology and processes, such as sampling, searching, or the use of selection criteria.

Principle 12 The production of electronically stored information should be made in the form or forms in which it is ordinarily maintained or that is reasonably usable given the nature of the electronically stored information and the proportional needs of the case.

Principle 13 The costs of preserving and producing relevant and proportionate electronically stored information ordinarily should be borne by the responding party.

Principle 14 The breach of a duty to preserve electronically stored information may be addressed by remedial measures, sanctions, or both: remedial measures are appropriate to cure prejudice; sanctions are appropriate only if a party acted with intent to deprive another party of the use of relevant electronically stored information.