

清末台湾之“淡新档案”及其整理

吴密察*

一、前言

所谓“淡新档案”，是目前典藏于台湾大学图书馆的一批清末台湾北部新竹县、淡水厅之州县官府档案。本文将介绍这批档案的大致轮廓与半世纪以来对这批档案的整理经验。这样的经验，可以说是不断尝试错误，也同时是因应科技的发展和时代需要而调整、深化的过程。希望这样的经验，可以成为目前各地正在陆续展开之清代州县官府档案整理工作的参考。

二、一九九〇年以前的整理

“淡新档案”系一八九五年日本占领台湾后，由日本政府接收并以其中内容多有涉及司法事务者而交由新竹地方法院管理，后转藏覆审法院（以后改为高等法院）^①。一九二八年，台北帝国大学成立，文政学部史学科之主要研究主题虽为以日本涉外史相关之南洋史与东洋史，兼及台湾原住民研究，但自

* 吴密察，台湾大学历史系教授。

① 戴炎辉：《清代淡新档案整理序说》，《台北文物》1953年第2卷第2期。

一九三六年六月进而开设台湾史料调查室,致力于调查史迹、搜集台湾史料^①。在这样的背景下,一九三七年六月二日,透过前台北地方法院长大里武八郎居中奔走,台湾史料调查室自高等法院(院长斋藤三郎)承接了清朝时代古文书一千一百四十袋数千件,当时对这批“古文书”的评价是:“视其内容、数量,不仅乃台湾史研究之根本资料,而且也是支那法制史研究上之贵重史料。”^②当时称此批档案为“台湾文书”,虽曾略作整理,但似未有人利用来从事学术专题研究^③。

一九四五年,国民政府接收台北帝国大学改称国立台湾大学(在极短的时期,曾自日本时代的校名径改为国立台北大学),这批“台湾文书”也便由台湾大学承接。这批档案虽然被台北帝国大学保存了下来,但战后接收时的保存情况却已非常堪忧,而且没有留下档案的目录。戴炎辉教授描述当时的档案情况是:“淡新档案内之案卷,本来每一案卷各有卷皮,案内各文件糊粘为一宗。日人承接此档案后,将一案卷收存于纸袋之内,袋面上用日文作简单案由,再装纸箱里。极少部分的案卷,日人抄作册子,又些许部分,作过简单的说明。……对此“淡新档案”,(日人)似乎尚未作过学术的检讨,也没有留下档案的目录。档案历时经久,收藏缺佳,致大部分为纸鱼所蚀。又少部分在抗战

① 台北帝国大学文政学部史学科设置台湾史料调查室之记事,见于《台北帝国大学文政学部史学科研究年报》第3辑(1936年),《汇报》栏,第381—382页。根据此记事,此调查室之工作在于“搜集、整理台湾关系资料,从事岛内史迹之调查,大致区别原住民时代、西班牙、荷兰时代、郑氏时代、清朝时代、日本领台以后,汇集关系各国之根本资料文献”。至于工作之实际展开,则可见《台北帝国大学文政学部史学科研究年报》第4辑(1937年),《汇报》栏,第566—578页。

② 《清朝时代古文书搜集》,《台北帝国大学文政学部史学科研究年报》第4辑(1937年),《汇报》栏,第579页。

③ 日本时代初期,殖民政府曾经参考这批档案了解新竹地区的地方行政。但是否有人曾经利用淡新档案从事研究,笔者寡闻,未能确定。唯戴炎辉教授则在不同时期有不同说法,在《清代淡新档案整理序说》(1953年)的阶段,戴教授表示:“余尚未发见日人用到此‘淡新档案’的迹象。故可以说,对此‘淡新档案’,似乎尚未作过学术的检讨,也没有留下档案的目录。”但在《淡新档案选录行政编初集》的《弁言》(1971年),戴教授则改口谓:“各书刊既录有清代档案,日人似亦曾利用过这淡新档案(不过未知这些档案是否原属淡新档案)。”

中，因搬移受炸或潮湿，致烂破、残缺，且很散乱。”^①

一九四七年底起，台大法学院戴炎辉教授，获得林熊征学田、中华文化教育基金会之经费支持，及法学院陈棋炎先生、学生曾琼珍、法学院教务分处讲义股同仁之帮助，在做了必要的杀虫工作之后，依照规划进行整理。戴教授自述整理工作之作法及其工作经过是：

1. **暂编号码** 全部案卷杀虫后，略加以检讨，可认为一个案卷者，随收入新纸袋，暂编其号码。

2. **裱褙修补** 案卷大破者，加以裱褙，小破者加以修补，而把一案卷内的文件粘连起来。此工作甚繁重，幸得法学院陈棋炎先生，学生曾琼珍君及假期中的法学院教务分处讲义股同仁的帮助。

3. **初步记录** 摘记各案卷的案由、年代月日及关系人，又每一案卷内的各文件，也摘记其年代月日、发文者及收文者，必要时又附以简注，以便于初步分类。

4. **初步整理** 凡案卷存有案由者，按照其案由；而无案由者，则检讨案卷的内容，大略分为行政、民事及刑事三门，做成整理表，附以整理号码，计有行政案六百八十六，民事案二百六十五，刑事案四百十五案。同时为防止混乱，且便于检索，另作暂编号码与初编号码对照表。

5. **初步分类** 经过初步整理，已有眉目，故据此将行政、民事及刑事各门，分作若干类，而制作初步分类表，以便于分离并合。

6. **分离并合** 大多数案卷，固然是完全的，但分散者，为数也不少。其分散程度，有达十余处之多的。又年代月日前后错杂，案卷的粘连，难免也有错误，故非经过一番分离并合不可。因此，此前所做的案卷初步记录表，也必须加以修正。此工作够繁重，非借分类之功，便不能把各文件妥归原处。

7. **缮抄工作** 修正记录表之后，随时加以缮抄。但仍可发见某案内之某文件不得其处，须移到该案卷内的另一处，或移到另一个案卷，遇到这种情形，随即加以补正或重抄。

8. **确定的分类** 初步分类，是便于分离并合的。全部案卷究有几种有几

① 前揭戴炎辉《清代淡新档案整理序说》，第2—3页。

案,尚未可知,又各案卷之正确年代月日,尚未查到。故经初步分类之后,尚须由学术上的观点,作有系统的分类。这确定的分类,仍分为行政、民事及刑事三门。门之下分为类;而类之下再分为款。各款下之各案卷,则按照年代月日序列之。又每一案卷内之各文件,也按照年代月日编号。三门总共案卷达一千一百六十三案之多。

9. 确定的记录表 每一案卷,按照分类表次序,记录其案由(原存有案由者照抄,无案由者或已散佚者,则检讨内容补充之)、案内每一文件之年代月日、发文者、收文者、内容概要及附注。

10. 简明分类表 为检索方便,另作简明分类表,摘记各案卷要项,尤其年代月日、关系人姓名、附注及简明案由。

11. 事项别卡片 已作的案卷记录表,不过摘记其形式上的属性而已。如果要把此档案供作学术研究的材料,须依实质的内容,制作事项别卡片。因此,须详查各案卷内的每一个文件,制作卡片,然后将卡片加以分类。但这工作需要很长时间,现在尚未着手。^①

经过戴炎辉教授整理后,这批古文书的轮廓已大致可以掌握。戴炎辉教授以这些古文书乃是清朝时代台湾淡水厅、新竹县的档案,故改名为“淡新档案”。从此,这个名称便被承袭了下来。

根据戴炎辉教授对于内容的简要介绍,我们得知这批古文书有几个部分:

1. 嘉庆十七年(1812)至光绪五年(1879)三月间的淡水厅档案。当时淡水厅的辖区,北自鸡笼(基隆),南至大甲溪北岸。

2. 光绪五年(1879)三月至光绪廿一年(1895)的新竹县档案。新竹县的辖区,在光绪五年(1879)三月至光绪十三年(1887),是原淡水厅范围割去桃涧堡(桃园)以北;光绪十三年(1887)后,又割去中港溪及支流南港溪以南。

3. 光绪初年的台北府档案。这是因为光绪元年(1875)设台北府后,台北府署未筑成之前,暂假旧淡水厅署(新竹城),而且新竹县也未派正印官,而由台北知府兼摄,台北府、新竹县可谓合署办公,所以这段短期间的台北府档案也留在新竹县。

^① 前揭戴炎辉《清代淡新档案整理序说》,第3—4页。

戴炎辉教授在整理告一段落之后，于《台北文物》第二卷第二期（1953年）发表《清代淡新档案整理序说》，简要地介绍了淡新档案的内容及其整理过程，并以其锐眼指出了淡新档案的价值：“本档案虽稍有损失，还算是整套的档案。又虽不是全台湾的档案，但从上级机关发下的‘札饬’或示谕，仍可窥测府政或省政之一端。何况，由此档案可以明了一厅一县的政治，（作）为研究数据很有价值”，而论断：“本档案不但是台湾法制史研究的好材料，对地方政治、社会及经济史研究，相信也有贡献”。^①

戴炎辉教授的整理工作，对于以后的研究与整理都具有重要的意义，值得稍加说明。

首先，戴炎辉教授确立了以“案”为单位的整理原则。从这批档案被发现时候的状况来看，清代衙门就曾经对档案进行过基本的整理。至少，每一个诉讼案件的相关文件是被汇编在一起的，而且这些相关文件还被以时间之先后，依序地黏贴成一长条状。即使不是诉讼案件的地方行政性相关档案，也已经将相同业务主题的文件汇合在一起，并依照时间顺序黏贴着。例如，编号31601便是有关验尸的历年（从同治十年到光绪十年）各种公文的汇辑，共15件。这种原来就以文件集存在的形式，可能就是台北帝国大学接手这批档案时号称有“一千一百四十袋”的原因，盖当时应该就是以清代衙门已经黏贴在一起的一份文件集为一袋。也就是戴炎辉教授说的“每一案卷各有卷皮，案内各文件糊粘为一宗”。戴炎辉教授整理时基本上尊重这种清代原有的整理，只是将部分显然错置或散页的文件加以调整。

其次，戴教授将这1100余“案”的档案，以“案”为单位，依照其内容之性质加以分类。戴教授以清代地方行政与近代法律的概念制作了一个以“编”统“类”，以“类”统“款”的三阶层分类表，并各赋予号码（表一《淡新档案分类表》）。因此，每一“案”都有一个相应的号码。例如，32410就表示是第3编（“刑事”编）第2类（“人身自由”类）第4款（“拐奸”款）的第10案。如果要更进一步进入“案”里的每一个文件，则在“案”码之后的小数点中表示。例如，32410.0就表示是32410案的“案由”，32410.2表示32410案的第2件；

^① 前揭戴炎辉《清代淡新档案整理序说》，第5页。

表一 淡新档案分类表

第1类 总务

第2类 民政第3类 财政

第4类 建设

第5类 交通

第6类 军事第7类 抚垦

●第2编 民事

第 1 款 失踪	第 2 款 结婚	第 3 款 离婚	第 4 款 收养	第 5 款 监护
-------------	-------------	-------------	-------------	-------------

第2类 田房

续表

	第 1 款 租借	第 2 款 抗租	第 3 款 霸收	第 4 款 霸占	第 5 款 争界	第 6 款 争财	第 7 款 公业	第 8 款 用水	第 9 款 抄押
第 3 类 钱债									
	第 1 款 买卖	第 2 款 典当	第 3 款 胎借	第 4 款 借贷	第 5 款 委寄	第 6 款 讨物	第 7 款 抗吞	第 8 款 抗算	第 9 款 匿契
第 4 类 商事									
	第 1 款 郊行	第 2 款 合股	第 3 款 倒闭	第 4 款 侵权					
●第 3 编 刑事									
第 1 类 总务									
	第 1 款 法令	第 2 款 册报	第 3 款 指挥	第 4 款 互助移解	第 5 款 通缉	第 6 款 验尸	第 7 款 定讞	第 8 款 执行	第 9 款 其他
第 2 类 人身自由									
	第 1 款 人命	第 2 款 殴伤	第 3 款 勒索	第 4 款 诱拐	第 5 款 略卖	第 6 款 掳禁	第 7 款 抢掳		
第 3 类 财产侵夺									
	第 1 款 窃盗	第 2 款 抢夺	第 3 款 强盗	第 4 款 强盗杀伤	第 5 款 海盗	第 6 款 恐吓	第 7 款 诈欺	第 8 款 盗卖	第 9 款 毁弃
第 4 类 公共秩序									
	第 1 款 匪徒	第 2 款 侵害水源	第 3 款 骚扰	第 4 款 谣言					
第 5 类 风化									
	第 1 款 忤逆	第 2 款 毁坟	第 3 款 通奸	第 4 款 拐奸	第 5 款 赌博				

戴教授透过上述整理的结果，统计出这批档案总共 1163 案，以行政编最多（574 案），刑事编次之（365 案），民事编最少（224 案）^①。如果将民事与刑事均视为司法关系案件，则司法相关者共有 589 案，其案数与行政编相当。戴炎辉教授并指出：行政编除了总务类之外，其重点主题在于财政与抚垦。其次

① 但以后陆续有所增损，因此数量上稍有变动。详见下文。

多为乡保行政。民事编则以田房事项为数最多,其中又以霸占与争界占大部分;其次则为钱债。刑事编除总务类之外,财产侵夺类占一大部分,其中窃盗案占多数;其次则为人身自由类^①。

如果以每一案来比较,每案所包含的件数因案情内容繁简或时间拖延久暂长短,而极为悬殊,少则仅数件(甚或1件),多则200余件,而且拖延多年(例如,编号32611案自光绪十三年诉讼至光绪二十一年割台前夕)。根据近年台湾大学图书馆的调查,超过100件的“大案”有31案,其中以编号22202(民事门·田房类·抗租款·第2案)最大,有223件;其次则是编号33319(刑事门·财产争夺类·强盗款·第19案),有217件^②。一般来说,行政门抚垦类及民事门田房类有较多“大案”。

但是,戴炎辉教授的上述对于各“案”之分类与各类“案”之统计结果,只能当成参考,不能径视为当然。例如,戴教授归入行政编的很多案就是诉讼案件,如果将之归入民事编或刑事编也未尝不可。甚至,戴教授对于每一个“案”的归类,也存在着各种问题。例如,一个“案”在诉讼之初控诉者号称被掳禁,因此被归类为刑事编的“人身自由”,但随着诉讼的进行、案情的发展,反而让人觉得应该归入民事编的“钱债”。因此,戴教授所做出来的分类,值得斟酌者,所在多有。因此直接以戴教授的分类为基础,进行加减统计甚至进一步解释,都无甚意义。但是,因为戴教授已经在这个分类的基础上,赋予每一“案”、每一“件”一个确定的号码,以后的整理者也就都尊重而不加以更动。

一九六八年七月,戴炎辉教授应美国西雅图华盛顿大学之邀,前往该校作专题研究,因此将淡新档案全部携赴该校拍摄为35 mm的黑白微卷(共33卷)^③。戴教授赴美途中也在日本东京大学法学部发表演讲,介绍淡新档案^④。事后,华

① 前揭戴炎辉《清代淡新档案整理序说》,第4—5页。

② 吴祖善:《清代“淡新档案”:台大图书馆的新特藏》,《“中国图书馆学会”会报》1987年第40期,表十三,第156—157页。

③ 高志彬:《淡新档案目录稿(一)》,《台北文物》1987年直字第79期。

④ 此次演讲内容曾被整理发表为戴炎辉:《清代台湾诉讼手续——淡新档案资料》,《“国家学会”杂志》1968年第81卷第3、4号。

大的 David C. Buxbaum 也以英文发表文章介绍这批档案^①。后来华盛顿大学曾允许研究机关在负担费用的条件下，从其负片制作正片副本。国内外几处研究机关（例如，东京大学法学部研究室、“中央研究院”历史语言研究所）便是从华盛顿大学购入正片副本的。因此，淡新档案不但广为内外学界所知，在 20 世纪 80 年代起也开始有一些研究者得以利用这批淡新档案的微卷从事专题研究。

自从战后戴炎辉教授着手进行整理之后，便一直由戴教授保管的淡新档案，在 1986 年 7 月转由台湾大学图书馆收藏。台大图书馆趁此机会曾重新装袋清点，结果与三十年前戴教授首次披露的数量稍有出入，实得 1139 案，19244 件。

	戴教授初步整理 (附注 1)	戴教授确定分类 (附注 2)	西雅图华大微卷	入藏台大之原件
行政	686 案	574 (+1) 案 (附注 3)	561 案	564 案 9207 件
民事	265 案	224 案	218 案	218 案 5150 件
刑事	415 案	365 案	361 案	357 案 4887 件
总计	1366 案	1163 (+1) 案 (附注 3)	1140 案	1139 案 19244 件

附注：(1) 戴炎辉《清代淡新档案整理序说》，《台北文物》1953 年第 2 卷第 2 期。

(2) 同上。

(3) 根据台大图书馆清点得知，戴炎辉教授因编号 11104 案只存“案由”，故当年未于统计中计入。若将此只存“案由”之编号 11104 案计入，则行政门应有 575 案，总计也将成为 1164 案。

(4) 吴祖善：《清代“淡新档案”：台大图书馆的新特藏》，《“中国图书馆学会”会报》1987 年第 40 期。

a) 淡新档案的公开

自从一九五三年戴炎辉教授撰文介绍之后，淡新档案作为一种台湾史甚至中国史研究之重要材料，已受到学术界的瞩目。尤其是戴教授此后随即利用淡新档案陆续进行了一系列有关清代台湾隘制、屯制及地方乡庄组织的研究^②，淡

① D.C. Buxbaum, *Some Aspects of Civil Procedure and Practice at the Trial Level in Tanshui and Hsinchu from 1789 to 1895*, *The Journal of Asian Studies*, vol. 30, no.2 (1971).

② 戴炎辉教授利用淡新档案的这一系列研究成果，后来结集出版戴炎辉：《清代台湾乡庄》，台北：联经出版公司，1979 年。

新档案之公开便是学术界所引颈期待的事了。

一九五五年,戴炎辉教授挑选淡新档案行政编民政类厚俗款的10种32件官府示禁告示在《台北文物》第四卷第三期用活字排印刊出^①,而且以照片的形式披露2张淡新档案的原件,这是外界首次得见淡新档案的样貌。虽然在此次刊出32件淡新档案内容时,戴教授于文中《例言》曾谓:“拟陆续将淡新档案资料公诸学界,供学术研究之参考”,而且刊出时也将题名拟作《淡新档案资料(一)》,但不知何故,却不见如《例言》所说的“陆续公诸学界”,因此其《淡新档案资料》(二)(三)……也便未曾出现了。此时期,外界只能从戴教授的诸多研究论文所征引的片段,略窥淡新档案的堂奥。

一九七一年,戴炎辉教授选录行政编的475件档案,交由台湾银行经济研究室编印成为《台湾文献丛刊》第295种《淡新档案选录·行政编初集》共四册594页。这是淡新档案比较有规模的出版。这使淡新档案可以被广泛地阅读,对长期企盼希望利用淡新档案的人来说,真是令人兴奋的事。

但是,或许是因为迁就《台湾文献丛刊》之体例或编辑方针(《台湾文献丛刊》对于较大部帙之书籍、史料,均以“选录”“选辑”之方式编辑出版),此次淡新档案以活字印刷出版时,并未如戴炎辉教授先前所强调的应该重视其以“案”为单位的重要性,而是就不同公文类型各选一、二“件”微例编印出版。因此,这些印刷出版的淡新档案的研究价值大大地打了折扣。因为,研究者将只能读取每“件”档案的信息,而无法研究一个具有丰富、完整信息,甚至一个有首有尾的个案。这种出版档案的方式,如果是在诉讼个案上就更会是严重的缺陷。因为这将会无法分析该诉讼案件的整体内容和诉讼过程中的具体攻防。

另外,这次的出版也具体地考验我们:在将清代档案以活字印刷时,到底可以做到如何程度的“如实重现”?在这种将档案原件印刷出版时,可以容许档案原件上之信息流失到什么程度?将原件排字出版的工作,不但费力、费时,而且也有很多极难克服的技术问题。例如,档案原件中的各种图形标记,甚至印章戳记便很难被忠实地原样印刷出来,因此经常必须以其他的形式呈现或完

① 戴炎辉、施绮云:《淡新档案资料(一)》,《台北文物》1953年第4卷第3期。

全舍弃；而且，淡新档案有甚多是淡水厅、新竹县衙门内的公文拟稿，同一文件经常经过多人或多次修改增删，这些修改、增删提供了宝贵的政治过程信息，但如何得以被忠实地在印刷过程中表现出来，却是很大的难题。为了应付这些难题，戴炎辉教授订立了繁复的《淡新档案选录体例》（共20条），企图在保存原件原貌与迁就活字印刷技术之间，求取适度的平衡，但也终不免牺牲了档案原件的诸多信息。其实，这个繁复的《淡新档案选录体例》已充分说明了以活字印刷清代官府源文件所面临的困局。后来，这个企图将淡新档案以活字印刷出版的计划，在只出版了4册（475件）《淡新档案选录·行政编初集》之后就没有再继续，其确实原因虽不得而知，但编辑出版印刷上难以克服的技术困难应该是可以想象的原因。

淡新档案的活字出版计划挫折之后，美国华盛顿大学（西雅图）以拍摄微卷的方式复制档案，无疑是突破上述“原样呈现”难题的好方法。但是，利用华大拍摄之微卷研究诉讼手续的日本东京大学滋贺秀三教授，对此微卷副本却还是认为不够理想：“大凡史料文献之原件与复本之间均有差异，而此淡新档案之微卷与其原件之落差，则属于较大的部类。原件中墨字与朱笔之交错，在黑白单调的写真中，便只变成些微的浓淡差异而已，而多难以判读。墨字之上按捺朱印之处，若是原件还可判读，但微卷则无法判读。文书上之附笺，时有未翻开拍摄者，亦有难以注意者。也偶有误摄而无法阅读者。另外，各影幅之拍摄比例尺不尽相同，这也必须留意。有几处缺号，但却有原件，不知到底是因何漏拍，抑或本无原件。”^①在此，我们感受到严谨研究者要求“原样呈现”档案原件的极致。

b) 一九九〇年代以后的整理

一九八六年，淡新档案入藏台湾大学图书馆后，台湾大学图书馆除了对档案的数量再次进行清点，和对一些破损的档案加以修补之外，并没有对这批档案进行整理。一九九〇年起，才在一项包含几种台湾史档案的目录整理

^① [日] 滋贺秀三：《淡新档案初步的知识：诉讼案件文书类型》，《岛田正郎博士颂寿纪念论集·东洋法史探究》，东京：汲古书院，1987年，第254页。

计划中,在原来戴炎辉教授所制作的目录之基础上,逐“件”地编制淡新档案目录。

此次编制目录时,我们从图书馆的目录卡获得启发,设计了一份包含有12个字段的《十二字段表》^①,企图想要以详细的 metadata 来描述这批档案。所谓 Metadata (中文翻译或为“诠释数据”“后设数据”“数据描述格”或是“元数据”),是叙述资源属性的数据,又称为“Data about data”(关于数据的数据)或是“Data describes other data”(描述其他数据的数据)、“Additional information that is necessary for data to be useful”(用来使资料成为有用的附加性数据)、“information about the data that helps in optimization and management of that data”(用来帮助使用、管理数据的信息)等广狭不一的定义。传统图书馆的目录,就是一种常见的 metadata,我们透过它所提供的信息(最常见的是“作者”“书名”“出版者”“出版年”“分类号”等),来从数量庞大的图书馆中找出我们所要的一本书,甚至利用它来找到相关的一些书。相同的道理,淡新档案也应该有这样的“诠释数据”,以便来管理、检索它。

但是,每一“件”淡新档案应该有哪些“诠释数据”呢?这却是见仁见智的问题,它将因为研究者的期待与需要之不同而有所不同。一般而言,档案的使用者(研究者)总是希望将档案原件里的信息尽量地表现在“诠释数据”里,而要求设计数量众多的“字段”。希望将档案原件里的信息尽可能地透过诠释数据域位表现出来,必然要多设字段数量或者增加字段的著录信息量。当然,这样也就连带着必然增加整理工作(制作诠释资料)的工作量。如果这样的要求无限扩大,则会使诠释数据的分量逼近档案原件的全文数量。档案使用者(研究者)希望以多数字段的《十二字段表》,配合详细的字段著录来增加信息量 and 使用的方便性,但它却必须花费相应的工作成本,这是整理档案时不能不考虑的。

① 所谓“字段”,就是图书馆学中指称目录卡中的各类信息,英文为 field,据说也有中文出为“字段”。《十二字段表》的12个字段,分别为:01.编号,02.名称,03.文书类别,04.时间,05.收发关系,06.关系人物,07.关系地区,08.关系地名,09.相关项目,10.典藏编号,11.附注,12.内容摘要。

另外，档案目录的制作工作看似简单，但真正执行开来却一定会遭遇到各种困难。即使决定了诠释数据应该要有的字段，但这也还只是决定了初期问题。接下来的是，如何实际注录诠释数据域位。图书馆用来管理、检索图书的目录，不论如何地简单也都会有“作者”“题名”“年代”三个字段。但即使只是这三个字段，用来整理像淡新档案这种州县档案时，却也不是那么地顺理成章、毫无困难。例如，一份新竹县知县将要送出去给台北府知府的禀文稿，应该如何注录呢？如果将“作者”当成该件禀文稿的制作者，那么是否应该将它注录成撰写该禀文稿的幕友或胥吏呢？或是应该注录最后送出禀文正本时候的名义人，即新竹县知县呢？至于“题名”，它并未见诸该禀文稿，档案原件应该只是在文件之最前头注明文件之类型“禀”接着是该禀文之“事由”，那么是否应该注录成“禀文稿”“禀文”，或其他（例如，将“事由”也当成是“题名”的一部分；甚或如我们在制作《十二字段表》时所采取的原则：“题名”乃是结合“发文者”“收文者”“事由”“文类”的总和）呢？至于年代，到底是这份禀文稿的拟稿日期，还是知县的批行日期，甚至是该禀文稿被清缮成正式公文用印后发出的日期呢？

也就是说，即使诠释数据的字段决定了，各字段应该如何著录也还是必须作出政策性的决定。诠释数据之字段应该如何著录，并不是不辩自明的。诠释数据之所以又称“后设数据”（其实，其英文原文 metadata 的直译反而应该是“后设数据”），乃是因为这些诠释资料多具有“后设”的性质。也就是说，它并不是原本就在那里的，而是后来的整理者、解读者、诠释者所给予的。那么，它也会因为整理者、解读者、诠释者的不同，或不同的脉络、目的而会有所不同。这样的不同，经常不是对错的问题，而是选择的问题。

因此，上述所指出的，诠释数据应该有什么样的字段、诠释数据域位应该如何注录，也便必须作出选择，即“政策性的决定”。它指的是研究者的目的（或研究主题），将会是决定应该有什么样的诠释数据域位和如何注录这些字段。当然，实际从事整理作业所需要的成本（人力、时间、经费等），也是决定如何选择的重要因素。

一九九〇年前期的这个以著录《十二字段表》为重心的淡新档案整理工作，最后几乎可以说是“白忙了一场”。一方面，将 19000 余件档案的每一件都各

填出一份《十二字段表》本身就是一项大工程，况且具体著录字段时多存在着难以简单地下笔的困境，甚至多有将字段空白下来的情况，而且这样庞大的工作势必由多人分头进行，如何让不同的字段著录者的著录达成一致，也是个难题。而且，由于这个工作结束之后不久，另一项将淡新档案全面性地打字、标点出版的计划已经展开了，因此这个《十二字段表》也就没有多大意义而被束诸高阁了^①。

一九九五年起，淡新档案的整理进入了新阶段。那就是将所有的档案以打字、标点的方式出版。这是一项工作量更大，困难度更高的计划。但是经过长达15年的持续努力，这项计划在2010年完成，总共出版《淡新档案》36册。

近百年来以规格化的字体、版式来出版原本非规格化的文献（例如，甲骨文、简牍、敦煌文书），研究者已经累积有相当经验，而且发展出来诸多方法（例如，以各种符号表示各种文献中的信息；以不同的字体、字的大小来表示不同性质的文句；以注释字的方式说明无法呈现于版面的信息……）。但是因为文献原件之歧异几乎有无限的可能，因此绝对无法全部用已经规格化了的字体、版式来呈现。淡新档案这种清代手写档案以现代之打字、排版方式出版时，最大的困难是如何保留档案原件并不均质，甚至格式与内容都变化无穷的信息。20世纪70年代《淡新档案选录·行政编初集》的出版计划无法持续的经验，已经说明了问题的困难性。

档案原件包含着该档案的物理性成分（例如，载体材料体，高、广、厚，格式等）之外，还有内容信息（文字与非文字）。档案原件中的非文字性信息，例如，图表、印章、画押等，固然多不是标准化的，也难以印刷出版；即使文字性的信息，也不见得是标准、一致性的（例如，有一些是已经印刷的格式性文字，但另外一些则是手写的文字；这些手写的文字，也经常会有字体大小、书写风格差异、俗字、异体字、错字等等各种几乎无法完全列举的变异）。这些档案原件中具有差异性质的信息，对于特定的研究者来说或许具

^① 目前经过打字、标点而以纸本出版的《淡新档案》，尚保留《十二字段表》中著录得相当长的每一件档案的“题名”。

有非常重要的意义。例如，淡新档案的诉讼案件中必然出现的“呈”（诉状）的档案原件，至少存在着几类不同性质的文字信息：（1）它是一张依官府规定印制的格式纸。其中已有一些文字信息。（2）提出诉讼者在印制的格式纸上所写的呈词。一般来说，这是研究者最关心的信息。（3）衙门接受呈状之后，衙役的登录注记。（4）主理官的批示。（5）事后衙吏在文书整理过程中所附加的注记。淡新档案中的“票”（主官签发衙役办事的命令书），也经常包括下列几种不同类型的信息：（1）幕友或书吏的拟稿。（2）主官在拟稿上的增删修改及批行。（3）销缴后的整理注记。这些文件中不同性质的信息形成了这个文件的“生命历程”。如果吾人直接阅览史料原件，便会觉察这些不同性质的信息，是以字体书风、墨色浓淡、格式、位置、大小等方式呈现的。这些可据以读取不同性质之信息、重建史料“生命历程”的差异，无疑会是研究者所珍惜的。但是在改用标准化的活字、版式编辑印刷时，这些在原件上本来并不均质的文字却多数要迁就标准化而不得而被割舍，甚至也无法完全如实呈现。即使在将档案原件以活字出版时，编辑者经常都会借助诸多记号、文字大小、不同字体，甚至批注来企图尽量将档案原件中的各种差异标示出来，但还是远远不足以将几乎无法穷举的档案原件中之差异完全呈现出来。

因此，档案原件的整理、出版工作之每一个环节，都必须作出妥协性的决策。例如，诉讼案件中的呈状在打字、印刷时，只挾取实质的呈词内容，而放弃事先印刷好的诉状纸里的信息；对于文件中的印章戳记，也只能选择部分加以呈现或批注，而无法全数表示；字体大小、抬头、格式，也多只能呈现其中一部分。也就是说，研究者所期待的“原样呈现”的理想，即使投入再多的努力，也势必无法达成而要被打折扣。因此，必须在顾虑研究者的期待之下决定哪些必须坚持，哪些必须妥协。如今，淡新档案已经全数打字、标点，以纸本的方式出版发行了。这套 36 册的《淡新档案》，就是在一连串的妥协之下所整理出来的成果。

c) 淡新档案的数字化整理及其展望

一九九五年淡新档案开始进行全文打字、标点出版纸本之后不久，另一个

数字化整理的工作，也在一九九七年展开了。

淡新档案的数字化整理，纾解了打字、标点出版无法达成“原样呈现”之缺憾。数字化整理所得的影像，几乎可以“原样呈现”档案原件的平面性信息（当然，目前的技术还是无法呈现材质、厚度、重量、气味、触感等），可以响应研究者“原样呈现”档案原件的期待^①。因此，打字、标点出版的《淡新档案》搭配档案原件的数字影像文件，不就让淡新档案完全公诸于世了吗？

淡新档案数字化整理，除了将档案原件进行数字化扫描之外，我们同时也仍然使用 metadata 来管理这些数字影像文件，并且将打字、标点后要出版为纸本的全文文件来形成全文数据库。因此，我们的淡新档案整理此时可以说是全面性地展开了：既有档案原件的数字影像，还有每件档案的 metadata，又有全文文字文件。于是，一个可以多方面借助数字技术来进行的淡新档案研究之基础建设已经大致完成。

不过，此时我们还是有一项尝试错误的经历。淡新档案之数字化整理的同时，我们也在进行其他台湾历史学、人类学、植物学、动物学资料的数字化，因此我们曾经企图发展出一个可以同时整合性地管理、检索这些各种不同学科性质的台湾资料之 metadata 字段表。这的确是个具有高度理想性的企图，但是经过一番尝试之后，我们不得不宣告放弃。因为每一种学科数据之性质不同，每一个学科之研究者使用数据的期待也不同，因此彼此之间很难达成一致的看法。所以，metadata 字段如何设计，除了有上述求其完备的要求之外，也有可能被要求跨越多种数据，达成整合管理数据的可能性。这种时候，勉强可以达成共识的字段，似乎就会只是最基本的几个字段（例如，作者、题名、时间、地点）而已。其实，即使只是这几个字段，不同的学科对于该字段如何定义、应该如何著录，也存在着不同的侧重或主张。例如，植物学者对于植物标本的“作者”这一个字段，一部分研究者可能希望著录该植物标本的采集者或制作者，也有一些研究者则主张应该著录该植物的命名者。由

① 1997 年数字化整理计划展开之初，因储存容量与传输瓶颈的原因，虽然决定数字化扫描规格为 300dpi，但是扫描成黑白影像文件（jpg 文件）。但以后随着储存容量、传输瓶颈的解放，2005 年以后，又重新以彩色扫描。

于，各学科之间的差异过大，无法简单地达成共识，因此我们放弃了想要发展出一套可以整合性的管理多学科资料之 metadata 字段的理想。其实，自从开始整理淡新档案以来，可以说就是一连串妥协的连续。这些妥协，或许来自于上面已经提到的成本考虑，也可能来自于技术上的限制，但同时也是研究者一开始总是习惯性地将特殊化的要求求诸于具有普遍性性质的资料整理工作所致。所以，数据的整理工作必须认清整理的目的，只作最基本的整理（尤其在数量庞大时，更应该如此）。如果在此基础之上，还有特殊化的要求时，才另作特殊化的处理。

淡新档案的数字化整理工作，在完成（1）每一件的数字影像、（2）每一件的 metadata、（3）每一件的全文数字文件之后，可以说已经有了得以利用信息技术来协助研究的基础设施^①。因此，目前台湾大学正在这样的基础之上，结合历史学者和资讯工程学者，一方面建设可供研究者使用的数字环境和工具，一方面透过科技的协助做出传统方法难以做出的研究。以下稍稍介绍目前台大研究团队的一些初步研究。

数据库的最基本功用之一是提供检索。因此，提升检索功能，并使检索方法更加方便，检索结果更加周遍、精准，就是努力的方向。一般来说，数据库的检索，经常是由使用者键入检索词，透过计算机在数据库中将该词找寻出来，但是检索者要能在计算机中键入一个检索词，显然这个检索词便要是检索者“已知的”。那么，是否可以让计算机不只是等待检索者键入一个已知的检索词，而是可以更积极地由计算机提示一些具有意义而却是检索者所未知的词汇呢？我们先由计算机自动从数据库中抽取关键类词，然后再由学科专家筛选确定而“发现”了不少新词汇。与此相关的是，为了使检索的机能更加精确有效率，也必须形成各种辞典（例如同音异字词典、同义辞典等）作为辅助检索工具。这些辞典也可以由计算机协助学科专家来完成。除此之外，我们也开发了以地理信息系统（GIS）的方法所支持的，靠着地图来检索数据库的办法。这种检索方法，可以不受地名变迁的影响，透过相同的地理坐标来检索出不同时代相同地点的数据。

^① 如果说还有什么必须做的基础建设，那就是以“案”为单位的 metadata 建置。

附录:

David C. Buxbaum 统计的淡新档案逐年案件数^①

每年案件数:

A: 占本表中的案件数百分比

B: 占每年的案件数百分比

年	行 政			民 事			刑 事			
	百分比			百分比			百分比			
	案数	A	B	案数	A	B	案数	A	B	总计
1789	1	0.18	100							1
.....										
1805	1	0.18	50	1	0.46	50				2
.....										
1812							1	0.28	100	1
1813										
1814										
1815	1	0.18	100							1
1816										
1817										
1818										
1819										
1820	1	0.18	100							1
1821										
1822										
1823										
1824										
1825										

① D. C. Buxbaum, *Some Aspects of Civil Procedure and Practice at the Trial Level in Tanshui and Hsinchu from 1789 to 1895*, *The Journal of Asian Studies*, vol. 30, no. 2 (1971), pp.265-266.

续表

年	行 政			民 事			刑 事			
	百分比			百分比			百分比			
	案数	A	B	案数	A	B	案数	A	B	总计
1826										
1827										
1828	1	0.18	100							1
1829										
1830										
1831										
1832										
1833										
1834	1	0.18	100							1
1835	1	0.18	100							1
1836	3	0.54	100							3
1837	1	0.18	100							1
1838	2	0.36	100							2
1839	1	0.18	100							1
1840	3	0.54	100							3
1841										
1842	3	0.54	100							3
1843	2	0.36	100							2
1844										
1845				1	0.46	100				1
1846										
1847	3	0.54	100							3
1848	1	0.18	100							1
1849	2	0.36	66.7				1	0.28	33.3	3
1850	2	0.36	40	2	0.92	40	1	0.28	20	5
1851	3	0.54	75	1	0.46	25				4
1852	4	0.72	40	3	1.38	30	3	0.83	30	10

续表

年	行 政			民 事			刑 事			
	百分比			百分比			百分比			
	案数	A	B	案数	A	B	案数	A	B	总计
1853	2	0.36	50				2	0.55	50	4
1854	2	0.36	100							2
1855	4	0.72	66.7				2	0.55	33.3	6
1856	3	0.54	50	2	0.92	33.3	1	0.28	16.7	6
1857	15	2.70	68.2	2	0.92	9.1	5	1.38	22.7	22
1858	2	0.36	33.3	2	0.92	33.3	2	0.55	33.4	6
1859	1	0.18	33.3				2	0.55	66.7	3
1860	1	0.18	33.3	2	0.92	66.7				3
1861	10	1.80	100							10
1862	5	0.90	71.4				2	0.55	28.6	7
1863	3	0.54	75				1	0.28	25	4
1864	3	0.54	75				1	0.28	25	4
1865	3	0.54	37.5	2	0.92	25	3	0.83	37.5	8
1866	9	1.62	90				1	0.28	10	10
1867	4	0.72	80				1	0.28	20	5
1868	7	1.26	63.6	3	1.38	27.3	1	0.28	9.1	11
1869				1	0.46	25	3	0.83	75	4
1870	6	1.08	50	3	1.38	25	3	0.83	25	12
1871	5	0.90	35.7	1	0.46	7.2	8	2.22	57.1	14
1872	6	1.08	35.3	4	1.83	23.5	7	1.94	41.2	17
1873	9	1.62	52.9	5	2.29	29.4	3	0.83	17.7	17
1874	14	2.52	45.2	5	2.29	16.1	12	3.32	38.7	31
1875	2	0.36	11.1	8	3.67	44.4	8	2.22	44.5	18
1876	9	1.62	36	6	2.75	24	10	2.77	40	25
1877	2	0.36	13.3	8	3.67	53.4	5	1.38	33.3	15
1878	14	2.52	38.9	15	6.88	41.7	7	1.94	19.4	36
1879	34	6.14	52.6	10	4.58	17.5	13	3.60	22.9	57

续表

年	行 政			民 事			刑 事			
	百分比			百分比			百分比			
	案数	A	B	案数	A	B	案数	A	B	总计
1880	12	2.16	48	6	2.75	24	7	1.94	28	25
1881	17	3.07	40.5	9	4.13	21.4	16	4.43	38.1	42
1882	31	5.58	42.5	17	7.80	23.3	25	6.93	34.2	73
1883	38	6.85	61.3	5	2.29	8.1	19	5.26	30.6	62
1884	19	3.43	46.3	8	3.67	19.5	14	3.88	34.2	41
1885	20	3.61	36.4	7	3.21	12.7	28	7.75	50.9	55
1886	47	8.47	51.1	14	6.41	15.2	31	8.59	33.7	92
1887	55	9.91	43.3	24	11.01	18.9	48	13.30	37.8	127
1888	33	5.94	61.1	13	5.96	24.1	8	2.22	14.8	54
1889	20	3.61	54.1	9	4.13	24.3	8	2.22	21.6	37
1890	18	3.25	62.1	5	2.29	17.2	6	1.65	20.7	29
1891	13	2.34	44.8	7	3.21	24.1	9	2.50	31.1	29
1892	7	1.26	31.8	2	0.92	9.1	13	3.60	59.1	22
1893	1	0.18	5.5	3	1.38	16.7	14	3.88	77.8	18
1894	8	1.44	50	2	0.92	12.5	6	1.65	37.5	16
1895	4	0.72	100							4
总计	555	100	48.9	218	100	19.2	361	100	31.9	1134

(原载《中国社会经济史研究》2017年第2期)

图书在版编目 (CIP) 数据

地方档案与文献研究. 第三辑 / 吴佩林主编. -- 北京: 国家图书馆出版社, 2017.11

ISBN 978-7-5013-6319-3

I. ①地… II. ①吴… III. ①地方档案—研究—中国 ②地方文献—研究—中国 IV. ① G279.27 ② G256

中国版本图书馆 CIP 数据核字 (2017) 第 299669 号

书 名 地方档案与文献研究 (第三辑)

著 者 吴佩林 主编

申 斌 本辑特邀主编

责任编辑 王亚宏

封面设计 程 言

出 版 国家图书馆出版社 (100034 北京市西城区文津街 7 号)

(原书目文献出版社 北京图书馆出版社)

发 行 010-66114536 66126153 66151313 66175620

66121706 (传真) 66126156 (门市部)

E-mail nlcpress@nlc.cn (邮购)

Website www.nlcpress.com → 投稿中心

经 销 新华书店

印 装 北京市通州兴龙印刷厂

版 次 2017 年 11 月第 1 版 2017 年 11 月第 1 次印刷

开 本 710×1000 (毫米) 1/16

印 张 22.5

字 数 370 千字

书 号 ISBN 978-7-5013-6319-3

定 价 58.00 元