

# 數據生態視域下的人文社會科學 撤稿問題分析\*

夏義堃

[提 要] 從數據生態視角考察人文社會科學撤稿問題,能夠全面反映撤稿背後的數據管理問題、學術誠信問題與評價問題。人文社會科學撤稿問題在撤稿原因以及學科、國家、期刊等要素分布上有其自身特點。當前,因數據問題的撤稿不斷增加,且類型複雜、界限模糊。究其原因,涉及到數據觀念局限、數據制度缺失、數據監管不力、數據基礎設施薄弱等問題。隨着數據驅動型研究範式的發展,應高度重視人文社會科學撤稿中的數據問題,把握數據管理、學術誠信管理以及學術評價之間的內在關係,以實現多主體數據管理制度的橫向貫通與撤稿管理的規範化。

[關鍵詞] 數據生態 撤稿 人文社會科學 學術誠信 學術評價

[中圖分類號] C01;G203 [文獻標識碼] A [文章編號] 0874-1824(2021)04-0120-13

撤稿一直被學術界視為糾正偏差的有效手段,是洞察學術不端、倫理失當等行為的有效工具。引文索引的創始人 Eugene Garfield 指出:撤稿是學術出版過程中必不可少的組成部分,學術界有義務監督其出版物。<sup>①</sup>首個撤稿記錄可追溯至 1756 年 6 月,英國皇家學會《哲學學報》上刊登的撤稿聲明。<sup>②</sup>與自然科學領域頻頻發生的撤稿事件相比,人文社會科學撤稿有其自身特點,站在數據生態視角,可以更為直觀、全面的揭示撤稿背後的驅動要素、發生機理與癥結誘因,進而找出解決問題的有效方法與對策。

## 一、研究背景

### (一)人文社會科學撤稿現象研究進展

儘管撤稿問題的高發學科集中在生命科學等自然科學領域,但人文社會科學的撤稿現象依然持續發生。1960 年《芝加哥大學法律評論》對五年前發表的一篇文章實施了撤稿;2012 年,美國本特利大學會計學教授 James E. Hunton 被迫辭職,其發表在《國際會計信息系統雜誌》等多家期刊的三十多篇文章因數據偽造問題而撤稿;<sup>③</sup>無獨有偶,內華達大學前英國文學教授 Mustapha Marrouch

\* 本文係國家自然科學基金創新群體項目“信息資源管理”(批准號:71921002)的階段性成果。

因剽竊他人成果,23篇文章被撤稿,並被解僱。<sup>④</sup>而荷蘭蒂爾堡大學心理學家 Diederik Stapel 因數據造假導致 58 篇文章被撤稿,一度引發了心理學的信任危機,涉及可複製性危機、理論危機、統計危機、抽樣危機等多方面問題。

從撤稿文章所涉及的學科來看,幾乎覆蓋人文社會科學領域的所有學科,其中,心理學、經濟學、管理學、社會學等社會科學撤稿數量較高。截至 2020 年 5 月底,撤稿觀察數據庫 Retraction Watch Database 撤稿人文社會科學論文數總計 951 篇,其中,商業經濟 372 篇,佔比 39.1%,社會科學 467 篇,佔比 49.1%,人文社科 112 篇,佔比 11.8%。<sup>⑤</sup>一項針對歐洲經濟學家的大型問卷調查顯示,“32% 的人報告曾經有選擇地提出實證結果以證實個人研究論點,3% 承認偽造了一些數據,這些數據表明,經濟學中現有的撤稿水平低估了舞弊的實際水平。”<sup>⑥</sup>當然,較高的撤稿率也反映了期刊、學界對學術不端等不當行為進行處理的強烈意願,並不只是論文檢測能力的問題。

在撤稿原因上,有學者根據 Retraction Watch Database 心理學撤稿聲明的調查,指出數據捏造/偽造/欺詐在心理學(48%)中比在商業和管理(33%)和經濟學(0%)中更為突出;<sup>⑦</sup>也有人通過對工商管理學科撤稿論文的追蹤,發現主要原因依次為數據欺詐、自我剽竊和剽竊、數據分析錯誤、作者身份爭議、不恰當的引用操縱等。<sup>⑧</sup>對此,Ribeiro 和 Vasconcelos 指出,雖然人文社會科學的撤稿數量最少,但在撤稿原因中,數據造假等學術不端佔比最高(58%)。<sup>⑨</sup>

在撤稿現象的深度分析中,普遍認為撤稿反映了學術研究的扭曲,既是市場功利性行為的滲透,也是學術出版過度求新的短視,並對學者聲譽造成惡劣影響。多位學者圍繞撤稿對撤稿人學術聲譽、職業發展的影響展開調查,發現撤稿對知名學者造成的負面影響遠大於普通研究人員,<sup>⑩</sup>第一作者的學術聲譽會顯著下降,加之媒體報道,會客觀上加重對撤稿人的處罰。同時,撤稿不會對資助撤稿學者的機構聲譽產生負面影響。<sup>⑪</sup>

對於撤稿管理,部分學者基於撤稿聲明、期刊撤稿流程、撤稿中的責任爭議以及撤稿後仍被引用等角度展開分析。魏眾、蔣穎聚焦中國人文社會科學的撤稿現象,指出存在着撤稿主體多樣、程序不明確、撤稿聲明缺乏統一規範以及撤稿後相應獲利行為終止與追溯困難等現實問題。<sup>⑫</sup>

## (二)數據生態與人文社會科學撤稿的內在關係

數據生態是相關主體在數據資源採集、加工、存儲、利用等環節,相互聯繫與互動交流所形成的數據關係和數據環境。直觀判斷,數據生態與人文社會科學撤稿之間並無緊密關聯,但數據生態由數據主體、數據節點、數據交換制度、行為和數據流構成,科學研究中的數據生態反映了人文社會科學數據在研究、出版與傳播等不同節點的流動與轉化,是學術生態的數據化展現,可清晰揭示撤稿論文數據問題的發生與演進過程,折射出撤稿人、期刊、高校等主體的學術動機、學術誠信管理與數據行為選擇等群象面貌。具體而言,二者間的內在關聯主要體現在以下方面:

(1)數據生態反映了撤稿現象的演進與發展。數據的記錄、描述、分析、共享、發布、存儲,不僅完整展示了科研過程中數據及成果醞釀、初創到完善的過程,也全面呈現了撤稿論文數據抽取加工直至形成的創作脈絡,為問題溯源及處理判斷提供了原始依據;因而,人文社會科學研究數據生命周期管理制度從無到有的確立、數據存儲共享平台從零到一的建構以及數據文化從封閉到開放的轉變,既代表了人文社會科學研究規範化、科學化水平的提升,也預示了學術偏差發現、糾正能力的提升。站在個體維度,數據生態也刻畫了撤稿人的學術認知及其數據行為選擇標準。據調查,意大利心理學家協會 277 名成員中有 88% 承認曾經參與過至少一次可懷疑的研究實踐,2.3% 的成員承認偽造過數據。<sup>⑬</sup>在現實中,數據獨佔的觀念窠臼、利益驅動的功利選擇都有可能改變研究人員的

治學態度進而引發數據造假等學術投機行為。

(2)數據生態揭示了撤稿現象的癥結與原因。以研究人員為核心的主體數據認知與數據行為既是構成數據可信度與科研完整性的關鍵性要素,也是考察撤稿問題的重要切入點。良好的數據生態既關注人文社會科學研究環境的變化,關注大數據環境下研究範式、方法和工具的重新定義與整體推進,也重視微觀環境的審視,強調以高質量數據驅動學術生態建設,並將孤立的撤稿問題與人文社會科學學術生態以及學術出版、學術服務與支持等聯結在一起,以從中探尋誘發撤稿問題的主客觀因素,以求構建自律、規範的人文社會科學學術生態合力。

(3)數據生態詮釋了撤稿現象的解決方案與策略選擇。數字人文的興起、數據驅動型研究範式的轉變、開放數據與開放科學進程的普及,使得人文社會科學數據生態與學術誠信、學術創新、學術出版乃至學術評價等學術生態體系相關要素之間,形成了互為支撐的複雜鏈式依存關係。因此,數據問題不僅反映了撤稿人的思維局限,還通過學術誠信、評價、出版等要素之間的交叉作用影響,揭示出數據偽造、數據誤用等現象背後的深層次利益關係,並通過開放學術、數據共享以及評價制度改革等找出標本兼治的有效方法,以實現對人文社會科學學術信任機制的重構。

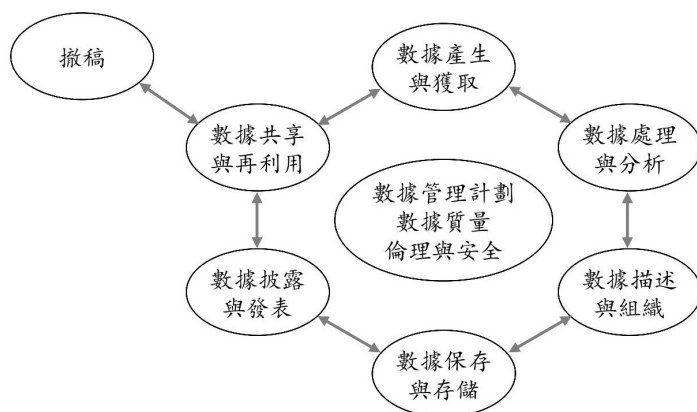


圖 1 數據生態與人文社會科學撤稿內在關係

當然,撤稿現象同時也暴露出人文社會科學數據管理的薄弱與不足。國內外多項研究表明,人文社會科學研究人員數據共享意願低、機構數據管護能力差、期刊數據存繳要求少,公共數據存儲開放平台弱等,不僅造成了人文社會科學數據資源彙聚融合、共享利用的現實瓶頸,客觀上也為數據造假、數據誤用等問題的撤稿打開了便利之門。由此可見,數據生態與人文社會科學撤稿現象之間存在着互為依存互為促進關係(見圖 1),其本質是激活撤稿主體的數據意識,強化研究過程中的數據共享,並通過負責任的數據管理來增強學術誠信,平衡對資助機構、高校/科研機構、研究人員、期刊編輯和其他參與者的學術倫理和數據義務,以消除或降低撤稿現象的發生。

## 二、人文社會科學撤稿現象概況及數據問題分析

### (一)數據來源

2010 年創建的 Retraction Watch 網站收錄了 1990 年以來的國內外撤稿文獻,可通過標題、學科、期刊、出版商、機構、撤稿原因、發表時間、撤稿時間、國家等字段對撤稿論文進行檢索,現已成為研究學術不端和學術出版問題的主要數據來源。本文聚焦 Retraction Watch 網站人文社會科學領域

撤稿論文,截至2021年6月10日,去除重複、學科關聯不高、非學術論文等干擾項,共採集到2,907篇相關撤稿論文。

## (二)基本概況

### (1)撤稿原因分布

Retraction Watch 網站共羅列出 100 條撤稿原因,除卻通知不全、通知用語模糊等原因不明確的記錄,主要撤稿原因可歸納為剽竊或重複發表、數據、作者、方法與結果、期刊/出版商、作者機構、第三方、同行評審、版權/法律糾紛、違背學術倫理共 10 個類別。由於一篇論文可能涉及多條撤稿原因,被撤稿的 2,907 篇學術論文共涉及 5,514 條撤稿原因記錄。其中,首要撤稿原因是剽竊或重複發表,佔撤稿記錄的 36.87%;其次是因圖像和數據錯誤、複製、偽造、不可信、缺失、操縱等數據問題導致的論文撤稿 676 篇,共對應了 827 條撤稿原因記錄,佔比 15.00%;再次為作者原因引發的撤稿,如違反期刊/出版商的規定或偽造作者、未取得原作者同意等,佔比 9.70%。此外,因研究方法或實驗設計不科學、不完整而導致的結果不可信,也即可懷疑的研究實踐,與數據問題關聯緊密,佔比達到 9.18% (參見表 1)。

表 1 人文社會科學撤稿論文原因歸類統計

序號	原因歸類	撤稿記錄(條)	佔比
1	剽竊或重複發表	2,033	36.87%
2	數據	827	15.00%
3	作者	535	9.70%
4	方法與結果	506	9.18%
5	期刊/出版商	497	9.01%
6	作者機構	357	6.47%
7	第三方	355	6.44%
8	同行評審	257	4.66%
9	版權/法律糾紛	115	2.08%
10	違背學術倫理	32	0.58%
合計		5,514	100%

### (2)時間分布

總體而言,人文社會科學撤稿論文數量呈現上升且加速增長的趨勢,具有較為明顯的階段式增長特點。第一階段(1982~2008年),撤稿論文數量較少,每年基本維持在個位數;第二階段(2010~2016年),撤稿論文數量緩慢增加,逐漸從兩位數增長到三位數;第三階段(2017~2020年),撤稿論文快速增加,2019年相較於2018年增幅高達72.13%(參見圖2)。自2012年起,因數據問題引發的撤稿開始呈現出緩慢增長的波動趨勢,2020年達到高峰,上升到149篇,佔該年度撤稿論文的30.0%。從發展趨勢來看,數據問題導致的撤稿演進曲線與總體撤稿演進曲線表現出較高的一致性。

### (3)時滯分布

從數據分析中發現,數據問題相較於剽竊或重複發表等具有更強的隱蔽性和複雜性,往往需要更長的時間才能被發現。人文社會科學總體撤稿時滯為 0~77 年,遠高於自然科學,其中 88% 的撤稿發生在 5 年內,平均值為 3.5 年;與數據問題相關的撤稿時滯 0~77 年,71% 的撤稿發生在 5 年內,撤稿時滯中位數為 2 年,平均值為 8.12 年。

#### (4) 學科分布

按照 Retraction Watch 網站的大類學科劃分,撤稿論文主要集中在商業和技術相關學科(53.1%),其次為社會科學(39.5%),人文學科數量最低,只佔撤稿總量的 7.4%。考慮到國內外學科劃分的差異,參考《普通高等學校本科專業目錄(2020 年版)》,本文將哲學、經濟學、法學、教育學、文學、歷史學、管理學和藝術 8 個學科門類的相關專業劃入“人文社會科學”學科範疇,並將 Retraction Watch 網站的學科領域手工對應到相應學科門類。

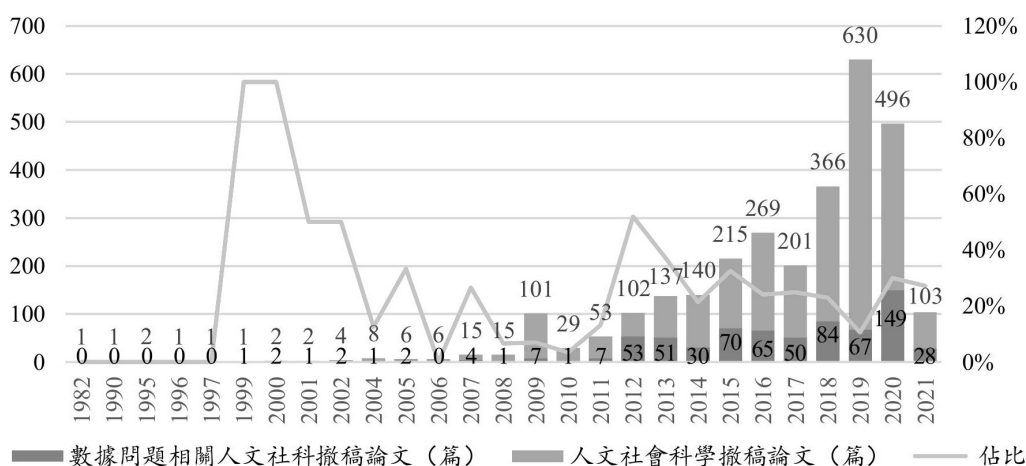


圖 2 人文社會科學總體撤稿論文及數據問題相關撤稿論文的时间分布

通過比較發現,撤稿論文的學科分布差異性特徵明顯,法學、管理學等社會科學撤稿記錄遠多於文學、歷史學等人文科學。其中,法學關聯撤稿記錄最多為 1,619 條,管理學(502 條)次之、其後為教育學(386 條)和經濟學(367 條)(參見表 2)。進一步細分各學科撤稿原因,因剽竊或重複發表引發的撤稿始終多於數據問題撤稿,表明人文社會科學的撤稿更多集中在學術誠信問題上。因數據問題導致的撤稿多集中在管理學(27.89%)、哲學(21.74%)、法學(16.49%)等學科,而文學(7.20%)、歷史學(6.06%)數據問題撤稿記錄佔比明顯偏低。顯而易見,不同學科研究範式對數據需求與處理的差異是造成這一現象的主要原因,法學、管理學、經濟學等需要以案例分析、問卷調查、訪談、田野調查等定量研究方法為支撐,對數據的依賴性更高,而文學、歷史學等學科更多依靠文獻、檔案、考古發現等展開研究,定量化數據需求相對有限。

#### (5) 期刊分布

2,907 篇人文社會科學撤稿論文共來自 1,503 個期刊,其中,676 篇因數據問題而撤稿的論文來自 367 個期刊。總體而言,無論是高影響力期刊還是普通學術期刊,撤稿論文分布較為均衡,15 個撤稿數量最多的期刊中,僅有 5 個 JCR Q1 區、2 個 Q2 區高影響力期刊,撤稿原因多為偽造專家評審和剽竊或重複。例如已停刊的 *Open Automation and Control Systems Journal* 撤回的 49 篇論文中,批量撤回了中國學者發表的 47 篇文章,主要原因是偽造同行評審專家;而處於 JCR Q1 區的期刊

*Multimedia Tools and Applications* 被撤稿的 39 篇論文,首要撤稿原因是剽竊或重複發表,其次為同行評審作假。

表 2 人文社會科學撤稿論文各學科撤稿原因比較

學科	法學	管理學	教育學	經濟學	哲學	文學	歷史學	藝術學
數據問題撤稿佔比%	16.49	27.89	11.87	15.53	21.74	7.20	6.06	15.56
剽竊/重複發表撤稿佔比%	56.58	44.02	58.59	58.04	50.00	65.60	65.66	53.33
其他問題撤稿佔比%	26.93	28.09	29.54	26.43	28.26	27.20	28.28	31.11

與之相對,數據問題引發的撤稿更多集中在高影響力學術期刊,撤稿數量最多的 15 個期刊中,有 10 個為 JCR Q1 區、3 個為 Q2 區,排在前五位的期刊中,有四個是心理學高影響力期刊,如 *Psychological Reports* (JCR Q2,撤稿 30 篇) *Journal of Personality and Social Psychology* (JCR Q1,撤稿 17 篇) *Journal of Experimental Social Psychology* (JCR Q1,撤稿 11 篇) 等。歸因探究,固然與高影響力期刊擁有更為嚴格、規範的論文審查和數據核驗機制有關,但不可否認的是,過度求新的期刊

表 3 撤稿論文 TOP 10 國家(地區)

總體撤稿			數據問題撤稿			
序號	國家(地區)	撤稿記錄(條)	序號	國家(地區)	撤稿記錄(條)	佔比
1	俄羅斯	697	1	美國	273	50.93%
2	美國	536	2	英國	98	56.65%
3	中國	498	3	荷蘭	90	70.87%
4	英國	173	4	中國	79	15.86%
5	伊朗	135	5	德國	40	50.63%
6	荷蘭	127	6	加拿大	26	44.83%
7	印度	122	7	印度	23	18.85%
8	德國	79	8	伊朗	20	14.81%
9	中國台灣	77	9	新加坡	16	48.48%
10	馬來西亞	68	10	澳大利亞	15	31.25%

用稿偏好以及職稱、項目等評價導向引發的學風浮躁、數據管理缺失也是撤稿論文不斷增加的催化劑。2011 年,荷蘭心理學家 Diederik A. Stapel 因實驗數據造假導致 33 篇文章被集中撤稿,面對質疑,Diederik A. Stapel 聲稱造假是為了得出學術期刊願意發表的漂亮數據與研究結論,<sup>④</sup> 揭示出學術期刊偏愛積極的發現以及高校/研究機構過度量化評價是誘發數據偽造、數據操縱等數據問題的直接驅動因素。2020 年,英國著名心理學家 Hans J. Eysenck 的 30 篇論文被 *Perceptual and Motor Skills* 批量撤稿,2019 年,倫敦大學國王學院對其論文數據的有效性,包括招募參與者、措施管理、確認結果的可靠性、數據收集中的偏見、缺乏相關協變量、對分析案例的選擇等展開調查,認定論文數據及結果不可信,建議倫敦大學告知相關期刊所發表論文中存在的數據與方法缺陷。<sup>⑤</sup>



### (6) 國家分布

就撤稿的絕對數量而言,最高的是俄羅斯、美國和中國,而數據問題引發撤稿最多的國家是美國、英國、荷蘭(參見表3)。英美等西方發達國家人文社會科學撤稿原因向更具隱蔽性的“數據問題”集中,荷蘭、英國、美國、德國因數據問題撤稿佔撤稿論文總數的50%以上,而俄羅斯(0.87%)、伊朗(14.81%)、中國(15.86%)、印度(18.85%)等發展中國家相應佔比偏低,撤稿原因集中在剽竊/重複發表以及同行評審造假、學術倫理等方面。

這一差異表明,健全的學術誠信管理制度在抑制顯性學術不端方面發揮了決定性作用,並使得學術誠信問題與數據操縱、數據捏造以及可懷疑的研究實踐等數據問題聯繫得更加緊密,而學術誠信管理制度的缺失則會助長剽竊、重複發表等撤稿現象的增長。2012年,俄羅斯提出建設科技強國的戰略任務,要求2020年至少有五所大學進入世界排名前100的大學,各領域學術論文激增,但俄羅斯科學院直到2019年才設立反學術造假委員會,滯後的學術管理制度設計與執行導致其論文剽竊與批量撤稿事件接連發生。<sup>⑥</sup>同時,不同國家數據問題撤稿數量的落差也折射出國家之間人文社會科學數據開發利用與管理水平的差異,發展中國家數據驅動型研究水平整體較低,人文社會科學數據管理制度、數據基礎設施以及數據共享利用水平相對薄弱,使得數據問題的撤稿尚未成為撤稿的主要原因。而完善的數據管理制度、開放的數據共享平台與基礎設施,既是發現數據誤用、數據造假等問題的重要手段,也是規避和抑制潛在學術不端的有效方法。

表4 人文社會科學撤稿論文數據問題細分一覽表

原因	原因描述	質量標準	撤稿佔比
數據/圖像擔憂	關於數據/圖像有效性的爭議	有效性	32.08%
數據/圖像錯誤	數據/圖像輸入、收集或識別中出現錯誤	準確性	20.68%
分析錯誤	評估數據或計算時出錯	準確性	17.92%
數據/圖像偽造	故意更改數據,使其不代表實際發現	可靠性	16.17%
數據/圖像不可信	數據/圖像準確性和有效性受到懷疑	準確性	5.89%
圖像重複	也稱“自我剽竊”,圖像被重複發表而沒有標引說明	規範性	2.39%
數據剽竊	數據被重複發表而沒有適當標引說明	規範性	2.39%
圖像操縱	通過反轉、旋轉或類似動作改變圖像的表現形式	規範性	1.13%
數據惡作劇	故意用虛假數據撰寫論文,以測試期刊接受政策	可靠性	0.74%
數據切片	為增加論文數量,將整體數據分解分別發表	可靠性	0.37%
未提供原始數據	沒有按照格式、標準等要求提供原始數據	規範性	0.24%

### (三) 撤稿論文的數據問題分析

人文社會科學數據管理邊界的複雜性、權屬關係的模糊性、主體認知與能力的不平衡性導致撤稿論文中的數據問題形式多樣、歸因複雜(參見表4)。從對象來看,既包括問卷與實驗調查數據,也涉及一定的圖像音視頻資料;從主體行為意願來看,既有數據偽造、操縱、分割利用等主觀故意(佔比25.88%),也有計算、分析錯誤等非主觀故意(佔比74.12%)。

從質量標準來看,數據問題導致的撤稿主要表現為未能滿足人文社會科學研究對數據準確性、

有效性、可靠性、規範性等質量要求。其中,首要問題是數據準確性問題,如在採集、加工與分析利用中出現的數據/圖像錯誤、分析錯誤、數據/圖像不可信等,佔比超過 43%;其次是數據有效性問題,包括對數據/圖像內容與形式的質疑或爭議,佔比 30% 以上;可靠性強調數據的“可信賴”或“可信任”程度,而因數據或圖像偽造、操縱以及文章惡作劇等導致的撤稿佔比接近 18%;規範性關注數據表達的規範化程度,由於作者的不規範引用、將大型數據集割裂使用導致的論文撤稿佔比接近 6%;完整性是期刊發表的基本要求,目前,未提供完整原始數據的撤稿僅有 2 篇。

### 三、數據生態視域下人文社會科學撤稿現象原因分析

將撤稿現象置於數據生態環境下,可從數據管理視角放大撤稿背後的學術問題,並在人文社會科學數據管理層面,將撤稿問題涉及到的學術誠信、學術評價、學術出版等一併納入數據生態建設範疇,重新理解撤稿現象,以尋求數據生態優化的可行路徑方法。此舉不僅可以避免將撤稿簡單歸結於學術誠信的思維局限,還可以擴展人文社會科學數據管理觸角,為撤稿問題包括學術誠信的綜合整治建立一個基於數據生態的參照體系。當前,人文社會科學撤稿現象所折射出的數據管理問題形式多樣,涉及多重利益關係與錯綜複雜的矛盾衝突。主要的癥結原因如下:

#### (一)數據觀念局限,價值驅動錯位

數據是科學研究的基石,人文社會科學研究需要運用大規模的數據、多種類的數字檔案、多模式的媒體文本、數據庫以及複雜的軟件和工具。不同主體的數據共享意願、數據利用態度,既決定了數據生態體系的開放與封閉程度,也直接影響到人文社會科學數據質量乃至研究成果的科學性、創新性。

撤稿論文的大量出現,尤其是有關數據問題撤稿數量的急劇增加,其一是功利化數據觀的體現,部分研究人員既撇開了數據真實性、準確性、完整性等質量要求,也放棄了嚴謹務實、追求真理的學術初心。雖然有些錯誤可以歸因於意外或不稱職的研究,但大量案例表明,很多撤稿是學者在日益激烈的競爭環境下不誠實的學術行為造成的。其二則是人文社會科學領域數據管理觀念的短視、偏見,導致數據採集、加工與利用、存儲的不規範。在數據擁有方面,許多人文社會科學研究人員習慣獨佔文獻資料並隱瞞所擁有的數據。在數據共享利用方面,美國圖書館和信息資源委員會的調查發現,社會科學研究人員對於修改和保護其科研數據的興趣不大,興趣點主要集中在數據的有用性以準備發表和出版方面。<sup>①</sup>

#### (二)數據規範缺失,制度執行遇阻

如果說數據觀念、數據管理意識是造成數據問題撤稿的內在動因,那麼,外在的數據採集利用等制度規定則對撤稿構成了無形的外在約束。在一定程度上,外在約束力的強弱決定了撤稿者數據博弈的機會成本,並對數據行為選擇產生直接影響。總體而言,人文社會科學數據管理的制度化水平滯後於自然科學,主要表現在:

(1)數據生命周期管理的制度體系不夠健全,數據行為規範缺失。美國政治科學協會(APSA)的道德準則規定,研究人員有道德義務通過數據訪問、生產透明度和分析透明度促進對其基於證據的知識主張的評估,以便他們的工作能夠得到測試或複製。但在現實中,無論是資助機構還是高校/研究機構,針對人文社會科學數據採集、加工、分析、利用的質量要求極為有限,數據管理計劃、數據管護職責以及存儲匯繳標準等缺乏統一的操作規程和強制性要求,寬泛的數據政策為各種理由的數據造假、數據誤用留有餘地。



(2) 不當數據行為界定模糊,存在制度執行盲區。從有預謀的數據偽造、數據操縱到因草率無知而出現的數據誤讀、數據不完整等非主觀行為,不當數據行為類型複雜。與自然科學研究高度依賴精確的實驗室數據相比,人文社會科學的實證、量化分析往往並非在嚴格控制變量的實驗室環境中開展,撤稿論文中存在大量有問題的研究或可疑研究,心理學、管理學、經濟學等研究方案設計、樣本選擇以及數據選取範圍等環節均容易出現紕漏。例如,根據個人偏好設定研究結論並採取靈活性數據分析獲取預期數據結果,則含有潛在數據不當行為,處在數據問題與學術不端的灰色地帶。由於處理手段有限,“偽造新藥的結果可能會導致病人死亡,而偽造商業期刊數據則不太可能直接影響特定業務”,<sup>⑧</sup>進一步加大了部分研究人員不當數據行為的投機心理。

(3) 期刊數據政策不完善,制約撤稿管理的實施。與自然科學學術出版平台嚴格的數據提交、發表、存儲等制度執行相比,人文社會科學研究成果學術出版環節對數據問題的規範籠統、關注重點局限在研究方案設計和數據生成方式的說明上,對數據提交、存儲、共享與核驗等要求的重視程度有限,僅少量期刊制定了數據匯繳與共享政策。例如,《會計研究雜誌》(JAR)要求提交的論文或附錄中應注明原始數據是如何獲得或生成的,包括數據來源、下載或獲取數據的具體日期以及用於生成數據的儀器,篩選數據的標準,數據轉換使用到計算機的程序或代碼,明確要求指明負責數據收集和管理的作者。<sup>⑨</sup>但總體而言,大多數人文社會科學學術期刊並未對作者做出數據資料提交的強制性要求,亦無統一的數據提交格式標準,更無專門的數據審查環節,同行評審聚焦在研究的理論性、新穎性和嚴謹性等方面,對數據資料核驗的重視度不高。

### (三) 數據監管不力,信息傳導低效

“學術記錄的誠信取決於驗證和自我修正機制的可靠性,而不是單一數據集的準確性”。<sup>⑩</sup>無論是主觀意圖上的數據造假,還是非故意的數據缺失、數據誤用,無論是有意的可疑研究實踐及結論,還是無意的資料記錄錯誤和數據遺漏等,撤稿論文中形形色色的數據問題清晰揭示了人文社會科學數據過程監管和學術誠信管理的薄弱。主要表現如下:

(1) 項目資助機構、高校/研究機構等相關主體未能對研究人員提交的科研項目數據管理計劃、數據管理實施進展等進行必要的審核和指導。許多學科,學術組織方式上基本屬於學者個體或者小型團隊自主性學術行為,外在的參與和監督十分有限,導致研究人員在科研組織以及數據準備和結果分析中享有過大的自由裁量權。

(2) 研究團隊數據管理職責的弱化。從數據分析中發現,數據問題撤稿的作者合作密度高於總體撤稿的平均作者合作密度,2,907篇撤稿論文中,有1,992篇(佔比68.52%)係多作者合著,有1,255篇(佔比43.17%)是多機構合著;而676因數據問題撤稿論文中,有570篇(佔比84.32%)是多作者合著,388篇(佔比57.40%)是多機構合著,心理學、經濟學、管理學等數據密集型學科大量撤稿論文屬於團隊研究成果,容易在數據分析處理過程中出現混亂和錯誤累積。同時,合作研究責任的分散,常常給人一種認知假設,即其他成員將會檢查和重新檢查設備、樣本、數據和分析,從而使各種問題從裂縫中消失。

(3) 期刊及審稿人的數據把關不到位。儘管2016年的COPE指南明確規定:“期刊應對其編輯的行為負責,保護研究記錄,並確保其發表的一切內容的可靠性”。<sup>⑪</sup>但在注意力稀缺的現實情境下,複雜數據問題的檢測需要審稿人和編輯高強度的細緻檢查,並以大量的時間成本和機會成本為保障,這與普遍奉行的效率導向原則構成衝突,一些期刊甚至沒有使用任何形式的抄襲檢測軟件,導致學術發表環節的數據監管流於形式。同時,部分撤稿聲明對問題陳述的審慎與含混不清,如研

究不可複製、研究可信度不足等溫婉表述客觀上麻痺了人們對數據問題嚴重性的認識。

#### (四) 基礎設施投入不足, 協同治理薄弱

準確、高效的數據管理需要以高水平的數據庫、數據存儲利用平台和便捷的應用工具等基礎設施為支撐。國外的調查指出, 項目資助機構和管理部門越來越深刻地意識到研究人員越來越依賴於功能完善的、可持續的科學數據管理基礎設施。<sup>22</sup>伴隨着數據科學、數字人文以及計算社會學等學科的興起, 以數據平台、數據存儲中心等為代表的基礎設施對於推進人文社會科學研究的支撐性和重要性進一步凸顯, 需求也越來越高。

當前, 即便有形式多樣的與人文社會科學相關的數據存儲庫或數據中心、數據平台, 但與現實需求相比, 人文社會科學研究既面臨着數據基礎設施不足的困境, 也面臨着研究人員對數據基礎設施建設參與不足、利用不高的雙重實踐難題。歐洲針對人文社會科學研究人員的調查結果表明, 55%的人回答沒有參與數據基礎設施建設, 主要理由是沒有興趣, 其次是不知道數據基礎設施對自己的研究有多大幫助,<sup>23</sup>人文社會科學領域還沒有形成專注於共享和使用開放數字資源的學術文化。對此, 德國北部的 University of Bielefeld 提出人文社會科學數據管理必須以技術基礎設施、政策和宣傳推介三個維度為基礎的三大支柱戰略來加以實施;<sup>24</sup>《倫敦大學 2020—2025 發展戰略》也指出要為跨學科調查建立知識基礎設施, 使廣大研究人員從中受益, 並支持藝術和人文學科的使用。<sup>25</sup>

撤稿中的數據問題不僅揭示了研究人員、數據管理人員以及期刊編輯等職責的缺位, 也映射出項目資助機構、研究機構等主體數據監管的協同不力。如果不是大規模的期刊撤稿引發各界關注, 一般性撤稿並不會及時反饋到作者所在單位和項目資助機構, 更不會引起學術誠信委員會/辦公室的注意, 對作者的影響特別是普通作者的影響相對有限。美國會計學會承認期刊出版部門應對數據完整性等問題負責, 但由於這些部門沒有能力進行深入調查與核驗, 認為該職責應由學院、大學和資助機構履行, 並鼓勵作者向其他研究人員分享數據。即便如此, 該學會並未指明數據保存平台和方法,<sup>26</sup>對數據問題的處理缺乏多主體的分工負責與整體協同。

當然, 人文社會科學研究的抽象性、思辨性等內在屬性決定其數據加工與解釋不同於自然科學的實驗記錄, 在細緻跟蹤、客觀描述、忠實記錄基礎上, 還需要更高的數據敏銳性和學術創造力, 不僅打破了學術創造、學術服務和學術傳播之間涇渭分明的界限, 也模糊了高校/研究機構、研究人員、數據管理者、期刊編輯、數據平台等傳統學術主體的角色和責任, 需要形成多主體協同參與人文社會科學學術研究與數據管理乃至學術誠信管理的新格局, 即學術研究的完整性和準確性取決於由個體研究人員、學術同行和學術界等多主體共同組成的互為支撐的生態體系。

## 四、思考與對策建議

數據既是數字化時代學術成果及其記錄的關鍵組成部分, 也是決定和考察學術誠信的決定性要素。“在科學研究中, 有三件事很重要: 數據, 收集數據的方法(這給了他們鑒定價值)以及將數據和方法與結論有機聯繫起來的邏輯”。<sup>27</sup>健全的數據管理制度、完善的數據流程、開放的數據存儲共享平台足以保證數據及研究成果的可發現性、可訪問性、互操作性與再利用性, 進而降低問題論文的出現, 規避有意或無意的數據錯誤。

儘管人文社會科學撤稿論文總量以及因數據問題而撤稿數量有限, 但並不意味着人文社會科學領域整體的數據管理規範、數據開放共享與學術誠信程度較高, 而是其研究範式與數據管理的複

雜性特殊性導致研究成果中的數據問題更加隱蔽,加之整體的數據意識薄弱、數據素養不足而難以驗證。例如,人文社會科學研究更加重視研究視角、理論、方法及結論的原創性、示範性,更加注重研究規律與發展趨勢的描述性總結與推廣,而對調查或實驗數據進行複製、驗證的需求相對較低,使得數據問題的發現具有明顯的滯後性。同時,由於研究成果的多樣性,書籍、手稿、詩歌、創造性寫作、地圖、照片、藝術、新聞、娛樂和許多其他類型的文本乃至考古發現等實物資料,事實上使得人文社會科學數據及成果的開放共享成本與技術實現難度更高,數據及其資料的核驗複製耗時費事。當然,人文社會科學學術研究本身的抽象性思辨性,客觀上使得研究成果的數據問題,即便是主觀故意,也有可能以不太嚴重的形式出現,如選擇性報告和遺漏數據,由於沒有明確的辨析界限,有時會處於數據處理不當和學術不端的模糊地帶。<sup>⑧</sup>無論動機如何,以不同形式出現的數據問題,本質上既是數據管理問題也是學風與學術誠信問題,一些國外學者甚至將可懷疑的研究實踐、迎合型或靈活型統計分析歸結為隱形學術不端行為。

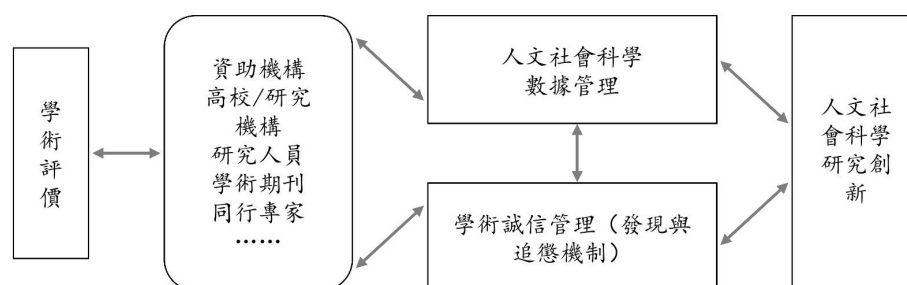


圖3 人文社會科學數據管理與學術評價、學術誠信的內在關係

在現實中,許多學術誠信準則等制度規範在應對快速發展的數字化變革帶來的管理挑戰上仍然存在缺陷,無論基於工具理性還是價值理性,學術誠信的倡導與推進都需要有具體明確的載體形式,而不是空中樓閣和枯燥說教。數據生態着眼於數據產生、流動的整體情境與價值實現,不僅是人文社會科學研究的存在狀態,也是學者賴以依存並運用數據流開展學術創新的操作系統,還是透視數據行為以檢驗學術誠信的試金石。通過對數據收集方式、研究設計、數據整合以及分析技術等綜合檢驗,可以將學術誠信管理中抽象的、框架性原則轉換為具體的、連續性數據行為,並對數據問題引發撤稿的主體、過程與後果進行綜合判斷。

當然,決定各主體數據行為選擇的既有學術創新的源動力,也有職務晉升、項目申請等評價指揮棒的外在壓力,當外在壓力形成人文社會科學學術創新的正向激勵時,研究人員的數據管理計劃及其數據採集、加工、傳播、共享等行為會沿着規範化軌跡,實現包括學術創新與學術誠信的自律;而一旦評價指揮棒異化,外在壓力就會對學術行為形成負向激勵,如“貨幣化”學術激勵、“數量化”考核壓力等就會為學術不端提供強大驅動力,並導致數據造假、數據誤用等關聯性實踐。有人指出,“在研究過很多科研不端案例後可以發現,研究人員面臨着巨大壓力,採用數據造假是一個很容易解決他們科研困境的途徑。”<sup>⑨</sup>可見,學術評價、數據生態與學術誠信之間是前後銜接的邏輯鏈條,考察撤稿論文的數據問題,可以更集中地反映學術評價與學術誠信之間、學術研究與學術傳播(期刊出版)之間的內在邏輯關係。涵養有利於人文社會科學研究創新的學術生態,關鍵着力點是在數據管理、學術誠信以及學術評價之間形成互為促進的正反饋效應(見圖3)。一旦孤立、片面地看待三者間關係,則難以切中要害無助於撤稿現象的根治與杜絕。

近年來,國家有關部門對學風建設與數據管理等方面均表示了高度的關注,並先後出台了相關的文件。2019年,國務院辦公廳印發的《關於進一步弘揚科學家精神加強作風和學風建設的意見》指出,論文等科研成果發表後1個月內,要將所涉及的實驗記錄、實驗數據等原始數據資料交所在單位統一管理、留存備查。2018年,澳大利亞研究理事會、澳大利亞大學聯盟等聯合推出了《澳大利亞負責任研究行為準則》,也將恰當地管理和保存研究數據納入到負責任研究行為的重要內容。上述政策文件表明,科研過程中的數據生態已成為學術誠信治理的重要組成部分。隨着數據驅動型研究範式在人文社會科學研究中的廣泛應用,數據生態及學術評價制度不僅成為根治撤稿現象的關鍵,也是恪守學術誠信的決定性要素。

將數據生態理念融入人文社會科學科研管理實踐,為解決當前的學術誠信管理問題提供了新的視角和方案,有必要實現政府、資助機構、高校、期刊等多主體數據管理制度的橫向貫通與撤稿權限、程序、責任的具體化規範化。同時,還要不斷提升研究人員的數據素養,健全人文社會科學數據的生命周期管理制度體系,建設高水平的數據存儲共享平台,進一步推動數據驅動型研究範式的普及,以此激發人文社會科學研究人員的學術自律,打造標本兼治的良好學術氛圍。

本文雖然從數據生態視角揭示了人文社會科學撤稿中數據問題的概況、成因與表現,但對於撤稿及其撤稿中數據問題與學術誠信管理之間的內在關係及作用機理並未展開深度分析與論證,限於篇幅原因有待於在後續研究中進一步論證和說明。

[致謝:本文的寫作得到了武漢大學信息管理學院董克副教授和博士研究生管茜同學在資料與數據加工方面的支持,在此表示感謝。]

①B. K. Redman & J. Merz, Policies and Quality Assurance Research Misconduct Policies of High Impact Biomedical Journals, *Accountability in Research* 92 (2), 2004, pp. 242-246.

②J. Brainard, et al., Rethinking Retractions, *Science* 362 (6413), 2018, pp. 390-393.

③趙河雨:《美國本特利大學教授因學術不端被撤稿31篇》,北京:科學網,2015-06-30, <http://news.sciencenet.cn/htmlnews/2015/6/321791.shtm>, 檢索日期:2021年8月6日。

④B. P. Schmidt UNLV Professor Is Investigated for Career-Spanning Plagiarism, *Chronicle of Higher Education* 61 (1), 2014, pp. A16-A16.

⑤蔣穎:《科研誠信視角下的人文社科國際學術論文撤稿特徵研究》,北京:《情報資料工作》,2020年第6期。

⑥S. Necker Scientific Misbehavior in Economics, *Research Policy* 43 (10), 2014, pp. 1747-1759.

⑦R. Craig, et al., Using Retracted Journal Articles in

Psychology to Understand Research Misconduct in the Social Sciences: What is to be done?, *Research Policy* 49 (4), 2020, 103930.

⑧Dennis Tourish, Russell Craig, Research Misconduct in Business and Management Studies: Causes, Consequences, and Possible Remedies, *Journal of Management Inquiry* 29 (2), 2018, pp. 174-187.

⑨M. D. Ribeiro, S. M. R. Vasconcelos, Retractions Covered by Retraction Watch in the 2013-2015 Period: Prevalence for the Most Productive Countries, *Scientometrics* 114, 2018, pp. 719-734.

⑩Pierre Azoulay, et al., The Career Effects of Scandal: Evidence from Scientific Retractions, *Research Policy*, 2017, 46 (9): 1552-1569.

⑪Xin Shuai, et al., A Multidimensional Investigation of the Effects of Publication Retraction on Scholarly Impact, *Journal of the Association for Information Science and Technology* 68 (9), 2017, pp. 2225-2236.

⑫魏眾、蔣穎:《中國人文社會科學學術期刊撤稿問

題研究》，澳門：《澳門理工學報》，2020年第4期。

⑬F. Agnoli et al., Questionable Research Practices Among Italian Research Psychologists, *PLoS ONE*, 2017, 12(3): e0172792.

⑭M. A., Edwards, & S. Roy, Academic Research in the 21<sup>st</sup> Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition, *Environmental Engineering Science*, 2017 (34), pp. 51-61.

⑮King's College London, King's College London Enquiry into Publications Authored by Professor Hans Eysenck with Professor Ronald Grossarth Maticsek. <https://www.kcl.ac.uk/news/statements/docs/hans-eyenck-enquiry-final-may-2019.pdf>, 檢索日期:2021年6月10日。

⑯楊梟:《俄羅斯近千篇論文被撤稿,未來還會有更多》, <http://new.qq.com/omn/20200119/20200119A07HX900.html>, 檢索日期:2021年6月24日。

⑰Nancy-Diana, Gómez et al., Social Sciences and Humanities Research Data and Metadata: A Perspective from Thematic Data Repositories, *Profesional de la Información* 25(4), 2016, pp. 545-555.

⑱B. Honig et al., Ethics in the Production and Dissemination of Management Research: Institutional Failure or Individual Fallibility?, *Journal of Management Study*, 51 (2014), pp. 118-142.

⑲Journal of Accounting Research, Author Guidelines, <https://onlinelibrary.wiley.com/page/journal/1475679x/homepage/forauthors.html>, 檢索日期:2021年6月10日。

⑳ Heather Coates, Ensuring Research Integrity: The Role of Data Management in Current Crises, *College & Research Libraries News* 75(14), 2014, pp. 598-601.

㉑E. Wager, S. Kleinert. Cooperation between Research Institutions and Journals on Research Integrity Cases: Guidance from the Committee on Publication Ethics, *Saudi Journal of Anaesthesia* 72(2), 2012, pp. 165-169.

㉒SCIENCE EUROPE.Funding Research Data Management and Related Infrastructures. [https://repository.jisc.ac.uk/6402/1/Funding\\_RDM\\_%26\\_Related\\_Infrastructures\\_MAY2016\\_v7.pdf](https://repository.jisc.ac.uk/6402/1/Funding_RDM_%26_Related_Infrastructures_MAY2016_v7.pdf), 檢索日期:2021年6月10日。

㉓DARIAH, Barriers and Pathways to Community Engagement, *Report on the Activities of the DARIAH VCC2 Community Engagement Working Group*, 2018.

㉔J. Schirrwagen, et al., Expanding the Research Data Management Service Portfolio at Bielefeld University according to the Three-pillar Principle towards Data FAIRness. <https://pub.unibiel-efeld.de/publication/2919659>, 檢索日期:2021年6月10日。

㉕University of London, University of London Strategy 2020-2025. <https://london.ac.uk/strategy/vision-and-objectives#researching-22805>, 檢索日期:2021年6月10日。

㉖American Accounting Association, AAA Publications Ethics Policy - Data Integrity. <https://meridian.allenpress.com/DocumentLibrary/AAAN/Data-Integrity.pdf>, 檢索日期:2021年6月10日。

㉗DB. Allison, Letter for Hearing on Pros and Cons of Restricting SNAP Purchases, Committee on Agriculture, House of Representatives, 115th Congress, 1st Session (U.S. Government Publishing Office, Washington, DC)2017.

㉘Jeremy Hall & Ben Martin, Towards a Taxonomy of Research Misconduct: The Case of Business School Research, *Research Policy* 48(2), 2019, pp. 414-427.

㉙王曉真:《完善學術誠信保護舉措》,北京:《中國社會科學報》,2019年8月28日。

作者簡介:夏義堃,武漢大學信息管理學院教授,信息資源研究中心副主任。武漢 430072

[責任編輯 劉澤生]