

# Scripts

Different types of scripts to extract data here and there.

- [Internet Archive Usage Data > CSV Python Script](#)
- [GraphQL Script to extract Author Affiliations from the Thoth Catalog](#)
- [GraphQL Script to Get Books Published per Year](#)
- [Extracting DOIs from a PDF \(requires manual check still because of line breaks\)](#)
- [Uploading extracted DOIs to Thoth](#)

# Internet Archive Usage Data

## > CSV Python Script

```
import requests
import csv
```

Define the list of item IDs and the time period

```
item_ids = [
    "f755b169-bedd-4b95-b95a-294f355abc71",
    "7fbc96cf-4c88-4e70-b1fe-d4e69324184a",
    "4da1ef38-e3d8-4384-8c73-82d6e36df49d",
    "f3c9e9d8-9a38-4558-be2e-cab9a70d62f0",
    "89990379-94c2-4590-9037-cbd5052694a4",
    "0090dbfb-bc8f-44aa-9803-08b277861b14",
    "1399a869-9f56-4980-981d-2cc83f0a6668",
    "6c51c8fa-947b-4a12-a2e9-5306ee81d117",
    "637566b3-dca3-4a8b-b5bd-01fcbb77ca09",
    "c00d9a0c-320d-4dfb-ba0c-d1adbdb491ef",
    "77e1fa52-1938-47dd-b8a5-2a57bfb9c91d1",
    "8377c394-c27a-44cb-98f5-5e5b789ad7b8",
    "c6125a74-2801-4255-afe9-89cdb8d253f4",
    "003137ea-4fe6-470d-8bd3-f936ad065f3c",
    "9e11adff-abed-4b5d-adeb-b0c4466231e8",
    "93330f65-a84f-4c5c-aa44-f710c714eca2",
    "b4d68a6d-01fb-48f1-9f64-1fcdaaf1cdfd",
    "11749800-364e-4a27-bf79-9f0ceeacb4d6",
    "a3779e1d-5b3a-4bd0-b193-6e8a0ad8bde3",
    "08543bd7-e603-43ae-bb0f-1d4c1c96030b",
    "60c43792-7509-4d10-bd72-5f7be5fc610a",
    "61b6e4fb-d99e-4b6e-beca-293c990cef44",
    "d6651c3c-c453-42ab-84b3-4e847d3a3324",
    "2b166028-700d-4c4b-bb43-b8fd6d6ba7e6",
    "7aa9a630-c5e0-42e8-825c-c854dfbffb56",
    "b761f18e-a2b6-4343-917f-646d64b54628",
    "417ecc06-51a4-4660-959b-482763864559",
    "5fe37152-6263-4293-ac3d-9799585563a6",
    "7eb6f426-e913-4d69-92c5-15a640f1b4b9",
    "200b11a8-57d6-4f81-b089-ddd4ee7fe2f2",
    "10a41381-792f-4376-bed1-3781d1b8bae7",
    "ee2cb855-4c94-4176-b62c-3114985dd84e",
```

"f5fb8a0e-ea1d-471f-b76a-a000edae5956",  
"911de470-77e1-4816-b437-545122a7bf26",  
"1d014946-aa73-4fae-9042-ef8830089f3c",  
"6ca16a49-7c95-4c81-b8f0-8f3c7e42de7d",  
"85a2a2fe-d515-4784-b451-d26ec4c62a4f",  
"61da662d-c720-4d22-957c-4d96071ee5f2",  
"d97b2e6e-8fab-40ee-a38d-5dc804b150ed",  
"0ff62120-4478-46dc-8d01-1d7e1dc5b7a6",  
"69c890c5-d8c5-4295-b5a7-688560929d8b",  
"419e17ed-3bcc-430c-a67e-3121537e4702",  
"438e0846-b4b9-4c84-9545-d7a6fb13e996",  
"f5217945-8c2c-4e65-a5dd-3dbff208dfb7",  
"34682ba4-201f-4122-8e4a-edc3edc57a7b",  
"fe8ddfb7-0e5b-4604-811c-78cf4db7528b",  
"29bae81a-80ad-470d-926a-c02ceef170dc",  
"97a2ac65-5b1b-4ab8-8588-db8340f04d27",  
"628bb121-5ba2-4fc1-a741-a8062c45b63b",  
"d8284a1f-51ac-4ecd-99a6-45e5644db4c8",  
"cc194166-c595-48aa-99f8-b5109fa2c2d0",  
"0ea7574d-2d74-4d69-8dd3-ce39f34cd9ce",  
"84bbf59f-1dbb-445e-8f65-f26574f609b6",  
"13390641-86f6-4351-923d-8c456f175bff",  
"dbe1ab13-7c70-444e-840d-20e46f04cc98",  
"1376b0f4-e967-4a6f-8d7d-8ba876bbbdde",  
"448c3581-9167-491e-86f7-08d5a6c953a9",  
"cc73eed0-a1f9-4ad4-b7d8-2394b92765f0",  
"a065ad95-716a-4005-b436-a46d9dbd64df",  
"e1f74d6b-adab-4e56-8bc9-6fbd0eaab89c",  
"58aadfa5-abc6-4c44-9768-f8ff41502867",  
"49ebcb4a-928f-4d83-9596-b296dfce0b20",  
"f37627c1-d89f-434c-9915-f1f2f33dc037",  
"d3cbb60f-537f-4bd7-96cb-d8aba595a947",  
"33917b8f-775f-4ee2-a43a-6b5285579f84",  
"0a8fba81-f1d0-498c-88c4-0b96d3bf2947",  
"eb8a2862-e812-4730-ab06-8dff1b6208bf",  
"efd22fcf-1ed6-420d-ad0b-144db9fa642d",  
"6fab7c76-7567-4b57-8ad7-90a5536d87af",  
"da814d9f-14ff-4660-acfe-52ac2a2058fa",  
"cd037a39-f6b9-462a-a207-5079a000065b",  
"f20869c5-746f-491b-8c34-f88dc3728e18",  
"f3db2a03-75db-4837-af31-4bb0cb189fa2",  
"119e45d6-63ab-4cc4-aabf-06ecba1fb055",  
"1950e4ba-651c-4ec9-83f6-df46b777b10f",  
"fccabf96-0c07-4554-835f-a83d788e1b42",  
"37536ae7-c4d2-428c-865e-bcb88b902371",  
"37cb9bb4-0bb3-4bd3-86ea-d8dfb60c9cd8",

"2658fe95-2df3-4e7d-8df6-e86c18359a23",  
"5402ea62-7a1b-48b4-b5fb-7b114c04bc27",  
"14a2356a-4767-4136-b44a-684a28dc87a6",  
"cd836291-fb7f-4508-bdff-cd59dca2b447",  
"eeb920c0-6f2e-462c-a315-3687b5ca8da3",  
"08788fe1-c17d-4f6b-aeab-c81aa3036940",  
"e6962291-bc96-4f3f-906e-0e1e1e317f58",  
"d5f5978b-32e0-44a1-a72a-c80568c9b93a",  
"ee2effea-4051-470b-b4cb-63bccfe30943",  
"2d522a5b-ecca-402f-8188-52c6100be235",  
"449add5c-b935-47e2-8e46-2545fad86221",  
"f131762c-a877-4925-9fa1-50555bc4e2ae",  
"df518095-84ff-4138-b2f9-5d8fe6ddf53a",  
"a9d68f12-1de4-4c08-84b1-fe9a786ab47f",  
"34584bfe-1cf8-49c5-b8d1-6302ea1cfcfa",  
"1a71ecd5-c868-44af-9b53-b45888fb241c",  
"bdfc263a-7ace-43f3-9c80-140c6fb32ec7",  
"06db2bc1-e25a-42c8-8908-fbd774f73204",  
"81dbcd46-54ce-4b6e-b5ec-2df4004947e1",  
"3b41b8de-b9bb-4ebd-a002-52052a9e39a9",  
"1547aa4b-7629-4a21-8b2b-621223c73ec9",  
"9e108419-fc08-440d-b1b5-eb8b5cd52c5e",  
"0a1b16f0-411d-4509-b50c-088a3495476e",  
"3cc0269d-7170-4981-8ac7-5b01e7b9e080",  
"ae9f8357-4b39-4809-a8e9-766e200fb937",  
"1c39ca0c-0189-44d3-bb2f-9345e2a2b152",  
"5f24bd29-3d48-4a70-8491-6269f7cc6212",  
"b950d243-7cfc-4aee-b908-d1776be327df",  
"46ab709c-3272-4a03-991e-d1b1394b8e2c",  
"60813d93-663f-4974-8789-1a2ee83cd042",  
"078ba632-782a-44ef-a75c-af87218a1908",  
"63e2f6b6-f324-4bdc-836e-55515ba3cd8f",  
"5b652d05-2b5f-465a-8c66-f4dc01dafd03",  
"895b271b-6669-43cc-842d-3f92a39fd463",  
"bb91585a-0ce3-4e37-b6e2-0bf160cedab0",  
"639a3c5b-82ad-4557-897b-2bfebe3dc53c",  
"c21f4155-1d84-4590-9c08-f67ac39f3d97",  
"b63ffeb5-7906-4c74-8ec2-68cbe87f593c",  
"7f1d3e2e-c708-4f59-81cf-104c1ca528d0",  
"13c12944-701a-41f4-9d85-c753267d564b",  
"680b88e2-8685-4d13-806d-b274a7b1afe6",  
"a23f9c80-72b4-4ccc-a6c7-22a261691da5",  
"75693fd0-e93a-4fc3-b82e-4c83a11f28b1",  
"435a1db3-1bbb-44b2-9368-7b2fd8a4e63e",  
"ed3ea389-5d5c-430c-9453-814ed94e027b",  
"ac612279-727a-4920-a9bf-95879369e7a8",

"4f46d026-49c6-4319-b79a-a6f70d412b5c",  
"8d778210-0ef3-47fa-b106-ea173874259b",  
"859e72c3-8159-48e4-b2f0-842f3400cb8d",  
"f6b4d156-1be6-41f0-b641-66bf50016b24",  
"5482b275-d814-4dc9-a429-92c0aeb05f3a",  
"b0257269-5ca3-40b3-b4e1-90f66baddb88",  
"96aee3d0-e618-4fe7-b6a8-87cd3c8f2f93",  
"f712541c-07b4-477c-8b8c-8c1a307810d0",  
"06b256bf-62cf-4e01-be37-44f250a8b920",  
"860d205d-1953-4524-b1a7-70c59c42bd0e",  
"0d485404-68e5-4f24-83e8-e993093f5f1d",  
"ed96eea8-82c6-46c5-a19b-dad5e45962c6",  
"006571ae-ac0e-4cb0-8a3f-71280aa7f23b",  
"f456bebb-2710-4140-a182-aa1d713460c7",  
"6dd15dd7-ae8c-4438-a597-7c99d5be4138",  
"ae6d88a5-a488-49b3-b8f8-8ed3e9b1e9c4",  
"d98dcf27-9898-4a48-92cf-18101631afc9",  
"ffb69824-12b3-4900-8aee-53d024151137",  
"0ff404f8-cd4a-43d8-a4f7-fd5843674e00",  
"a871cb31-e158-401d-a639-3767131c0f34",  
"88887777-f901-4ee7-97d8-b72b498dbb99",  
"f2a2626b-4029-4e43-bb84-7b3cacf61b23",  
"41508a3c-614b-473e-aa74-edcb6b09dc9d",  
"861ea7cc-5447-4c60-8657-c50d0a31cd24",  
"47cd079b-03f3-4a5b-b5e4-36cec4db7fab",  
"a022743e-8b77-4246-a068-e08d57815e27",  
"0e43d1f5-30f5-4001-88c0-9b99d00a43dc",  
"20dab41d-2267-4a68-befa-d787b7c98599",  
"196206d8-879e-4737-a0d7-26b576224def",  
"8f2609a9-b833-4ec0-aeda-b844f6a319e0",  
"381b2f4f-0b6a-45cd-89ef-da7fd6e7b662",  
"0cd80cd2-1733-4bde-b48f-a03fc01acfbf",  
"20d15631-f886-43a0-b00b-b62426710bdf",  
"fa4bc310-b7db-458a-8ba9-13347a91c862",  
"f02786d4-3bcc-473e-8d43-3da66c7e877c",  
"7fe2c6dc-6673-4537-a397-1f0377c2296f",  
"0985e294-aa85-40d0-90ce-af53ae37898d",  
"4fc74913-bde4-426e-b7e5-2f66c60af484",  
"5f441303-4fc6-4a7d-951e-5b966a1cbd91",  
"19b32470-bf29-48e1-99db-c08ef90516a9",  
"65ba0747-ff5b-4299-9d41-7e940e19b8be",  
"d2e40ec1-5c2a-404d-8e9f-6727c7c178dc",  
"84447325-88e2-4658-8597-3f2329451156",  
"ed1a8fb5-8b71-43ca-9748-ebd43f0d7580",  
"d0d59741-4866-42c3-8528-f65c3da3ffdd",  
"bbe77bbb-0242-46d7-92d2-cfd35c17fe8f",

"cb11259b-7b83-498e-bc8a-7c184ee2c279",  
"9e6bb4d8-4e05-4cd7-abe9-4a795ade0340",  
"7e2e26fd-4b0b-4c0b-a1fa-278524c43757",  
"b43ec529-2f51-4c59-b3cb-394f3649502c",  
"5a78355d-4b21-40a3-a379-8fefe1ecff9f",  
"2d88c98b-7a96-4e4d-a462-bda17a91c376",  
"6dcf29ea-76c1-4121-ad7f-2341574c45fe",  
"3d78e15e-19cb-464a-a238-b5291dbfd49f",  
"a8e6722a-1858-4f38-995d-bde0b120fe8c",  
"18d3d876-bcaf-4e1c-a67a-05537f808a99",  
"2a360648-3157-4a1b-9ba7-a61895a8a10c",  
"177e3717-4c07-4f31-9318-616ad3b71e89",  
"53cd2c70-eab6-45b7-a147-8ef1c87d9ac0",  
"431b58fe-7f59-49d9-bf6f-53eae379ee4d",  
"5dca2af4-43f2-4cdb-a7a5-5654a722c4e0",  
"b2d1b2e3-226e-43c2-a898-fbad7b410e3f",  
"e770d0ad-6124-4ff3-b21b-784dd172668d",  
"a3f50086-75c6-4f69-aace-7e498e21beb6",  
"f4d42680-8b02-4e3a-9ec8-44aee852b29f",  
"b605be50-a3bb-4d1f-a341-95dd486c63a9",  
"14f2b847-faeb-43c9-b116-88a0091b6f1f",  
"46344fe3-1d72-4ddd-a57e-1d3f4377d2a2",  
"35ed7096-8218-43d7-a572-6453c9892ed1",  
"c5b21f1e-990e-4d24-a0da-40cac057b175",  
"30fa2ca0-d82f-4981-bf56-0481ac6c58be",  
"9839926e-56ea-4d71-a3de-44cabd1d2893",  
"9bd104d3-59e6-4ac0-9ccd-96079902769e",  
"6cf8d32a-92bc-4e43-8881-0ea812d78374",  
"3f78b298-8826-4162-886e-af21a77f2957",  
"5ec826f5-18ab-498c-8b66-bd288618df15",  
"13a03c11-0f22-4d40-881d-b935452d4bf3",  
"763ca210-ebe8-4a54-8e98-d57acd0d5e2f",  
"992f0e59-77d2-4c46-bc15-e85404660d8a",  
"1e5ea7bf-a13b-4bfa-834d-1c9209f99c19",  
"0fa87d43-ad8f-4bde-b70e-af249e3a2537",  
"992c6ff8-e166-4014-85cc-b53af250a4e4",  
"f01cb60b-69bf-4d11-bd3c-fd5b36663029",  
"48e2a673-aec2-4ed6-99d4-46a8de200493",  
"28a0db09-a149-43fe-ba08-00dde962b4b8",  
"4a42f23b-5277-49b5-8310-c3c38ded5bf5",  
"b50aca26-6527-4db3-8ab9-57ba2f3cb14f",  
"26d522d4-fb46-47bf-a344-fe6af86688d3",  
"69365c88-4571-45f3-8770-5a94f7c9badc",  
"6a109b6a-55e9-4dd5-b670-61926c10e611",  
"e81ef154-5bc3-481b-9083-64fd7aeb7575",  
"7be9aa8c-b8af-4b2f-96ff-16e4532f2b83",

"8b77c06a-3c1c-48ac-a32e-466ef37f293e",  
"c2c22cdf-b9d5-406d-9127-45cea8e741b1",  
"d24a0567-d430-4768-8c4d-1b9d59394af2",  
"df9bf011-efaf-49a7-9497-2a4d4cfde9e8",  
"bc283f71-9f37-47c4-b30b-8ed9f3be9f9c",  
"0cb39600-2fd2-4a7a-9d3a-6d92b8e32e9e",  
"a9b27739-0d29-4238-8a41-47b3ac2d5bd5",  
"ebeae9d6-7543-4cd4-9fa9-c39c43ba0d4b",  
"119f1640-dfb4-488f-a564-ef507d74b72d",  
"dfe575e1-2836-43f3-a11b-316af9509612",  
"b55c95a7-ce6e-4cfb-8945-cab4e04001e5",  
"367397db-bcb4-4f0e-9185-4be74c119c19",  
"c3c09f99-71f9-431c-b0f4-ff30c3f7fe11",  
"6c30545e-539b-419a-8b96-5f6c475bab9e",  
"edf31616-ea2a-4c51-b932-f510b9eb8848",  
"90b5558c-19fe-4ee2-8a7c-5ea6a6bb8d2f",  
"ce38f309-4438-479f-bd1c-b3690dbd7d8d",  
"781b77bd-edf8-4688-937d-cc7cc47de89f",  
"bd78286d-43e0-4561-8279-94b629ddb4fd",  
"a6e8459d-9451-44bd-ad5e-d3962c652ea0",  
"ea367445-3c67-499c-84f9-0b611766f2cb",  
"5dda1ad6-70ac-4a31-baf2-b77f8f5a8190",  
"ac6acc15-6927-4cef-95d3-1c71183ef2a6",  
"1ef505f6-6c8d-4875-8be2-c3328b6c2ce5",  
"a2ce9f9c-f594-4165-83be-e3751d4d17fe",  
"3c5923bc-e76b-4fbe-8d8c-1a49a49020a8",  
"1b870455-0b99-4d0e-af22-49f4ebbb6493",  
"1cfca75f-2e57-4f34-85fb-a1585315a2a9",  
"571255b8-5bf5-4fe1-a201-5bc7aded7f9d",  
"1e0c7c29-dcd4-470d-b3ee-8c4012ac79dd",  
"571d5d40-cfd6-4270-9530-88bfcfc5d8b5",  
"dc622348-8ef4-48a3-9ab0-cf401d7afc2c",  
"3da27fb9-7a15-446e-ae0f-258c7dd4fd94",  
"f6afff19-25ae-41f8-8a7a-6c1acffa39",  
"0862d673-39ec-4103-8a6b-66a936b14435",  
"e1a60651-bb32-4ec1-a39f-d8ffe20bd238",  
"e7a5ceb2-b1ad-4381-8fb2-2b3a409da406",  
"1d30497f-4340-43ab-b328-9fd2fed3106e",  
"ca9f46ad-5b1b-4991-90d8-ac616a302225",  
"4d40aa92-380c-4fae-98d8-c598bb32e7c6",  
"ecfe6c89-6faa-46e2-bdfd-c1fb0f0088b2",  
"7f72c34d-4515-42eb-a32e-38fe74217b70",  
"ceffc30d-1d28-48c3-acee-e6a2dc38ff37",  
"1445b0f3-7e0d-4ebc-a066-d214ceb6b51d",  
"427eb4d3-482d-48d7-9094-823526880304",  
"27e17948-02c4-4ba3-8244-5c229cc8e9b8",

"ff1fa956-ac18-4220-ab97-fc6624fc3dae",  
"5f1db605-88b6-427a-84cb-ce2fcf0f89a3",  
"baf524c6-0a2c-40f2-90a7-e19c6e1b6b97",  
"f3294e78-9a12-49ff-983e-ed6154ff621e",  
"66af728c-958b-4575-9e5e-25209343cd66",  
"241f9c62-26be-4d0f-864b-ad4b243a03c3",  
"d890e88f-16d7-4b75-bef1-5e4d09c8daa0",  
"aa2b4fca-a055-4ce9-ac77-1c8ff8b320b9",  
"88c47bd3-f8c9-4157-9d1a-770d9be8c173",  
"1b3a402c-796d-4cdf-b6c8-ce204b2d19e6",  
"f94ded4d-1c87-4503-82f1-a1ca4346e756",  
"b6c3d0ad-493e-4aec-8bc3-56f4db4d9346",  
"26d4676f-258c-4559-94d5-90bb4535070e",  
"9bdf38ca-95fd-4cf4-adf6-ed26e97cf213",  
"e8e31797-be17-4151-8f08-8008ba800992",  
"940d0880-83b5-499d-9f39-1bf30ccfc4d0",  
"a4cfb06e-a5a6-48cc-b7e5-c38228c132a8",  
"63b0e966-e81c-4d84-b41d-3445b0d9911f",  
"811fd271-b1dc-490a-a872-3d6867d59e78",  
"8a03e8d9-6cae-4794-b109-3d33c7988ef6",  
"0f26068a-b9fe-480b-9502-eb984cd6e21e",  
"488c640d-e742-465a-98b4-1234bb09d038",  
"aeed0683-e022-42d0-a954-f9f36afc4bbf",  
"960cb748-a130-4e31-9abf-486c686e12ad",  
"0d707430-18c2-4828-aa4c-a9055204166e",  
"2c57948e-d33b-4717-b1ef-1603c2c011bb",  
"d312911b-9bee-4fff-97fb-ffa509c1fd5c",  
"a8bf3374-f153-460d-902a-adea7f41d7c7",  
"2f93b300-f147-48f5-95d5-afd0e0161fe6",  
"6092f859-05fe-475d-b914-3c1a6534e6b9",  
"3f5061bc-fc92-40eb-835c-0d0452e68c67",  
"636a5aa6-1d37-4cd2-8742-4dcad8c67e0c",  
"a603437d-578e-4577-9800-645614b28b4b",  
"97019dea-e207-4909-b907-076d0620ff74",  
"b3e72102-3ae0-4fc0-ae5b-f7128afeba0d",  
"e0f748b2-984f-45cc-8b9e-13989c31dda4",  
"2ab1260f-3a24-4793-94a2-f05669a9fcd9",  
"fdd9e45a-08b4-4b98-9c34-bada71a34979",  
"89f9c84b-be5c-4020-8edc-6fbe0b1c25f5",  
"3eaa2b26-1c54-4527-8b97-3c34f4856e5b",  
"fd67d684-aaff-4260-bb94-9d0373015620",  
"79464e83-b688-4b82-84bc-18d105f60f33",  
"5147a952-3d44-4beb-8d49-b41c91bce733",  
"5bc52c24-cb4c-4625-a883-53af8d40ed0f",  
"754c1299-9b8d-41ac-a1d6-534f174fa87b",  
"b6c3d0ad-493e-4aec-8bc3-56f4db4d9346",



"757ff294-0fca-40f5-9f33-39a2d3fd5c8a",  
"229ac088-257d-48f3-9fdf-033fade595b7",  
"98ce9caa-487e-4391-86c9-e5d8129be5b6",  
"a4ff976a-ac8a-49b8-a89c-f52f3030ccaa",  
"45aa16fa-5fd5-4449-a3bd-52d734fcb0a9",  
"15ab17fe-2486-4ca5-bb47-6b804793f80d",  
"00e012c5-9232-472c-bd8b-8ca4ea6d1275",  
"501a8862-dc30-4d1e-ab47-deb9f5579678",  
"735d8962-5ec7-41ce-a73a-a43c35cc354f",  
"8cd5ce6c-d604-46ac-b4f7-1f871589d96a",  
"1d9045f8-1d8f-479c-983d-383f3a289bec",  
"f940465f-2a20-40b4-bc1c-cd4670893ce4",  
"c80467d8-d472-4643-9a50-4ac489da14dd",  
"f4a04558-958a-43da-b009-d5b7580c532f",  
"e106fb4d-743a-4cf8-96ff-4bae8d3319e2",  
"af1d6a61-66bd-47fd-a8c5-20e433f7076b",  
"ce7ec5ea-88b2-430f-92be-0f2436600a46",  
"7c1149e7-7e2b-426f-acd1-281d4273e02d",  
"b904a8eb-9c98-4bb1-bf25-3cb9d075b157",  
"ffa5c5dd-ab4b-4739-8281-275d8c1fb504",  
"3ccdbbfc-6550-49f4-8ec9-77fc94a7a099",  
"2beff5ba-a543-407e-ae7a-f0ed1788f297",  
"03da5b84-80ba-48bc-89b9-b63fc56b364b",  
"aa9059ba-930c-4327-97a1-c8c7877332c1",  
"97d205c8-32f0-4e64-a7df-bf56334be638",  
"7a4506ac-dfdc-4054-b2d1-d8fdf4cea12b",  
"5a1cd53e-640b-46e7-82a6-d95bc4907e36",  
"c8245465-2937-40fd-9c3e-7bd33deef477",  
"1bba80bd-2efd-41a2-9b09-4ff8da0efeb9",  
"aa431454-40d3-42f5-8069-381a15789257",  
"f58100fb-f99d-4b1e-9f4f-f851e34fe06f",  
"f8b57164-89e6-48b1-bd70-9d360b53a453",  
"22c74596-3585-4f6a-b98b-3083bb8b11e1",  
"5bd6eb77-5c19-40cd-8f46-dc631d726358",  
"dd9008ae-0172-4e07-b3cf-50c35c51b606",  
"11eb327c-40fa-496d-9af5-3be8bdb95e93",  
"1809f10a-d0e3-4481-8f96-cca7f240d656",  
"27602ce3-fbd6-4044-8b44-b8421670edae",  
"245c521a-5014-4da0-bf2b-35eff9673367",  
"c3d008a2-b357-4886-acc4-a2c77f1749ee",  
"3eb0d095-fc27-4add-8202-1dc2333a758c",  
"bc0440f4-df75-452d-b6ba-5cdd8ac0f7ba",  
"678f4564-d01a-4ffe-8bdb-fead78f87955",  
"b5c810e1-c847-4553-a24e-9893164d9786",  
"6813bf17-373c-49ce-b9e3-1d7ab98f2977",  
"47c71c05-a4f1-48da-b8d5-9e5ba139a8ea",

"8a27431b-b1f9-4fed-a8e0-0a0aad9d98c",  
"833eb820-8a1d-4ef1-b277-76d3a7cc877f",  
"8f256239-8104-4838-9587-ac234aedd822",  
"d4a3f6cb-3023-4088-a5f4-147fb4510874",  
"a86b8055-acbf-4e8f-a93e-cb517de68160",  
"a01f41d6-1da8-4b0b-87b4-82ecc41c6d55",  
"4009b316-9098-4130-aa70-d644d798278b",  
"bb4a0068-d7e5-4728-bf8f-9ef000574633",  
"afeac3d3-5370-4f40-8a7d-53df8aa62899",  
"0a419ac2-6c33-48cc-a9f7-32977e3b4d08",  
"010b08ba-0eed-4d21-8104-0a7913e10909",  
"8b481da1-c3ac-424f-ba50-a8354daee6b8",  
"88a8657d-d73d-4be1-b893-678b35fd709a",  
"e03a77a9-4f3f-4c09-95f6-6a02391a6525",  
"a57c6271-1af0-42c7-ac8c-e11e8a64009f",  
"dfaab9b1-940c-432f-85bf-13d387595689",  
"cb58dd26-1e37-45ab-b36a-db4e4d5e8596",  
"2594554e-02ef-4015-b619-2645020b0ac1",  
"4c6edfeb-226d-4e99-b077-a68042caa2ca",  
"82102cd2-5cb9-4f26-bf75-f27ff7476ffc",  
"f1963ed9-1b8d-43f7-af89-ee589b6e7116",  
"7c2700ae-b3ba-4771-8c2d-991309577ca6",  
"87e58255-cc79-4fd5-adba-c579493b3d35",  
"ca2f9ad1-a5fd-45b6-a662-4991cf1767a3",  
"2af98e3e-7caa-4bb8-8097-df3b0ded803f",  
"39784316-7e34-42dc-8e13-32b3a1f01a00",  
"4d0430e3-3640-4d87-8f02-cbb45f6ae83b",  
"c9957c97-8a34-4394-bb0b-aee6b4f42279",  
"1603556c-53fc-4d14-b0bf-8c18ad7b24ab",  
"534c3d13-b18b-4be5-91e6-768c0cf09361",  
"ca875697-7ffc-4b38-8868-51de34465969",  
"a17ee594-7fe8-49fe-96cf-7a464f9c5c13",  
"70562164-2f8a-46d1-9823-c71f69f258b1",  
"466dfea0-8fa6-4f10-a48e-00d5a6c8d92c",  
"cd7c07af-840b-483b-af13-94320cb9aaaf",  
"1628a7f6-38e8-443f-8849-acbb2268b4e8",  
"0aeaad1b-4c63-4ebc-9fde-2f75939d65e9",  
"34f68e0d-fafc-4ae6-9ced-2706aad79ea5",  
"b8580994-0b8b-42ba-b5a8-3e7e331aa948",  
"7e078f07-4424-4a5b-bbe9-fb597498f64f",  
"242cdb6d-1d99-41ac-b9d9-4dfdb403cc2b",  
"ebf014ee-502d-4ed9-ad1e-90c3fcf20435",  
"3a80885a-8d06-4fd8-8dcc-1eed76b582f5",  
"797b87d1-f6ef-4081-bab4-c4ad66d8b315",  
"c7f76fbe-f90e-4663-ae9c-108921027ac7",  
"2253da3a-0c12-4f5b-b21d-828d4720859c",

```
"3b817d8c-864e-4050-90e2-371c64a3e22e",  
"fecc033f-919b-4d98-ab32-c6fb19dfb7e8",  
"31423bb7-3107-47f5-a24f-1de070c2ed47",  
"0e5b4753-8bfe-4311-bc2a-13bf5f52d167",  
"6da2123c-c8ae-4c61-b40d-2080474e2efb",  
"aa6c36bf-0d86-4768-8575-f4c630efdfa1",  
"e20a5d20-78a2-47b4-b15e-2daab8effb56",  
"7145c373-a655-4d83-bfa2-9cdfae8184b9",  
"0f0b5d24-cc43-411e-8a3b-902e9d7e796b"  
"c754cfd7-a34c-4cd5-ad6b-ea28437f0741",  
"7fd1f5fd-1006-4b40-9e4c-1de386a7c813",  
"33d0a518-f82b-46e0-939b-bb4c7b9fb62f",  
"64ac7d98-d8bd-41e7-a661-bef4f090b439",  
"11374c79-b2b1-464e-95c7-94aec6f24099",  
"824e1b3c-5641-4545-a438-723f74bf4bf2",  
"f1badc5e-5363-45e7-8577-845a32ea340c",  
"6f0a56be-6b4f-43c7-a775-4167cbea9504",  
"19d21671-246e-4387-9ac7-7e1a938240f9",  
"2adf5b44-a6ae-4877-9963-3286ed9428f7",  
"e8d8f5d7-3348-4f25-8716-5807c0350899",  
"d3cb2d79-746d-4a75-b773-e58b46353bcf",  
"34d9a282-6232-4bb2-8b55-9ccf5a751b0f",  
"f6de421b-4eac-4a59-8138-e8716509f557",  
"d9766275-a9ba-4276-9188-0005fd7086af",  
"1b4311ff-58e8-43ef-bb1b-b42d25bc4b2d",  
"612a3506-2ec0-4ae3-baef-3f4f4c9d7772",  
"4071e86c-433e-4a46-a84a-47f69ac945ac",  
"ec0cc90a-1765-4ef2-8829-9f2342c9c922",  
"a973a771-e305-4582-9f2e-f81c4d925172",  
"c4cc39f9-d72b-4e83-9813-e98f87186331",  
"8cdf585-a3bf-41ac-969c-5f29d6993604",  
"1ebc4e44-344c-42a7-876a-85e2b07a3bec",  
"6244a086-636e-4756-bf5e-f803a78ffbc4",  
"d8b26c64-3054-4f6a-a632-1225fa769dc8",  
"29a54313-4659-4a83-825c-fa01b3284a21",  
"5629d4a6-95ab-4561-8de6-21a674eff06d",  
"412acefd-6a82-429c-b9c5-9e5d96a27afb",  
"521d09b6-6a91-496b-8f6f-1947d3968a23",  
"0604d827-a0f8-4e31-9644-17c9c854c5be"]  
start_date = "2022-11-01"  
end_date = "2022-11-30"
```

```
# Create a list to hold the access data  
access_list = []
```

```

# Iterate over the list of item IDs and retrieve the access data for each item
for item_id in item_ids:
    # Build the API call URL for the current item ID
    url = f"https://be-api.us.archive.org/views/v1/detail/item/{item_id}/{start_date}/{end_date}"

    # Make the API call and extract the relevant access data
    response = requests.get(url).json()
    access_data = response["counts_geo"]

    # Add each count_kind in the access data to the list as a separate row
    for count_data in access_data:
        count_kind = count_data["count_kind"]
        geo_country = count_data["geo_country"]
        sum_count_value = count_data["sum_count_value"]
        access_dict = {"start_date": start_date, "item_id": item_id, "count_kind": count_kind, "access":
sum_count_value, "country": geo_country}
        access_list.append(access_dict)

# Write the access data to a CSV file
filename = f"access_data_{start_date}.csv"
with open(filename, 'w', newline='') as file:
    writer = csv.writer(file)
    # Write the header row
    writer.writerow(["start_date", "item_id", "count_kind", "access", "country"])
    # Write the data rows
    for access_dict in access_list:
        access = access_dict.get("access")
        country = access_dict.get("country")
        writer.writerow([access_dict["start_date"], access_dict["item_id"], access_dict["count_kind"],
access, country])

```

# GraphQL Script to extract Author Affiliations from the Thoth Catalog

```
{ affiliations(publishers: ["9c41b13c-cecc-4f6a-a151-be4682915ef5"], limit: 9999) { institution { institutionName countryCode updatedAt } contribution { fullName contributionType contributor { orcid } work { title subtitle workType workStatus publicationDate landingPage updatedAtWithRelations} } } }
```

When you want these data for all publishers:

```
{ affiliations(limit: 9999) { institution { institutionName countryCode updatedAt } contribution { fullName contributionType contributor { orcid } work { title subtitle workType workStatus publicationDate landingPage updatedAtWithRelations} } } }
```

# GraphQL Script to Get Books Published per Year

```
{  
  books(order: {field: PUBLICATION_DATE, direction: ASC}, publishers: ["9c41b13c-cecc-4f6a-a151-be4682915ef5"], limit: 9999, workStatus: ACTIVE) {  
    fullTitle  
    doi  
    publicationDate  
  }  
}
```

# Extracting DOIs from a PDF (requires manual check still because of line breaks)

```
import pdfplumber
import re
import csv
import os
from datetime import datetime

def extract_dois_from_pdf(pdf_path):
    """
    Extracts DOI numbers from the given PDF file and returns a list of DOIs.
    Removes the final period if it's part of the DOI.
    Fixes DOIs split across line breaks.
    """
    dois = []
    with pdfplumber.open(pdf_path) as pdf:
        for page in pdf.pages:
            # Extract text from the page
            text = page.extract_text()
            if text:
                # Step 1: Join the lines into a single block of text
                text = ' '.join(text.splitlines())

                # Step 2: Fix broken DOIs that split after a period (.) or a slash (/)
                # We look for a line break after a period or slash and then merge it with the next part.
                text = re.sub(r'([\.\V])\s*(?=\d)', r'\1', text) # Merge split DOI after period or slash

                # Step 3: Use regex to find DOI numbers
                doi_pattern = r'\b10\.\d{4,9}/[-._;()/:A-Z0-9]+(?:\b|\s|$)'
                found_dois = re.findall(doi_pattern, text, re.IGNORECASE)

                # Step 4: Remove the final period if it's part of the DOI
                cleaned_dois = [doi.rstrip('.') for doi in found_dois]
                dois.extend(cleaned_dois)

    return dois
```

```

def save_dois_to_csv(dois, pdf_path):
    """
    Save the extracted DOIs to a CSV file with consecutive numbering.
    The output file name contains the input PDF name and a timestamp.
    """
    # Get the base name of the input PDF file (without extension)
    base_name = os.path.splitext(os.path.basename(pdf_path))[0]

    # Get the current timestamp in the format YYYY-MM-DD_HH-MM-SS
    timestamp = datetime.now().strftime('%Y-%m-%d_%H-%M-%S')

    # Generate the output CSV file name
    output_csv_path = f"{base_name}_{timestamp}_dois.csv"

    with open(output_csv_path, mode='w', newline="", encoding='utf-8') as file:
        writer = csv.writer(file)
        # Write DOIs with consecutive numbering in the first column
        for idx, doi in enumerate(dois, start=1):
            # Prefix "https://doi.org/" to each DOI
            doi_url = f"https://doi.org/{doi}"
            writer.writerow([idx, doi_url]) # Write the index and the prefixed DOI

    print(f"DOIs have been saved to {output_csv_path}")
    return output_csv_path

def main():
    pdf_path = '0446.1.00.pdf' # Path to the input PDF file

    # Extract DOIs from PDF
    dois = extract_dois_from_pdf(pdf_path)

    if dois:
        print(f"Found {len(dois)} DOIs. Saving to CSV...")
        # Save the extracted DOIs to a CSV file
        save_dois_to_csv(dois, pdf_path)
    else:
        print("No DOIs found in the PDF.")

if __name__ == "__main__":
    main()

```



# Uploading extracted DOIs to Thoth

```
import thothlibrary
import csv
```

```
# Log in to Thoth
email = "XXX@punctumbooks.com" # Replace with your email
password = "XXX" # Replace with your password
```

```
thoth = thothlibrary.ThothClient()
thoth.login(email, password)
```

```
# Define the work_id (this should be specific to the work you're uploading references for)
work_id = "6244a086-636e-4756-bf5e-f803a78ffbc4" # Replace with the actual work_id
```

```
# Path to your CSV file containing the DOI data
csv_file_path = "0446.1.00_2025-02-04_11-33-41_dois.csv" # Replace with your CSV file path
```

```
# Read and process the CSV file
with open(csv_file_path, mode='r', newline='', encoding='utf-8') as file:
    reader = csv.reader(file)
```

```
for row in reader:
    if row: # Ensure the row is not empty
        reference_ordinal = row[0] # First column: reference ordinal
        doi = row[1] # Second column: DOI
```

```
# Construct the reference dictionary
reference = {
    "workId": work_id,
    "referenceOrdinal": reference_ordinal,
    "doi": doi,
    "unstructuredCitation": None,
    "issn": None,
    "isbn": None,
    "journalTitle": None,
    "articleTitle": None,
    "seriesTitle": None,
    "volumeTitle": None,
```

```
"edition": None,  
"author": None,  
"volume": None,  
"issue": None,  
"firstPage": None,  
"componentNumber": None,  
"standardDesignator": None,  
"standardsBodyName": None,  
"standardsBodyAcronym": None,  
"url": None,  
"publicationDate": None,  
"retrievalDate": None,  
}
```

```
# Upload the reference to Thoth and get the reference ID  
reference_id = thoth.create_reference(reference)  
print(f"Reference ID for DOI {doi}: {reference_id}")
```