

Israeli Supreme Court Database 1997-2022 and Beyond

Final Report

31 May 2023

PIs: Dr. Ayelet Sela, Bar Ilan University Faculty of Law

Dr. Jonathan Schler, Holon Institute of Technology, School. of Computer Sciences & Bar Ilan University, RESPONSA Project

Overview

This grant supported the creation of a high-quality data infrastructure of all publicly available and digitally archived Israeli Supreme Court cases that were opened between the years 1997 and 2022. The database includes both the full texts of the Court's final decision in each case and its metadata, comprising a large number of automatically extracted and computed variables pertaining to the case. The database is meant to serve as a resource for researchers and policy makers by enabling computational and empirical analyses of judicial behavior and decision-making, the state of the law, and the Supreme Court as an institution; as well as to advance the creation of natural language processing (NLP) models for Hebrew legal data. Methodologically, the project relies on the development of an automatic method for extracting reference data from the online case database of the Israeli Supreme Court and providing two formats for accessing and using the data.

Description of the Database

The database includes data on all digitally archived Israeli Supreme Court cases that were opened between 1 January 1997 and 31 December 2022 - a total of 249,115 cases. The database includes the full text of the final decision of each case, as well as up to 135 metadata and biographical variables.

- 1) **The full text of the final decision in each case:** For each case, the database includes the full text of the final decision of the court (if available) or the latest decision rendered in the case (if a final judgment is not available), thereby creating the first comprehensive corpus of Israeli Supreme Court final decisions. The full text is available in two formats:
 - a) HTML formatted text preserving the original formatting of the digitally published decision, typically usable for naive reading.

- b) Plain text without any formatting, typically usable for text processing.
- 2) **80 case-related meta-data variables:** including all the scrapable data from the Court’s website, data automatically extracted from the text of the decision, and data computed from the above-mentioned variables. There are 80 meta-data variables pertaining to the cases and decisions, which are summarized in Table 1 in the Appendix. Naturally, cases include information only on variables that are relevant to them, thus, not all cases include data on all variables. All variables were subject to validation; relevant case variables have also been normalized. Appendix 3 describes the validation and normalization procedures.
 - 3) **55 judge-related biographical variables:** We supplemented the database with biographical data on all Israeli Supreme Court Justices, a total of 55 demographic variables which are summarized in Table 2 in the appendix. The biographical variables originate from a database created by the researcher Alon Jasper,¹ who generously shared it with us. As with case data, some variables are relevant only to a subset of the justices.

The database is stored as individual JSON files for each case that contain all the above mentioned data points. The database is currently located in an AWS cloud-based server.

Summary Statistics

In the table below we report several key attributes of the database.

Item	Count
Cases	249,115
Meta-data variables	135
Full text documents*	198,176
Full text total words (excluding metadata)	133,397,671
Corpus size	18.6 GB
Years covered	25

* For some cases only some metadata variables are available since the court did not make the full text of the decisions and other identifying data publicly available on its website. Judges have the discretion to classify cases as confidential based on various policy considerations, such as litigants’

¹ Documentation regarding these variables is available in Jasper’s publication: Alon Jasper, “A Place at the Table: On the Social and Professional Composition of the Judiciary **33 Bar Ilan Legal Studies** 593 (2022)

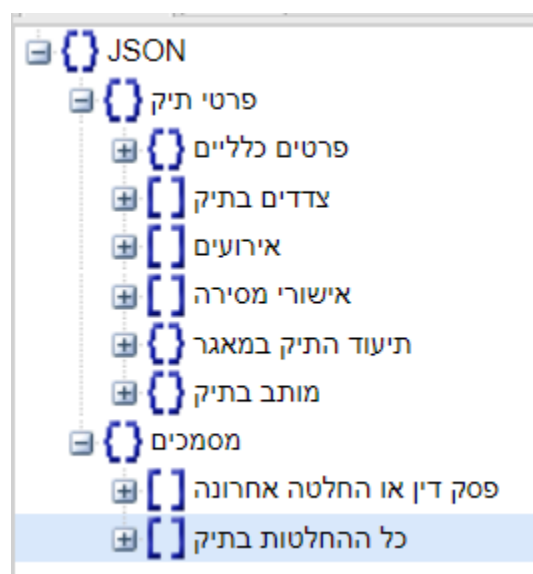
privacy and national security. Figure 1 in Appendix 4 reports the rate of confidential cases by year, showing a steady growth over time.

The procedure by which researchers may gain access to the database

We created a dedicated public facing website for the database, at the address: <https://schler.wixsite.com/il-supreme-court-db>. The database will eventually present a description of the contents of the database (in the vein provided in the previous section), documentation of all variables (in the vein of Appendix 1 and Appendix 2), a description of the procedure and data formats by which researchers may gain access to the data (see below), as well as a request form.

Researchers may request access to the data in two formats:

1. **JSON files:** A format reflecting the hierarchical nature of the data, and containing all the available variables for each case. This format is likely suitable for data scientists who are familiar with the required tools for handling it. Figure 1 below demonstrates how the JSON format appears in a hierarchical JSON viewer.



2. **Excel format:** In this format, data is provided in a tabular format as an Excel (xlsx) file, where each row represents the case number (a unique identifier), and the columns contain a subset of XX of the variables. Figure 2 demonstrates how this format looks.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
1	מספר הליך	מדור	תקנתאריך הגשה	סטטוס	מספר תיק	סוג הליך	הוצאת סוג הליך	פתיחת התיק	ראשון	ב' ראשון	צד ב' ראשון	תיק חסוי	אירוע אחרון	אירוע אחרון	בן התיק	צדדים בן התיק	מספר צדדים	מספר דלמטר
2	7622/97	אזרחי ע"א	1997-12-25	סגור	7622/97	ערעור אזהר ע"א	1997	מנורה חבה אבי דרגאקי	FALSE	מסירת פסי	1999-03-0	432	3	1	2	1	2	
3	7633/97	אזרחי רע"א	1997-12-25	סגור	7633/97	רשות ערעור רע"א	1997	קופת חוליך עידן שינדלו	FALSE	בקשה לעיני	1998-01-2	24	3	2	1	1	1	
4	3035/98	בג"ץ	1998-05-17	סגור	3035/98	בית משפט בג"ץ	1998	שר הפנים זינב אבו-זה	FALSE	מסירת פסי	1998-09-1	122	2	0	4	4	4	
5	3041/98	אזרחי ע"א	1998-05-17	סגור	3041/98	ערעור אזהר ע"א	1998	בנק דיסקונ אלטון מוצנ	FALSE	ערבון/ערוב	2000-01-2	605	5	1	4	4	4	
6	3187/99	בג"ץ	1999-05-12	סגור	3187/99	בית משפט בג"ץ	1999	איתן ברושי אבו גאסר ד	FALSE	בקשה לביני	2000-08-3	476	9	0	16	16	16	
7	1448/00	פילי ע"פ	2000-02-22	סגור	1448/00	ערעור פילי ע"פ	2000	מדינת ישרא איימן חוסני	FALSE	שכ"ט סגנו	2001-02-0	307	5	1	5	5	5	
8	1465/01	פילי	2001-02-21	סגור	1465/01	בקשות שונבש"פ	2001	מדינת ישרא מיכה בוחב	FALSE	מסירת החי	2001-02-2	4	2	0	1	1	1	
9	5933/01	פילי	2001-07-26	סגור	5933/01	בקשות שונבש"פ	2001	איברהים קי מדינת ישרא	FALSE	הוגש הליך	2001-07-3	4	2	0	1	1	1	
10	1802/02	פילי	2002-02-26	סגור	1802/02	בקשות שונבש"פ	2002	נסים כהן מדינת ישרא	FALSE	הוגש הליך	2002-03-0	5	2	1	1	1	1	
11	1808/02	בג"ץ	2002-02-26	סגור	1808/02	בית משפט בג"ץ	2002	מדינת ישרא מחמוד אלכ	FALSE	הוגש הליך	2002-12-2	280	7	0	5	5	5	
12	373/03	פילי	2003-01-12	סגור	373/03	בקשות שונבש"פ	2003	עזריה אילני עזריה ויקטו	FALSE	הוגש הליך	2003-01-1	-2	2	0	1	1	1	
13	380/03	אזרחי רע"א	2003-01-13	סגור	380/03	רשות ערעור רע"א	2003	דובק בע"מ פרוץ יובלי	FALSE	החלטה / ה	2004-08-1	580	7	0	11	11	11	
14	385/05	פילי	2005-01-12	סגור	385/05	בקשות שונבש"פ	2005	נימר אבו ע מדינת ישרא	FALSE	החלטה / ע	2005-01-1	1	4	0	1	1	1	
15	8650/05	פילי	2005-01-12	סגור	8650/05	בקשת התו בה"נ	2005	נימר אבו ע מדינת ישרא	TRUE	החלטה / ע	1900-01-0	-2	4	0	1	1	1	
16	371/06	פילי	2005-01-12	סגור	371/06	ערעור פילי ע"פ	2006	נימר אבו ע מדינת ישרא	TRUE	החלטה / ע	1900-01-0	-2	4	0	2	2	2	
17	6446/09	פילי	2005-01-12	סגור	6446/09	ערעור פילי ע"פ	2009	נימר אבו ע מדינת ישרא	TRUE	החלטה / ע	1900-01-0	-2	4	0	5	5	5	

Note: Whenever there are multiple variables in one column, this data is represented as a list, which the order in it is critical - each attribute belongs to the instance in the respective order.

CF	CG	CH
שם צד	באי כוח צד	מזהה צד
! 'אבנר איגוד	! 'ארד בעד', 'לוי עוזי', 'לוי עוזי']	! 'מערער 1', 'משיב 1', 'משיב 2']
! 'ממון חיים', 'רקובר עוזד יצחק', 'רקובר עוז	! 'ממון חיים', 'רקובר עוזד יצחק', 'רקובר עוז	! 'מערער 1', 'משיב 1', 'משיב 2']
! 'יזנב אבו-זהרייה', 'שר הפנים']	! 'כהן מאיר', 'אין מידע']	! 'עותר 1', 'משיב 1']

Data Request Procedure

In order to gain access to the data, researchers will be asked to fill out a form on the website detailing their institutional affiliation and contact information, specifying the title of the research and the data requested (including format), as well as provide proof they obtained the approval of their institutional review board (IRB) for their research project. For the reasons detailed in the following, we believe that an IRB approval is a desirable ethical safeguard, a means to ensure the appropriate handling of data by researchers, as well as a necessity given the current legal state of affairs.

While the information on each individual court case is publicly available on the Court's website to the extent sanctioned by the Court, providing electronic access to aggregate case data raises some ethical and potentially legal concerns. Specifically, the database includes personally identifiable information of litigants, lawyers, judges and other stakeholders involved in proceedings which may lead to privacy harms. Moreover, in aggregate, the data can be used for behavioral analysis of the said stakeholders, in ways which may be used to manipulate future proceedings. Indeed, the Supreme Court of Israel has long opined that technological developments stemming from the digital era pose complex issues regarding the scope of the right to review court

files, and that it may require “rethinking of the appropriate arrangement for reviewing court files.”² The Court reaffirmed this stance in the specific context of indexing and searching of electronic case files.³ Recently, after over a decade of deliberations, the issue of making such information publicly available has become particularly sensitive legally. A public committee that was appointed by the Minister of Justice over a decade ago finally issued its official recommendation to change the default publication format of court cases, such that litigants’ names will not appear on electronic case files, barring few exceptions, in order to protect their privacy.⁴ The issue of analyzing judicial big-data with quantitative tools and advanced computational tools in order to detect judicial trends and make predictions has also become controversial, to the extent that in France it is now a criminal offense to do so.⁵

In light of the above mentioned consideration, we believe that for the time being it is advisable to require researchers to obtain IRB approval for their research before granting them access to the database.

The mechanics by which the database will be maintained

We will run an annual process to update the database by running our protocol on the Supreme Court Website. The update will include new published cases and completing missing historic data (e.g. final texts of judgments, in cases that were previously pending).

We note that currently, Israeli Supreme Court decisions and case data are stored and presented on a dedicated website, independent of the general case management system which serves all other courts – “Net Hamishpat.” Our database was created by scraping and extracting data from this

² H CJ 5917/97 **Association for Civil Rights in Israel vs. Minister of Justice** (Rendered on 8.10.2009) at pp. 28-29 (“The issue of the scope of the right to review court files... requires an up-to-date examination... Today, this mainly stems from the technological developments in the computerized systems of the courts of Israel... [I]t poses complex issues concerning the rights and interests of parties to these proceedings and chief among them the constitutional right to privacy... it is possible that this situation requires a rethinking of the appropriate arrangement for reviewing court files...”).

³ H CJ 5870/14 **Hashavim HPS Business Data Ltd. vs. Israeli Court Administration** (12.11.2015) at p.17. The Court noted that “Indexing judgments is in itself a “step up” in the accessibility of legal material to the general public, thus serving the purpose for which the Court’s database was established. Indeed, the “step up” in the accessibility of judgments creates a parallel increase in the violation of the privacy of the litigants, with the human sensitivities involved, and it is certainly possible that the administrative authorities should take this into account and look for solutions” (

⁴ While we could not obtain a copy of the report itself, the following newspaper article describes its main recommendations: Hen Maanit, “A Public Committee Recommends Concealing the Identity of Litigants from Published Court Cases **TheMarker** (16 May 2023), available at: <https://www.haaretz.co.il/news/law/2023-05-16/ty-article/premium/00000188-20eb-d18b-a79d-24fba6b30000>

⁵ See Article 33 of the French Justice Reform Act, which states that: “The identity data of magistrates and members of the judiciary cannot be reused with the purpose or effect of evaluating, analyzing, comparing or predicting their actual or alleged professional practices.” The full text of the legislation is available at: https://www.legifrance.gouv.fr/jorf/article_jo/JORFARTI000038261761

website. It has recently come to our knowledge that the Israeli Court Administration intends to converge the Supreme Court data into “Net Hamishpat” system. Once this migration is complete, a different data scraping and extraction protocol will need to be developed, based on “Net Hamishpat” architecture. We believe that many of the metadata variables that we were able to scrape from the current Supreme Court website will no longer be scrapable from “Net Hamishpat.”

At the same time, the knowledge we gained in this project will undoubtedly serve us in developing any future protocol for data published on “Net Hamishpat.”

Appendix 1 - Court Metadata Variables

פרטי תיק - {Dict}					
פרטים כלליים - {Dict}					
פרטים כלליים כפי שמופיעים באתר ביה"מ העליון במידה והתיק חסי - למשתנה זה יקראו "פרטי תיק כלליים" ויהיה בו פחות מידע (אין METADATA באתר)					
משתנה	הסבר	מקור	סוג	דוגמא	ערכים אפשריים (מתוך רשימה)
מספר הליך	מספר הליך	מטה-דאטה	ראשי תיבות + שנה\מספר רץ - text	בג"ץ 2\00	
מדור	מדור	מטה-דאטה	ראשי תיבות - text	בג"ץ	[אזורי, בג"ץ, פלילי]
תאריך הגשה	תאריך הגשת התיק	מטה-דאטה	YYYY-MM-DD- text	31-12-1992	
סטטוס תיק	האם התיק פתוח או נסגר	מטה-דאטה	פתוח\סגור - text	סגור	[פתוח, סגור]
מערער	צד א' - אדם או גוף	מטה-דאטה	שם מלא - text	ישראל ישראלי	
משיב	צד ב' - אדם או גוף	מטה-דאטה	שם מלא - text	מועצה אזורית דרום השרון	
מספר תיק מלא	מספר תיק וראשי תיבות	מחושב - זהה למספר ההליך	ראשי תיבות + שנה\מספר רץ - text	בג"ץ 2\00	
מספר תיק	מספר תיק - חיתוך הספרות בלבד	מחושב מתוך מספר ההליך	שנה\מספר רץ text	2\00	
סוג הליך	ראשי התיבות מנורמלות (בלי אותיות סופיות ועם גרש)	מספר ההליך	ראשי תיבות - text	בג"ץ	רשימת סוגי ההליכים
פירוט סוג הליך	פירוט ראשי התיבות מתוך טבלת פירוט סוגי ההליכים	מחושב	טקסט ראשי תיבות - text	בקשות שונות פלילי	רשימת פירוט סוגי ההליכים
שנת פתיחת התיק	שנת התיק המלאה	מספר ההליך	YYYY - text	2000	
אירוע אחרון	סוג אירוע אחרון בתיק	מטה-דאטה	סוג האירוע - text	החלטה; החלטה נשלחה לצדדים	
תאריך ארוע אחרון	תאריך האירוע	מתוך מטה-דאטה - אירועים	YYYY-MM-DD- text	09/03/2006	
משך התיק בימים	כמה זמן עבר בין פתיחת התיק לסגירתו	מחושב	מספר שלם - int	24	
כמויות - {dict}					
משתנה	הסבר	מקור	סוג	דוגמא	
מספר תיקי דלמטה	כמות תיקי דלמטה	מחושב-סכום	מספר שלם - int	2	
מספר דיונים	כמות דיונים	מחושב-סכום	מספר שלם - int	1	
מספר אירועים	כמות אירועים	מחושב-סכום	מספר שלם - int	4	
מספר אישורי מסירה	כמות אישורי מסירה	מחושב-סכום	מספר שלם - int	2	
מספר בקשות בתיק	כמות בקשות	מחושב-סכום	מספר שלם - int	5	
מספר החלטות בתיק	כמות החלטות	מחושב-סכום	מספר שלם - int	1	

צדדים בתיק - [list]					
פירוט צדדים בתיק כפי שמופיעים באתר ביה"מ העליון (רשימה בעלת גודל משתנה בהתאם למספר הצדדים)					
משתנה	הסבר	מקור	סוג	דוגמא	ערכים אפשריים (מותר רשימה)
סוג צד	סוג צד	מטה-דאטה	סוג הצד - text	עותר	[מבקש, מערער, העותר, מאשים, התובע, עורר, [נתבע, נאשם, משיב]
מספר	מספר רץ של אותו סוג צד	מטה-דאטה	מספר שלם - text	1	
שם	שם הצד - אדם או גוף	מטה-דאטה	שם מלא - text	ישראל ישראלי	
באי כוח	שם המייצג-ת בהליך	מטה-דאטה	שם מלא - text	משה כהן	
צד	סוג צד ומספר הצד	מחושב	סוג צד + מספר - text	עותר 1	

תיק דלמטה - [list]					
תיק דלמטה כפי שמופיעים באתר ביה"מ העליון (רשימה בעלת גודל משתנה בהתאם למספר התיקים)					
משתנה	הסבר	מקור	סוג	דוגמא	ערכים אפשריים (מותר רשימה)
שם בית המשפט דלמטה	שם בית המשפט	מטה-דאטה	שם בית משפט - text	מחוזי מרכז-לוד	רשימת בתי משפט דלמטה
מספר תיק דלמטה	מספר תיק דלמטה	מטה-דאטה	ראשי תיבות + מספר סידורי - text	עמל 03-013271-14	
תאריך החלטה דלמטה	תאריך החלטה	מטה-דאטה	YYYY-MM-DD - text	21-03-2010	
גורם שיפוט דלמטה	האם שופט או הרכב	מטה-דאטה	שופט/הרכב - text	שופט	[שופט, הרכב]
פירוש סוג תיק דלמטה	פירוש ראשי התיבות	מחושב	השוואה מול רשימת ראשי התיבות ופירוטן מאתר ביה"מ העליון - text	ערעור פלילי תעבודה	רשימת פירושים
סוג בית משפט דלמטה	פירוט סוג בית המשפט	מחושב - מתוך "שם בית המשפט דלמטה"	השוואה מול רשימת בתי המשפט - text	בית הדין המשמעתי	רשימת בתי משפט
איזור בית משפט דלמטה	פירוט איזור בית המשפט	מחושב - מתוך "שם בית המשפט דלמטה"	השוואה מול רשימת בתי המשפט - text	חיפה	רשימת בתי משפט

דיונים - [list]					
דיונים כפי שמופיעים באתר ביה"מ העליון (רשימה בעלת גודל משתנה בהתאם למספר הדיונים)					
משתנה	הסבר	מקור	סוג	דוגמא	ערכים אפשריים (מותר רשימה)
תאריך דיון	תאריך הדיון	מטה-דאטה	YYYY-MM-DD - text	16-08-2006	
שעת דיון	שעת הדיון	מטה-דאטה	HH:MM - text	11:00	
אולם דיון	מיקום הדיון	מטה-דאטה	מספר\אות האולם - text	ב'	רשימה סגורה?
גורם שיפוט דיון	שופט יחיד או הרכב	מטה-דאטה	דן יחיד/הרכב - text	דן יחיד	[דן יחיד, הרכב]
סטטוס דיון	האם הדיון התקיים	מטה-דאטה	התקיים/נדחה/אין מידע - text	התקיים	[התקיים, נדחה, אין מידע]

אירועים - [list]				
אירועים כפי שמופיעים באתר ביה"מ העליון (רשימה בעלת גודל משתנה בהתאם למספר האירועים)				
משתנה	הסבר	מקור	סוג	דוגמא
מספר	מספר האירוע	מטה-דאטה	מספר שלם - text	2
אירוע ראשי	סוג האירוע הראשי	מטה-דאטה	כותרת האירוע - text	הוגשה הודעה
אירוע משני	פירוט האירוע	מטה-דאטה	פירוט האירוע - text	הודעה מטעם המערער
תאריך	תאריך האירוע	מטה-דאטה	YYYY-MM-DD- text	09/03/2006

אישורי מסירה - [list]				
אישורי המסירה כפי שמופיעים באתר ביה"מ העליון (רשימה בעלת גודל משתנה בהתאם למספר אישורי המסירה)				
משתנה	הסבר	מקור	סוג	דוגמא
מספר	מספר אישור המסירה	מטה-דאטה	מספר שלם - text	123757
נמען	אליו נמסר האישור	מטה-דאטה	אדם/מייצג/גוף - text	ע"ד משה כהן
תאריך חתימה	תאריך חתימה	מטה-דאטה	YYYY-MM-DD- text	08/10/2006

בקשות - [list]				
בקשות כפי שמופיעות באתר ביה"מ העליון (רשימה בעלת גודל משתנה בהתאם למספר הבקשות)				
משתנה	הסבר	מקור	סוג	דוגמא
מספר	מספר הבקשה	מטה-דאטה	אות עברית המייצגת מספר רץ - text	ב
תיאור הבקשה	תיאור הבקשה	מטה-דאטה	טקסט - נושא הבקשה - text	למתן החלטה או פסק דין
תאריך	תאריך הבקשה	מטה-דאטה	DD-MM-YYYY - text	10/10/2006
שעה	שעת הבקשה	מטה-דאטה	HH:MM - text	6:09:00 PM
מגיש	מגיש הבקשה	מטה-דאטה	סוג המגיש (צד) ומספרו - text	משיב 1
נדחה מהמרשם	נימוקי מזכירות לחיית מסמך שהוגש	מטה-דאטה	טקסט חופשי - text	אין מידע

תיעוד התיק במאגר - [list]				
משתנה	הסבר	מקור	סוג	דוגמא
תיק חסוי	האם התיק חסוי	מחושב	בוליאני	TRUE
תאריך יצוא התיק	מתי חולץ מהאתר	מחושב	YYYY-MM-DD - text	2022-11-030
תאריך יצוא הקובץ	תאריך בו הקובץ נשמר	מחושב	YYYY-MM-DD - text	13-01-2002
תאריך נרמול הקובץ	תאריך סידור ונרמול של הקובץ	מחושב	YYYY-MM-DD - text	2022-12-303
שם הקובץ	שם הקובץ ע"פ שם התיק	מחושב	מספר תיק שנה - text	2006_6466
קישור לתיק	קישור לתיק באתר ביה"מ העליון	מחושב	text - HTTPS	ecisions.court.go

מסמכים - {dict}

פסק דין או החלטה אחרונה - [list]

שליפת נתוני המסמך (פסק הדין, אם אין או החלטה האחרונה שהתקבלה בתיק)

משתנה	הסבר	מקור	סוג	דוגמא	ערכים אפשריים (מתוך רשימה)
לפני	שמות השופטים	מסמך פס"ד או החלטה אחרונה (מנורמל ל"לפני")	שמות מלאים + תואר - list	כבוד השופט א' רבינשטיין	
תאריך הישיבה	תאריך הישיבה	מסמך פס"ד או החלטה אחרונה	תאריך עברי מלא + text - (DD.MM.YYYY)	כ"ב באב התשס"ו (16.8.06)	
כל קוד הטמל	כל קוד HTML במסמך	Crawling	HTML - text		
מסמך מלא	כל הטקסט מהמסמך - פרטי ההליך + טקסט ההחלטה	מסמך פס"ד או החלטה אחרונה	טקסט - text		
גוף המסמך	טקסט ההחלטה\פס"ד בלבד	מסמך פס"ד או החלטה אחרונה	חלק הסיכום של המסמך - text		
קישור למסמך					
סוג מסמך	פסק דין \ החלטה אחרונה	Crawling	פסק דין \ החלטה אחרונה - text	פסק דין	
מספר השופטים	מספר השופטים	מחושב	מספר שלם int	3	
שמות השופטים	שמות השופטים - חיתוך משדה "לפני"	מחושב	שם מלא - text	א' ריבלין	
תפקידים	תוארי השופטים - חיתוך	מחושב	שם תואר - text	כבוד המשנה לנשיאה	
צד א'	צד א' מנורמל	מחושב	שם מלא - text	שמיצקו השקעות בע"מ	
צד ב'	צד ב' מנורמל	מחושב	שם מלא - text	יוסף מזרחי	
צדדים נוספים	צדדים נוספים (ידיד, פורמאלי)	מחושב	שם מלא - text	דוד אברהם	
בשם צד א'	מייצגי צד א'	מחושב	שם מלא - text	עו"ד משה כהן	
בשם צד ב'	מייצגי צד ב'	מחושב	שם מלא - text	עו"ד שרה ישראלי	
סוג הליכים	האם ההליך מאוחד	מחושב	האם ההליך מאוחד\ערעור וערעור שכנגד\תיק יחיד	תיק יחיד	הליך מאוחד, ערעור [ערעור שכנגד, תיק יחיד]
שמות הליכים המאוחדים	שמות התיקים המאוחדים במידה ויש	מחושב	שמות התיקים\שם התיק+ התיק שכנגד	דנ"א 101/08 ו"דנ"א 101/08 שכנגד	
תאריך ההחלטה/פסק הדין	תאריך החלטה אחרונה / פסק דין	מחושב מתוך מטת-דאטה החלטות	YYYY-MM-DD - text	21/03/2010	
אין מסמכים להצגה	במידה אין מסמכים להצגה		בוליאני		מופיע רק במידה ואין מסמכים להצגה

פרטי השופטים					
משתנה	הסבר	מקור	סוג	דוגמא	ערכים אפשריים (מתוך רשימה)
שם השופט	שם מלא	מתוך טבלת השופטים	שם פרטי ושם משפחה	נאור מרים	
מגדר	מגדר	מתוך טבלת השופטים	מגדר	נקבה	[זכר, נקבה]
לאום	לאום	מתוך טבלת השופטים	לאום	יהודי	[יהודי, מוסלמי, מצרי]
יישוב מגורים מרכזי בילדות	עיר מגורים מרכזי בילדות	מתוך טבלת השופטים	עיר בישראל \ מחוץ לישראל	ירושלים	ירושלים, מחוץ לישראל, תל אביב-יפו, וכו'

כל ההחלטות בתיק - [list]					
רשימת כל ההחלטות שהתקבלו בתיק לפי אתר ביה"מ העליון					
משתנה	הסבר	מקור	סוג	דוגמא	
תאריך החלטה	תאריך החלטה\פסק דין	מטה-דאטה	YYYY-MM-DD - text	21-03-2010	
סוג מסמך	פסק דין\ החלטה	מטה-דאטה	פסק דין\ החלטה - text	פסק דין	
מספר עמודים	מספר העמודים במסמך	מטה-דאטה	מספר שלם - int	12	
קישור למסמך הטמל	קישור למסמך ההחלטה\פס"ד	מחושב	קישור - text - HTTPS	https://supremede	

Appendix 2 - Judges' Biographical Variables

Full documentation regarding judges' biographical variables is available in: Alon Jasper, "A Place at the Table: On the Social and Professional Composition of the Judiciary" 33 **Bar Ilan Legal Studies** 593 (2022)

שם	שנת מינוי למשרה שיפוטית 3
מגדר	משרה שיפוטית 4
לאום	שם בית המשפט במשרה שיפוטית 4
דת	שנת מינוי למשרה שיפוטית 4
קבוצה דתית	משרה שיפוטית 5
שיוך עדתי	שם בית המשפט במשרה שיפוטית 5
ארץ/איזור לידה	שנת מינוי למשרה שיפוטית 5
שנת לידה	משרה שיפוטית 6
יישוב מגורים מרכזי בילדות	שם בית המשפט במשרה שיפוטית 6
דירוג חברתי כלכלי של יישוב מגורים מרכזי בילדות	שנת מינוי למשרה שיפוטית 6
בית ספר תיכון	משרה שיפוטית 7
סוג החינוך	שם בית המשפט במשרה שיפוטית 7
מוסד חינוך משפטי עיקרי	שנת מינוי למשרה שיפוטית 7
שנת סיום התואר הראשון במשפטים	משרה שיפוטית 8
סוג השכלה משפטית מתקדמת	שם בית המשפט במשרה שיפוטית 8
מוסד חינוך משפטי מתקדם	שנת מינוי למשרה שיפוטית 8
שנת קבלת התואר המתקדם	שנת מינוי ראשון לבית המשפט העליון
מקום התמחות	שנת סיום מינוי ראשון לבית המשפט העליון
שנת הסמכה כעורך/ת דין	שנת מינוי שני לבית המשפט העליון
המשרה האחרונה טרם המינוי לשיפוט	שנת סיום מינוי שני לבית המשפט העליון
משרה שיפוטית 1	כהונה כרשם/מת?
שם בית המשפט במשרה שיפוטית 1	כהונה כשופט/ת מחוזי?
שנת מינוי למשרה שיפוטית 1	כהונה כשופט/ת שלום?
משרה שיפוטית 2	שנת סיום מינוי כשופט/ת
שם בית המשפט במשרה שיפוטית 2	גיל מינוי לשיפוט
שנת מינוי למשרה שיפוטית 2	ותק במשפטים בעת מינוי ראשון
משרה שיפוטית 3	גיל בעת הפרישה משיפוט
שם בית המשפט במשרה שיפוטית 3	ותק שיפוטי בעת המינוי לביהמ"ש העליון

Appendix 3: Validation and Normalization Procedures

Data Validation

RAs from the law faculty in Bar Ilan university, with an affinity for software and data were recruited to verify the accuracy of the data downloaded to the JSON files, the quality of the data and its correlation with the relevant case on the Supreme Court website. Crawler malfunctions were reported to the PIs for repair. This process ensured that data were captured correctly and in an intelligible and useful manner for future end users.

Data validation also included correcting human data entry errors that originate from the court systems themselves. For example, we observed instances in which a calculated field we created indicating the duration of the procedure (derived from the date the case was opened until the date the final decision was rendered) yielded a negative value. Data validation revealed the dates were erroneously entered into the case management system and appear as such on the court's website. In these cases, we adjusted the crawler to discover the mistake, and derive the duration of the case from an alternative source: the dates listed in the text of the final decision

Dataset Normalization

For data consistency all data goes through a normalization layer. This layer normalizes words that appear in various formats in the meta data or text (i.e. different words that reflect the same variable/attribute) in order to create uniformity. In addition, we added an indication for variables with missing values in a given case.

In order to create uniformity in the data, the normalization module works on two levels:

1. Normalizing the content of different rewritten fields into one uniform language. For example – [ביה"מ, ביהמ"ש, בית המשפט, בית משפט, בימ"ש] are all normalised as "בית המשפט"). This will assist search and analysis later. A similar approach is used also for judges' names and other common variables.
2. Generating new meaningful normalized values for data that represent a broader meaningful category (legal concept). For example – [המערער, המערערת, התובע, התובעת, המבקש, המבקשת, המאשים, המאשימה] are all categorized into an additional label as "צד א"). This normalization will enable streamlined analyses of parties in Supreme Court cases, regardless of their role or case type.

Appendix 4 - Rate of Confidential Cases in the Supreme Court Website by Year

