Cellebrite

Physical Analyzer 7 VS. Cellebrite Inseyets.PA 10

Comparing Output Results



Introduction

Since it's introduction, Physical Analyzer (PA) has gone through some significant enhancements, moving from a tool that handled everything in RAM, to a brand-new infrastructure built upon a PostgreSQL database to work faster and make for a better user experience.

As with any journey, there are always challenges. A persistent issue relayed by users is that processing an extraction in Cellebrite Inseyets.PA (formerly known as Physical Analyzer Ultra - hereafter referred to as Inseyets.PA) resulted in a varying number of records decoded when compared to the same extraction in PA7.x. This is the kind of data integrity issue that can cause issues and concerns and it's important to understand it fully and learn how to avoid any confusion.

This document is designed to assist organisations in recognizing the differences between the data decoded in PA7.x and Inseyets.PA and the reasons that differences may occur.

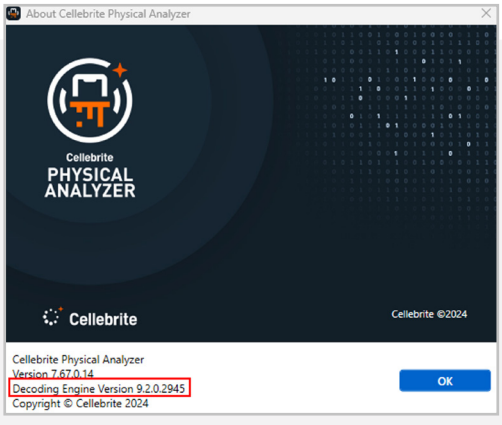
What does Cellebrite Inseyets include?

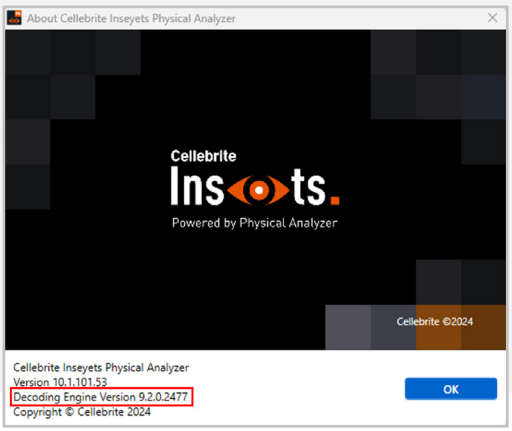
Both PA7.x and Inseyets.PA rely on a separate code block called the Decoding Engine which handles the processing and parsing functions. The main benefit of having the Decoding Engine sit outside of the main tool is that it is accessible by any Cellebrite application that requires decoding capabilities and provides consistency between them.

Since the release of PA Ultra 8.2, we have always ensured it is the same Decoding Engine used in the PA7 and Inseyets.PA releases. This was meant to ease the transition from PA7 to Inseyets.PA and make sure that regardless of which version of PA in use, the decoded data was consistent.

In most cases, a newer Decoding Engine will decode more data than its predecessor. Comparing two versions of PA that do not have the same Decoding Engine is expected to have discrepancies and is the first thing to check. Ideally, both PA versions being compared should share the same Decoding Engine version, which can be found in the "About" screen of both tools.

These screenshots below show that even though the version of PA is different, the Decoding Engine Version is 9.2 in each.





Decoding and Enrichment Options

PA presents you with Decoding and Enrichment Options when you first start a project, and it is important to ensure both tools are utilizing the same options for a fair comparison.

This table shows the available features in each version of PA that could affect the number of records parsed.

OPTION	PA 7	INSEYETS.PA
Recover Data from Archives	✓	✓
Carve Locations	✓	✓
Selective Decoding	✓	Not Yet
Cryptocurrency Scanner	✓	✓

Recover Data from Archives

This enrichment option will search inside any zip or tar files found in the extraction and may result in finding additional records. Both tools should treat archives the same, but if this option is selected in one version and not the other, the difference could be profound.

Cryptocurrency

This is the internal Cryptocurrency feature which will search for any strings that appear to be a Cryptocurrency wallet address. As stated above, the results of running this feature should be the same in each tool but running it in only one version could result in a difference in the number of artifacts found.

Carve Locations

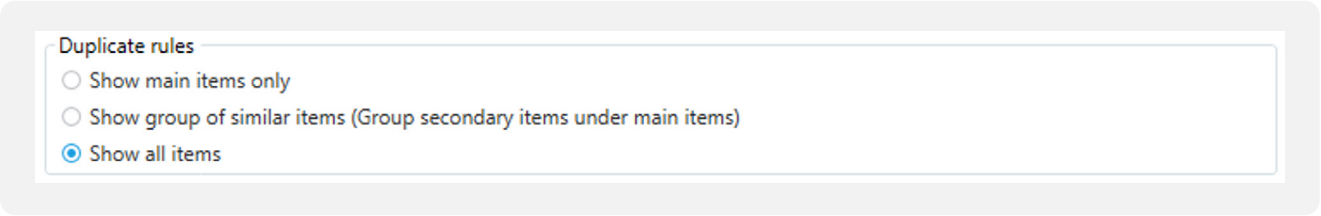
This enrichment option will seek any data on the device that appears to be a latitude/longitude and timestamp. This feature has further options such as carving for locations around a specific GPS coordinate or around the most visited locations. It may result in thousands of results and must therefore be handled the same in each tool.

Selective App Decoding

The Selective Decoding option in PA7 allows you to decode only the apps that are important to you while excluding other applications. This provides much faster access to data but only provides a fraction of the data on the device. As Inseyets.PA doesn't have this feature yet, all applications will be decoded, resulting in many more results. For comparison purposes, this option should be omitted.

Deduplication Options

One of PA's longstanding features is the options related to displaying duplicate data.
The same options are in both PA and Inseyets.PA and can be found in the settings.



Show Main Items Only

PA will only show one version of the duplicated item and hide the rest. For example, a user takes a photograph and sends it to several contacts.

With **Show Main Items Only** selected, the Images tab will only include one version of the image and it will not be obvious that there are duplicates. As only one version of the image is being shown, **the omitted duplicates do not factor into the count.**

Show Group of Similar Items

PA will only show one version of the duplicated item, but the rest will be viewable. Using the same example as above, a user takes a photograph and sends it to several contacts.

With **Show Group of Similar Items** selected, the Images tab will only include one version of the image but will show that several duplicates exist and allow you focus on them if required.
The duplicated items will be included in the count.

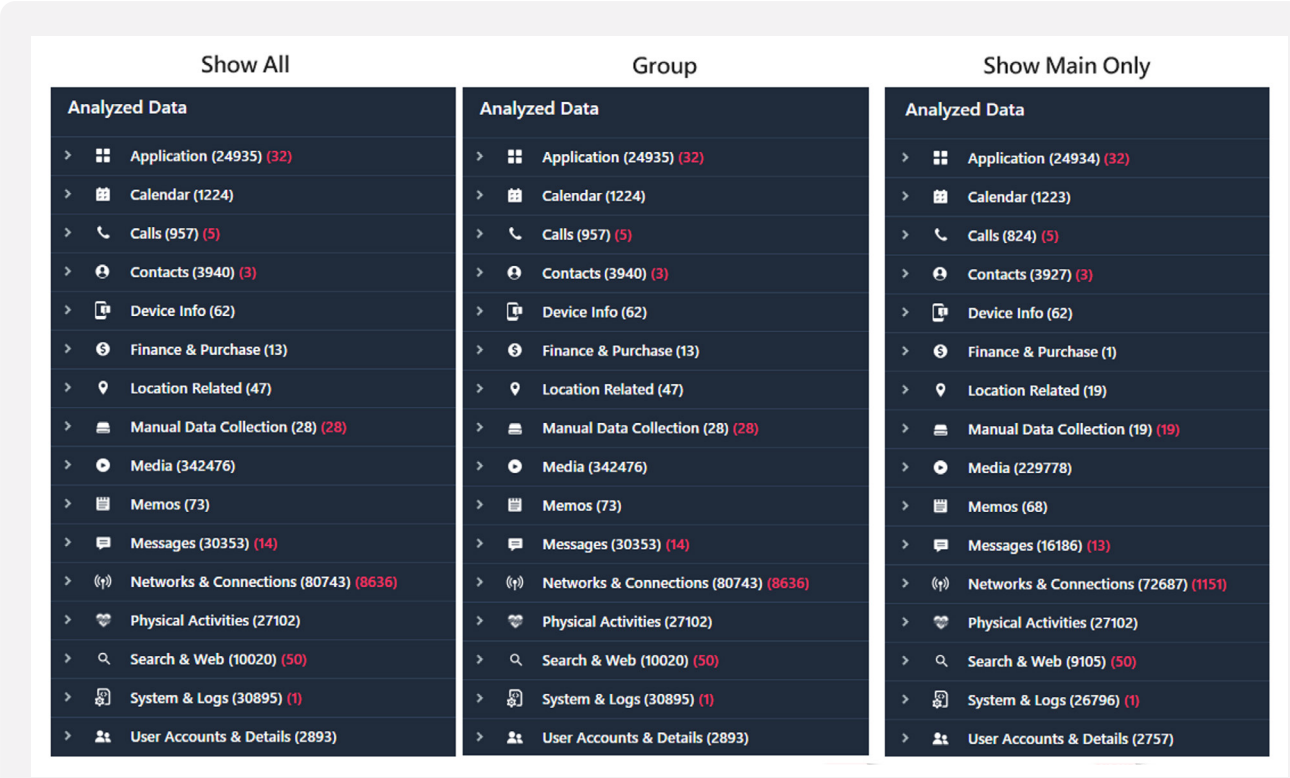
Show All Items

PA will show all versions of the duplicated item. With the same example above, **PA will show all copies of the image and the duplicated items will be included into the count.**

You may have noticed in PA7 that changing this option requires the case to be reparsed because the deduplication rule is applied at the time of decoding.

But in Inseyets.PA, this option can be modified on the fly. All records are processed at the time of decoding and the deduplicate rule is applied when the results are being shown. This means that you can alter this setting and immediately see the difference it makes without having to reprocess the case.

A difference in the deduplication rules being applied between versions will likely result in a difference in the record count, as can be seen in the below screenshot where the same device is analyzed using the different deduplication options.



While Show All and Group by Main still present the same number of records, the Show Main Only count shows substantially less records in many models.

This deduplication setting can be easily understood to result in large differences, as all duplicate records are omitted. However, there have also been small changes in the deduplication logic that may also account for differences even when the deduplication rules are set the same.

A great example of this is within the Messages > Chats model when using the Group by Main deduplication rule.

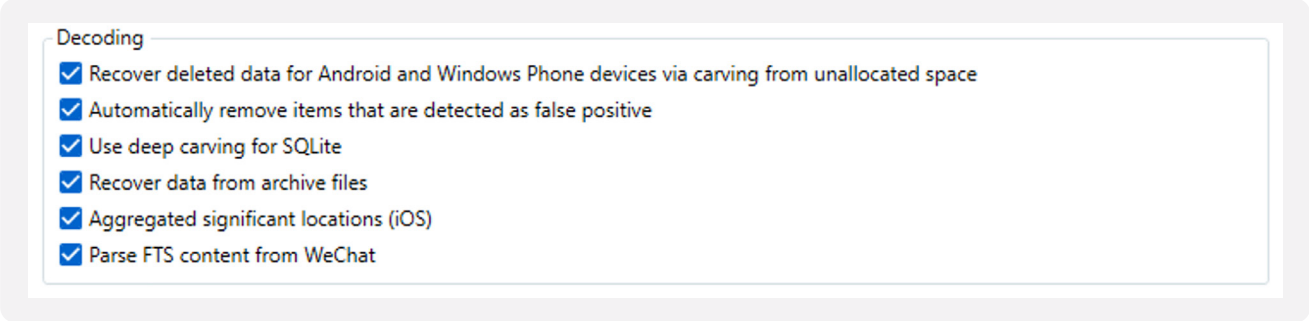
PA 7	INSEYETS.PA
Chats (896) (1) (111166 messages)	Chats (896) (1) (113950 messages)

In these screenshots, PA7.x and Inseyets.PA both have 896 Chats, but Inseyets.PA is showing 2,784 more Messages. This is not because there is a different number of messages found, but it is because of differences in how the counts are displayed.

In PA7, the Messages count of 111,166 is the number of messages found **after** deduplication, whereas in Inseyets.PA, the count of 113,950 is **before** deduplication.

Decoding Options

In addition to the Decoding and Enrichment Options available during case intake, PA has additional Decoding Options available from the Settings window and when comparing results between PA7 and Inseyets.PA it is important to ensure these settings are also aligned.



Recover deleted data from Android and Windows Phone devices via carving from unallocated space
This option is only relevant for older devices where a physical extraction was obtained, and PA can carve records from unallocated space on the image.

Automatically remove items that are detected as false positive
PA will detect False Positives from the carved data and exclude them.

Use deep carving for SQLite
This option is disabled by default. If it is enabled, each parser will run additional logic and process free pages from databases differently, resulting in many more records.

This feature may generate a large amount of false positive records.

Recover data from archive files
This is the same option as is presented when opening a new project. The difference being that this is the default behaviour whereas the option during the Decoding and Enrichment Options is relevant for that single case only.

Aggregated significant locations (iOS)
iOS devices may contain hundreds of thousands of location records and the sheer number can impede examinations and reporting this important artifact.

The Aggregation option (enabled by default) will reduce the number of records by grouping the results based on physical and temporal proximity. For example, rather than show 1000 records at the owner's home address, PA will group these 1000 records into a single marker that has a Start Time equal to the first record and an End Time equal to the last record.

Parse FTS content from WeChat
FTS (Full Text Search) is an indexing service related to SQLite databases aimed at speeding up the searching function.

FTS is only available on databases whose developers have opted to use it. Since it doesn't result in entire messages, the option to process this data is left to the user to decide.

Other Changes

Other changes introduced in Inseyets.PA were made to the grouping of artifacts or to give additional focus and help where it would be beneficial. These changes resulted in some nodes within the Analyzed Data tree appearing different to what is seen in PA7. It's less about a different number of records found, and more about where in the user interface they are displayed.

Extraction Comparison



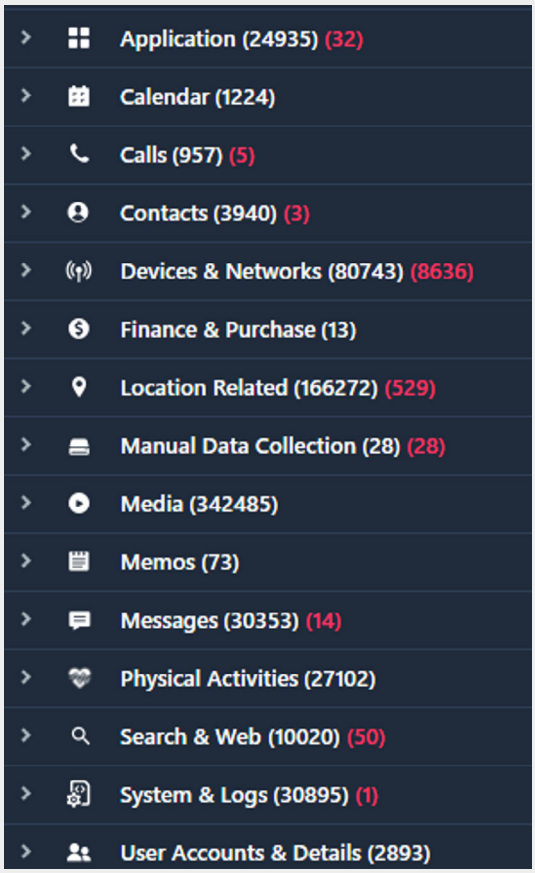
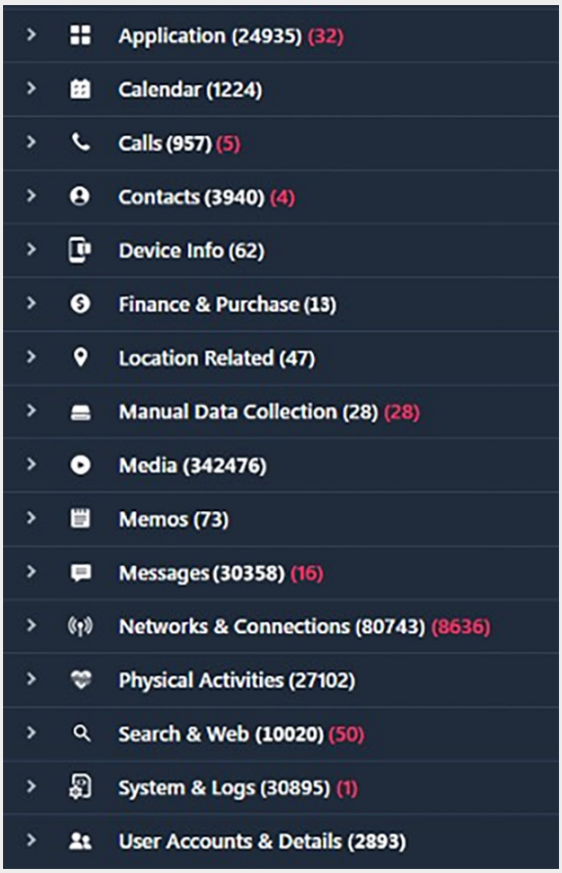
Extraction

Device : iPhone 14 Pro (iOS 16.3)
Type : Full File System
Size : 155GB

Machine

Processor : Ryzen ThreadRipper 3960x (24-Core)
RAM : 128GB DDR4
Hard Drive(s) : NVMe for System Drive.
NVMe for PA Database.
Extraction on External T5.

PA 7 7.66	INSEYETS.PA (10.1)
HashSets: On **Carve Location:** Off **Recover Data from Archives:** Off **Selective Apps Decoding:** Off **Media Classification:** Off **Cryptocurrency Scanner:** Off **Deduplication Rule:** Show All Items **Use Deep Carving for SQLite:** Off **Total Load Time:** 2:33hrs	**HashSets:** On **Carve Location:** Off **Recover Data from Archives:** Off **Media Classification:** Off **Media Origin:** Off **Cryptocurrency Scanner:** Off **Chainalysis Enrichment:** Off **Deduplication Rule:** Show All Items **Use Deep Carving for SQLite:** Off **Total Load Time:** 2:50hrs

PA 7.66	INSEYETS.PA (10.1)
	

As can be seen from the above screenshots, many of the items are identical between versions. A more detailed overview is provided here:

PA 7.66	NODE	INSEYETS.PA (10.1)	DELTA
24,935	Application	24,935	-
1,224	Calendar	1,224	-
957	Calls	957	-
3,940	Contacts	3,940	-
N/A	Device Info	62	-
80,670	Devices & Networks	N/A	-
13	Finance & Purchase	13	-
166,272	Location Related	47	-166,225
28	Manual Data Collection	28	-
342,485	Media	342,476	-9
73	Memos	73	-
30,353	Messages	30,353	-
N/A	Networks & Connections	80,743	-
27,102	Physical Activities	27,102	-
10,020	Search & Web	10,020	-
30,895	System & Logs	30,895	-
2,893	User Accounts & Details	2,893	-

Here, we dig in deeper to each artifact type.

APPLICATION	**PA7**: 24,935	**INSEYETS.PA**: 24,935	-
No differences.			

CALENDAR	**PA7**: 1,224	**INSEYETS.PA**: 1,224	-
No differences.			

CALLS	**PA7**: 957	**INSEYETS.PA**: 957	-
No differences.			

CONTACTS	**PA7**: 3,940	**INSEYETS.PA**: 3,940	-
No differences.			

DEVICE INFO	**PA7**: N/A	**INSEYETS.PA**: 62	-
Device Info is a new Analyzed Data node containing information about the device. This node is similar to the information found on the Device Summary in PA7 but is now presented in a table which can be searched, filtered, tagged and reported on.			

DEVICES & NETWORKS	**PA7**: 80,670	**INSEYETS.PA:** N/A	-
The Devices & Networks node has been replaced by the Networks and Connections node in Inseyets.PA. It contains Cell Towers, Wi-Fi Networks, Device Events, Devices and Device Connectivity. A direct comparison between "Devices and Networks" and "Networks and Connections" will be made later in the document.			

FINANCE & PURCHASE	**PA7**: 13	**INSEYETS.PA**: 13	-
No differences.			

LOCATION RELATED	**PA7**: 166,272	**INSEYETS.PA**: 47	**DELTA:** -166,225
At first glance, the Location Related node is the most considerable difference; however, it is important to remember that Inseyets.PA has placed a renewed focus on Location data and has moved most location data to its own tab, accessible from the main menu that runs vertically down the left side of the window. The reasoning behind the move to its own tab was to improve the semantics used to describe the locations and improve the overall user experience.			

While this move does make it harder to compare the numbers between the Locations Node and the Locations Tab, it's not impossible.

Within the Analyzed Data tree in PA7, the Locations node contains Locations and Journeys, and the locations are grouped by source application. In Inseyets.PA locations are grouped into categories such as Visited, Point of Interest and Media.

This new grouping allows an examiner to quickly find and focus on the location data that matters most while excluding records that are irrelevant to the goals of the exam.

PA 7		INSEYETS.PA
47 (56 Waypoints)	Journeys	47

Whereas in PA7, locations are grouped by their source application, (i.e. Apple Maps, Native, Weather etc.), Inseyets.PA groups into categories such as Visited, Point of Interest and Media.

This allows an examiner to quickly find and focus on the location data that matters most while excluding records that are irrelevant to the goals of the exam.

PA 7 7.66		INSEYETS.PA	
Locations	166,216	Visited	144,259
		Point of Interest	1564
		Media	8,392
		Other	13,196
		Total	166,716

As can be seen, Inseyets.PA reports 500 more records that PA7 which deserves some further clarification.

Many of the data sources between the two applications matched perfectly and will not be investigated any further.

All data sources that presented differently will be broken down below.

CACHE_ENCRYPTED B.SQLITE	**PA7:** 13072	**INSEYETS.PA:** 13095	**DELTA: +23**

Inseyets.PA recovered an additional 23 deleted Cell Tower records.

PA 7		PA10	
Select All		Select All	
Deleted: Yes	(75)	Deleted: Yes	(98)
Deleted: No	(962)	Deleted: No	(962)

GOOGLE MAPS	**PA 7**: 4	**INSEYETS.PA**: 5	**DELTA:** +1

Inseyets.PA was able to recover one additional deleted record.

NATIVE MESSAGES	**PA 7**: 24,935	**INSEYETS.PA**: 24,935	**DELTA:** -1

PA7 included a record with no actual location data.
This record was omitted from Inseyets.PA due to not containing any actual location information.

EMAILS	**PA 7**: 24,935	**INSEYETS.PA**: 24,935	**DELTA:** -7

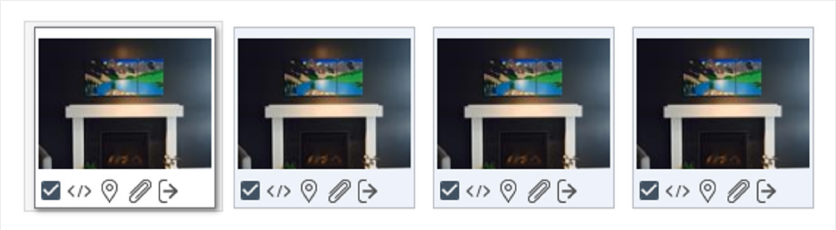
PA7 returned 7 records which did not contain any actual location data.
Inseyets.PA had omitted these records.

MEDIA FILES	**PA 7**: 24,935	**INSEYETS.PA**: 24,935	-

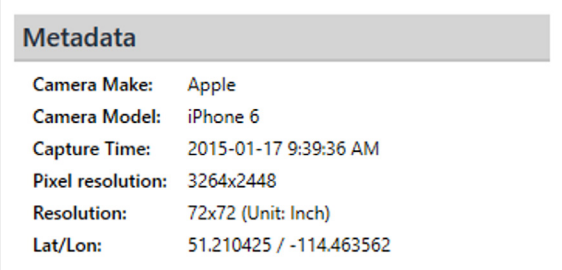
Regardless of deduplication settings, PA7 was deduplicating the location data obtained from media files whereas Inseyets.PA was not.

Media items can be duplicated many times across a file system as part of both user actions and normal file system operation. For example, Photographs that were taken and subsequently sent by a messaging service.

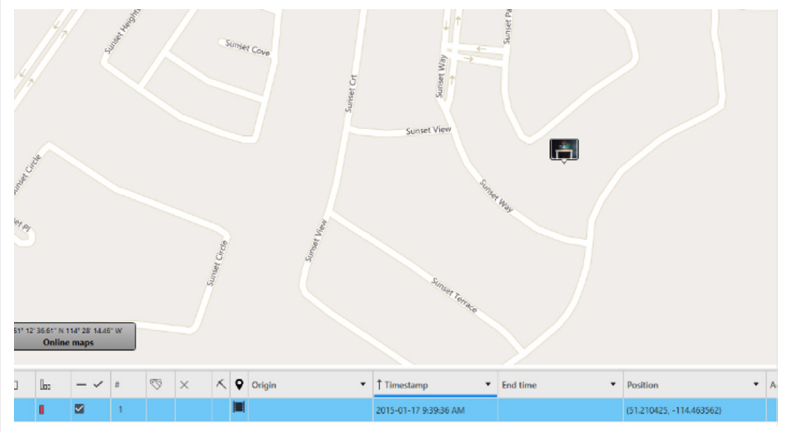
In this case, PA7 would show 1 location record, even though the image exists multiple times within the Thumbnail gallery, each with a GPS tag.



The Metadata for the images shows the date, time and location of the image.



But when Locations are filtered to this date, only 1 record is found, even though deduplication was not being utilized.



Looking in PA10 and filtering Media Locations to the same timestamp results in 4 identical records.

Timestamp	End time	Position	Map Address	Description
2015-01-17 9:39:36 AM		(51.210425, -114.463562)		FullSizeRender.jpg
2015-01-17 9:39:36 AM		(51.210425, -114.463562)		FullSizeRender.jpg
2015-01-17 9:39:36 AM		(51.210425, -114.463562)		FullSizeRender.jpg
2015-01-17 9:39:36 AM		(51.210425, -114.463562)		FullSizeRender.jpg

While this may seem strange, the link back to the image is important, and immediately shows all instances of the photo which is at this location.

Overall, most differences between PA7 and Inseyets.PA relates to the inclusion of a location record for each duplicated media item.

While there are some models where PA7 has more records that Inseyets.PA, these are blank and are omitted from Inseyets.PA on purpose.

MANUAL DATA COLLECTION	PA7: 28	INSEYETS.PA: 28	-
No differences.			

MEDIA	PA7: 34,2485	INSEYETS.PA: 34,2476	DIFFERENCE: -9

PA 7	PA10
	

As can be seen above, the difference in the Media Count could be broken down to 6 images and 3 videos missing from Inseyets.PA.

These items were found to be hard links within a tar archive which was part of the WhatsApp backup data.

Hard links are used in Tar files to reduce duplicated data within the archive and reduce the files overall size. For example, if an image exists 3 times within the data being written to a Tar file, then by default the file will only be written once and the additional 2 files will be written as pointers (hard links) to the full version of the file in the archive.

Within PA7, these hard links are shown as 0-byte files within their respective Images and Videos model. Within Inseyets.PA these are shown as Shortcuts under the Shortcuts model.

Ultimately, in Inseyets.PA, only the full file in the archive exists in the Images/Videos model and the hard links do not.

More information about Hard Links can be found here: www.gnu.org/software/tar/manual/html_node/hard-links.html

MEMOS	PA7: 73	INSEYETS.PA: 73	-
No differences.			

MESSAGES	**PA7**: 30,353	**INSEYETS.PA:** 30,353	-
No differences.			

NETWORKS & CONNECTIONS	**PA7**: N/A	**INSEYETS.PA:** 80,743	-
The Networks & Connections node is a new label in Inseyets.PA, taking the place of the Devices & Networks tab. It contains Cell Towers, Wi-Fi Networks, Device Events, Devices and Device Connectivity. A direct comparison between "Networks and Connections" and the "Devices and Networks" will be made later in the document.			

PHYSICAL ACTIVITIES	**PA7**: 27,102	**INSEYETS.PA**: 27,102	-
No differences.			

SEARCH & WEB	**PA7**: 10,020	**INSEYETS.PA**: 10,020	-
No differences.			

SYSTEM & LOGS	**PA7**: 30,895	**INSEYETS.PA**: 30,895	-
No differences.			

USER ACCOUNTS & DETAILS	**PA7**: 2,893	**INSEYETS.PA**: 2,893	-
No differences.			

Devices & Networks vs Networks & Connections

As mentioned above, Inseyets.PA has renamed the Devices & Networks node to Networks and Connections, but the data it contains is still the same.

PA 7	INSEYETS.PA
Devices & Networks	Networks & Connections
80,743	80,743
No differences.	

Conclusion

At first glance, it can seem that both PA7 and Inseyets.PA 10 are parsing some artifacts differently. But all things being equal (specifically options, enrichments, and settings), this isn't really the case, and the difference can be explained in large part to the presentation layouts between the two application versions.

The shared Decoding Engine ensures that the data being parsed is identical between tools, including PA7 and Inseyets.PA, and, while there are differences in how the applications display some of the decoded data, there should be no differences between the tools that cannot be explained.

Of course, the examples outlined above are from a single case and is non-exhaustive. You may see other differences in your own tests that aren't described here. But with a little investigation, the reason behind the differences should become apparent.

If you have any concerns, questions, or require assistance, feel free to contact us at **stumpus@cellebrite.com**. Our team of experts is ready to help you navigate through this transition. Thank you for choosing Cellebrite, and we eagerly anticipate working together to accelerate justice.

Cellebrite

About Cellebrite

Cellebrite's (Nasdaq: CLBT) mission is to enable its customers to protect and save lives, accelerate justice, and preserve privacy in communities around the world. We are a global leader in Digital Intelligence solutions for the public and private sectors, empowering organizations in mastering the complexities of legally sanctioned digital investigations by streamlining intelligence processes. Trusted by thousands of leading agencies and companies worldwide, Cellebrite's Digital Intelligence platform and solutions transform how customers collect, review, analyze and manage data in legally sanctioned investigations.

- To learn more visit us at www.cellebrite.com
- Contact Cellebrite globally at www.cellebrite.com/contact

