

Accounting for Secrets: Data Appendix

Mark Harrison

As described in Table 1, by May 2011 the Hoover Archive had acquired more than five thousand files of the KGB of Soviet Lithuania, containing just over one million microfilmed pages (see Table 2).¹ Table 2 presented results of an analysis of their content, based on keywords and keyword clusters. The appendix describes the sources and methods underlying Table 2.

The first step was to download the electronic descriptions of the KGB's management files (opisi 2, 3, 10, and 14, as explained in the text), and merge them. The size of the resulting file was 86,225 words and numbers.

The second step was to derive keywords. Initially I eliminated all numbers, all words of one or two letters, and conjunctions, pronouns, and other common non-substantial words (thus *dlia*, *drugie*, *ego*, *est'*, *kotoroi*, *nad*, *nim*, *nimi*, *nikh*, *pered*, *pod*, *posle*, *pri*, *protiv*, *sredi*, *tak*, *cherez*, *eto* and derivatives, *razn*-and *proch*-). I converted three words that were sometimes spelt out in Roman to Cyrillic (*Abstelle*, *LSSR*, and *Ostland*).

Because nouns decline and verbs conjugate in the Russian language, I sorted the remaining words by their first five letters (or fewer). At this point many words required disambiguation, for example, *agent* from *agentura*, *akt* from *aktiv*, and *antisovetskii* (955 occurrences) from *antisemiticheskii* (one occurrence). I went through the list by hand, sorting shorter words on the first 4 or 3 letters and longer words by up to 11 letters. I grouped words reflecting the changing historical nomenclature of unique organizations such as the NKGB-MGB-KGB. The final list comprised 1,400 unique keywords.

At this point I assigned each keyword to one of three headings, "operational" (oriented towards external goals, such as the suppression of hostile activity), "organizational" (oriented towards internal goals, such as integration with other party and state institutions), and "procedural" (oriented towards compliance with internal norms of bureaucratic behaviour). For general interest, Table A-1 reports the top 40 keywords under each heading and the frequency with which they occurred.

Based on the ranking of different keywords in Table A-1 and my own reading of the material, I set out to determine the overall frequency of a number of keyword clusters. Final results are reported in Table 2.

¹ These files are described at <http://www.hoover.org/library-and-archives/collections/east-europe/featured-collections/lietuvos-ssr> (accessed April 8, 2011).

Intermediate results are reported year by year in Table A-2; this table is also the basis of Figures 2 to 4.

Table A-1. Lithuania KGB management files, 1940 to 1991: Frequency of keyword stems

Rank	Operational	N	Organizational	N	Procedural	N
1	партизан– (partisan)	2373	НКГБ, МГБ, КГБ– (People's Commissariat– (Ministry, Committee) of State Security)	4398	документ– (document)	1667
2	борьб– (struggle)	1404	ЛССР– (Soviet Lithuania)	3094	отчет– (report)	1154
3	антисов– (anti-Soviet)	955	отдел– (department)	1909	справ– (reference report)	1116
4	подпол– (underground)	616	МВД– (People's Commissariat– (Ministry) of Internal Affairs)	1774	сообщ– (communication)	991
5	националист– (nationalist)	567	рабо– (worker, employee)	1430	дел– (file)	868
6	настроен– (mood)	460	управлен– (administration)	1353	спис– (list)	523
7	провер– (verification)	421	уезд– (rural district)	1205	акт– (deed)	521
8	ЛНД– (Lithuanian Popular Movement)	308	лиц– (person)	1059	перепис– (correspondence)	465
9	иностран– (foreign)	228	агентур– (agent network)	813	свод– (briefing)	424
10	автор– (author)	204	проявл– (manifestation)	794	данны– (data)	414
11	уби– (kill)	204	район– (urban district)	747	журнал– (ledger)	405
12	преступ– (criminal)	192	Литовск– (Lithuanian)	711	регистр– (register)	382
13	аноним– (anonymous)	189	оператив– (operational, operative)	592	стат– (statistical)	381
14	молод– (young, youth)	176	СССР– (USSR)	515	план– (plan)	352
15	бандит– (organized criminality)	175	формирован– (formation, usually military)	496	переда– (transfer)	307
16	служ– (service, usually religious)	170	сет– (network)	461	уничтож– (destruction)	284
17	духов– (ministry, religious)	135	аппарат– (apparatus)	438	прием– (acceptance)	279
18	бойц– (fighter)	134	уполномоч– (plenipotentiary)	410	материал– (documentation)	190
19	защит– (defence)	131	розыск– (manhunt)	373	протокол– (minutes)	178
20	немец– (German)	125	спец– (special-purpose)	361	опис– (description)	163
21	оккупаци– (occupation)	92	организ– (organized, organization)	349	заявлен– (request)	154
22	Литв– (Lithuania, pre-Soviet)	91	действ– (action, activity)	337	жалоб– (complaint)	150

Rank	Operational	<i>N</i>	Organizational	<i>N</i>	Procedural	<i>N</i>
23	поль– (Polish)	74	город– (city)	313	информ– (information)	137
24	ценност– (valuable, n.)	65	выявлен– (revealed)	312	учет– (account)	132
25	выбор– (electoral)	61	разраб– (processing, processed)	309	перевод– (translation)	110
26	вооружен– (armed)	58	арест– (arrest)	294	вопрос– (question)	106
27	промышл– (industry)	55	деятельност– (actuality, reality)	289	главн– (chief)	105
28	заключ– (prisoner, imprisoned)	54	област– (province)	254	секрет– (secret)	94
29	военн– (military)	51	государ– (state)	232	результат– (result)	80
30	профилакт– (preventive warning)	50	указ– (instruction, decree)	220	Приказ– (decree)	79
31	заграниц– (trans-border)	45	групп– (group)	200	распоря– (instruction)	79
32	листово– (leaflet)	45	агент– (agent)	200	ответ– (answer)	75
33	пересыл– (resettled)	45	оперучет– (operative)	181	повод– (subject)	72
34	транспорт– (transport)	42	развед– (intelligence)	180	факт– (fact)	58
35	возвра– (return, usually from abroad)	41	населен– (population)	178	делопроизв– (file-keeping)	53
36	легенд– (caption)	39	полити– (political)	178	недостатк– (deficiency)	53
37	америк– (American)	35	след– (clue, investigation)	172	ведомост– (gazette)	52
38	дорог– (road)	34	обстановк– (circumstance)	170	полож– (position)	51
39	желез– (iron, railroad)	34	назыв– (called, as in "so-called")	165	пис– (letter)	50
40	поезд– (train)	31	ликвид– (liquidation)	162	издан– (published)	49

Source: Lietuvos SSR Valstybės Saugumo Komitetas selected records, 1940-91, held at the Hoover Archive and described at <http://www.hoover.org/library-and-archives/collections/east-europe/featured-collections/lietuvos-ssr> (accessed April 8, 2011).

Table A-2. Lithuania KGB management files, 1940 to 1991: number and distribution by keyword cluster and over time

	All files	Counter-insurgency	Secret file management	Of which: Storage security	Of which: Distribution security	Police work	Matters relating to foreigners	Complaints and petitions	Economic matters	Anonymous circulars	Matters relating to young people	Preventive work	Matters relating to Jews	Not classified
1940	5	0	0	0	0	1	0	0	2	0	0	1	0	1
1941	13	3	0	0	0	3	4	0	0	0	0	0	0	5
1942	2	1	0	0	0	0	1	0	0	0	0	0	1	0
1943	7	5	1	0	1	0	1	0	0	0	0	0	0	0
1944	133	106	3	3	0	15	13	0	0	2	1	0	4	6
1945	259	232	7	7	0	6	4	0	0	1	1	0	0	13
1946	224	189	10	8	2	7	6	0	2	0	0	0	0	14
1947	232	170	39	38	1	6	4	1	2	1	0	0	0	12
1948	217	183	7	6	1	12	1	0	1	0	0	0	1	14
1949	187	159	11	8	3	8	4	0	1	0	0	0	1	4
1950	165	133	13	11	2	9	4	0	3	1	3	0	1	8
1951	133	93	4	3	1	9	16	0	0	2	2	0	0	14
1952	144	104	10	9	1	14	13	0	2	3	2	0	0	6
1953	271	136	50	50	2	66	9	0	2	35	16	0	1	15
1954	50	15	21	14	7	6	3	0	0	1	3	0	0	7
1955	50	4	23	20	3	14	4	0	0	4	3	0	0	9
1956	45	4	19	17	2	11	3	0	3	1	4	0	0	8
1957	62	16	16	14	2	16	6	0	3	4	5	0	0	11
1958	30	5	17	15	2	5	2	0	1	1	1	0	0	2
1959	57	11	22	20	2	10	5	0	2	1	1	3	2	8
1960	46	0	27	25	2	6	3	0	1	2	0	2	0	11
1961	31	0	18	14	4	6	1	0	0	1	1	2	0	7
1962	21	0	9	9	0	3	0	1	1	2	1	2	0	5

	All files	Counter-insurgency	Secret file management	Of which: Storage security	Of which: Distribution security	Police work	Matters relating to foreigners	Complaints and petitions	Economic matters	Anonymous circulars	Matters relating to young people	Preventive work	Matters relating to Jews	Not classified
1988	10	0	6	1	5	0	0	0	0	0	0	0	0	4
1989	3	0	2	0	2	0	0	0	0	0	0	0	0	1
1990	4	0	2	0	2	0	0	0	0	0	0	0	0	2
1991	3	0	2	0	2	0	0	0	0	0	0	0	0	1
Total	3433	1597	805	436	371	392	199	108	103	101	78	47	19	351

Notes: The assignment of files to categories is to some extent lexicographic, but does not exclude overlaps as noted below. A file is assigned to secret file management (distribution security) if its description has at least ONE occurrence of журнал– AND one of регистрации; OR one of реестр–; AND one of документ– OR приказ– OR указ– OR распоряж–; and to secret file management (storage security) if its description has at least ONE occurrence of акт (as the beginning of a word) OR акты OR ведомост– OR опис–; AND one of документ– OR материал– OR дел–; AND one of прием– OR передач– OR архив– or уничтож–.

Assignment to distribution and storage security may overlap and two files are assigned to both.

A file is assigned to counter-insurgency, provided it has not already been assigned to secret file management, if its description has at least one of бандит–, партизан–, подпол–, or защит–.

Files are assigned to other categories, provided they are not already assigned to secret file management or counter-insurgency, based on the occurrence of specific word stems in their descriptions as follows: police work (BOTH of спис– AND лиц–; OR ONE of след– OR розыск– OR антисов– OR националист– OR преступ–); matters relating to foreigners (ONE of загран– OR уехать– OR иностран– OR америк– OR немец– OR поль– (as the beginning of a word) OR изра– OR турист –); economic matters (ONE of промышл– OR транспорт– OR поезд– or OR перевоз–); anonymous circulars (ONE of листово– OR аноним–); complaints and petitions (ONE of заявлен– OR жалоб–; AND one of жителей); matters relating to young people (молод–); Preventive work (профилак–); matters relating to Jewish people (ONE of евре– OR сион–). Not classified: NONE of the above.

Categories other than counter-insurgency and accounting for secrets may overlap, and 281 files are assigned to more than one category.