

构筑日本未来的学术信息流通体系

——NACSIS-CAT/ILL 轻量化·合理化方案(CAT2020)简述

北京大学图书馆 力恺

1 NACSIS-CAT/ILL 轻量化·合理化计划产生的背景

1980年日本学术审议会基于对当时日本学术信息流通体系的分析发布了名为《关于未来学术信息系统的形态》的报告,提出一系列对日本未来学术信息系统功能的构想^[1]。依照该构想设计的联合目录系统 NACSIS-CAT/ILL 一直以来是支撑日本学术信息流通体系的核心系统。自 1984 年完成系统开发、1985 年正式投入使用以来,历经 30 余年,其数据结构、运行方式等基本未发生重大的更改。

但是,相比 NACSIS-CAT/ILL 设计之初,近年来日本的学术信息流通环境产生了几十个重要的变化:其一,资源数量不断增长,其中电子资源所占比例越来越高。其二,以在校生为主的用户利用信息资源学习研究的方式日益趋向数字化^[2]。其三,用户对信息资源的检索行为不再局限于 NACSIS-CAT/ILL 系统之内,而是通过搜索引擎、数据库、文献资源发现系统等不同渠道检索资源^[3]。最后,各大学图书馆从事编目工作的人员数量基本呈现逐年递减的态势^[4]。面对上述变化,NACSIS-CAT/ILL 已经不能充分满足当前环境下的学术信息资源组织与服务需求。

为了让 NACSIS-CAT/ILL 跟上时代的发展,NACSIS-CAT/ILL 的运行维护机构日本国立情报学研究所(National Institute of Informatics, NII)于 2015 年会同高校图书馆的专业馆员代表,在“未来学术信息系统构筑研讨委员会”(これからの学術情報システム構築検討委員会,以下简称委员会)的框架下设立了“NACSIS-CAT 研讨工作会议”(NACSIS-CAT 検討作業部会),正式展开对 NACSIS-CAT/ILL 进行全面升级的讨论。经过数次研讨,委员会先后公开了《关于 NACSIS-CAT/ILL 的轻量化·合理化(基本方案要点)》^[5]、《关于 NACSIS-CAT/ILL 的轻量化·合理化(基本方案)》^[6]、《关于 NACSIS-CAT/ILL 的轻量化·合理化(实施方案)》^[7]这几个文件,由抽象到具体初步确立了 NACSIS-CAT/ILL 的升级方案内容。随后,委员会针对实施方案的内容向 NACSIS-CAT/ILL 成员馆广泛征求意见并据此进行了调整,发布了《关于 NACSIS-CAT/ILL 的轻量化·合理化(实施方案)的变更说明》^[8]。2018 年 10 月,委员会发布了题为《关于 NACSIS-CAT/ILL 轻量化·合理化(最终方案总结)》^[9]的文件,正式确定了 NACSIS-CAT/ILL 系统的升级方案。为了方便与旧 NACSIS-CAT/ILL 系统进行区分,2020 年升级后的新一代系统被称为 CAT2020。

2 CAT2020 升级方案的主要内容及其分析

NACSIS-CAT/ILL 系统的设计基于上世纪八十年代的计算机处理能力,以初步实现计算机辅助处理的文献资源组织与服务为目标。在这套系统中,编目员除了要进行著录、标引等工作,还要建立和维护书目记录之间的关联,通过构筑复杂的书目记录层级结构来实现对书目信息的组织与呈现。在这个过程中,计算机系统仅负责书目记录数据的简单存取与直观再现,大量复杂而繁琐的文献资源组织工作主要依靠人工进行。随着前述学术信息流通环境的变化,这种单纯倚重编目员人工进行的文献资源组织方式难以为继。而书目记录结构过于复杂也制约了计算机对于书目信息的处理能力。为了解决上述问题,委员会对于 NACSIS-CAT/ILL 系统升级方案确立了“轻量化”

与“合理化”两个基本方针。“轻量化”是指减轻书目记录编制及维护工作中人作业的负担;“合理化”则是指合理改进书目记录结构,使其更加适应计算机处理的需求,同时提升与外部系统的兼容性从而强化书目信息服务的可扩展性。遵循上述基本方针,委员会制定了如下升级方案。

2.1 书目记录前置导入

委员会在设计 CAT2020 书目记录结构时,将继承自 NACSIS-CAT/ILL 的书目记录集合,即已经有馆藏记录的书目记录集合设为“BOOK”。此外还添加了一个“PREBOOK”集合——将由其它机构(如日本国立国会图书馆(NDL)、图书馆流通中心(TRC)、美国国会图书馆(LC)等)编制的书目记录提前导入 CAT2020 系统以供后续套录使用。导入时系统会抽取 ISBN 等信息对记录进行判重,如果判定与“BOOK”或“PREBOOK”中的既存书目记录一致则视为重复记录不予导入,但系统会从不被导入的书目记录中自动提取 OTHN(其他编号)、NBN(日本全国图书编号)、LCCN(美国国会图书馆书目记录编号)、GPON(美国政府印刷局编号)、SH(主题)、CLS(分类)字段中的数据,将其补充到系统中对应的书目记录之中。

CAT2020 通过设置“PREBOOK”来保存尚未有馆藏的外部书目记录,并与“BOOK”中已有馆藏的正式书目记录区分开。在图书馆的资源发现、馆藏目录及馆际互借等服务环境下,读者仅检索“BOOK”中的书目记录。而在编目作业环境下,编目员则可以统一检索“BOOK”及“PREBOOK”中所有的书目记录。当发现在编文献在“PREBOOK”中有对应书目记录,并向其添加馆藏记录时,该记录将自动迁移到“BOOK”之中。这种设计让外部书目记录与 CAT2020 的编目环境无缝对接,使编目员能够便捷地利用更多的套录数据,从而提高编目工作效率。

2.2 对于书目记录的结构进行调整

2.2.1 以资源的出版物理单位为准编制书目记录

NACSIS-CAT/ILL 在资源的单位划分上使用了 3 个概念:物理单位、出版物理单位、书目记录单位^[10]。单个资源保持一个完整的物理形态就可以视为一个物理单位,比如图书馆书架上的某一册书。相同的物理单位(即我们所说的复本)组成的一个集合即为出版物理单位,比如出版社发行的某本书。而在 NACSIS-CAT/ILL 中,一个书目记录单位则是由汇聚在一个特定题名下的一个或多个出版物理单位构成的。例如,作为单行著作发行的某本书,或是拥有共同题名但没有各自不同分辑题名的多卷书,都被视作一个书目记录单位进行编目。在 CAT2020 中,书目记录单位的划分不再依据出版物理单位所拥有的题名形式进行判断,而是统一按照出版物理单位划分。即在新标准下书目记录单位等于出版物理单位。新的划分标准在 CAT2020 系统切换之日起开始实施,但不对以往的书目记录进行回溯性更改。由于古籍以及缩微文献在资料形态上的特殊性,在编目时允许继续对其进行集中著录。

在 NACSIS-CAT/ILL 中集中著录的多卷书,多个分辑的书目信息被放在一个书目记录中描述,书目记录与资源实体不是一一对应的关系。各分辑的出版时间、页码、所包含的章节题名或是其对应的数字化资源的 URI 等信息在集中著录的记录里往往得不到详细、专指的描述。这些信息或者被简化、省略,或者被罗列于附注字段。这种对资源内容和载体形态描述方面的局限性,限制了书目信息的检索精度与呈现深度。而 CAT2020 将书目记录的编制单位统一到出版物理单位,使资源描述与资源实体做到一一对应,从而突破了上述局限。此外,在这种设计下借阅次数、预约次数等读者服务数据和书目订单等资源建设数据也与资源实体实现了一一对应的关系,使计算机能够针对不同资源实体准确抽取和统计相关数据,这对于图书馆通过数据分析提供

高水平的读者服务是很有意义的。

2.2.2 不再强制编制高层记录与维护其链接

NACSIS-CAT 对分散著录的多卷书规定必须建立高层记录,并且每个分辑的书目记录都要与高层记录建立连接。CAT2020 则将其改为可选操作,各机构在编制多卷书的书目记录时,如果发现没有相应的高层记录,既可以选择编制高层记录并在丛编题名字段建立连接,也可以选择仅著录丛编题名字段而不再连接高层记录。

日本编目界原有的高层记录规则主要是为了应对日本出版物中丛书较多的现象,在计算机处理能力有限的时代通过建立高层记录将相关的书目记录汇聚在一起加以揭示。而如今随着技术进步,已经可以由计算机通过对书目信息的分析和自动处理,将丛书各分辑的书目信息按不同的检索需求汇聚与揭示,编制高层记录的必要性也因此消失。

2.2.3 废除非均衡书目记录结构的概念

在 NACSIS-CAT 中,非均衡书目记录结构是指丛编中各分辑在书目记录层级上存在不统一的情况。例如丛编《中心地論》先出版了卷 1、卷 2 且没有独立的分辑题名,其后又追加出版了卷 3,且该卷有分辑题名《西ドイツにおける地域政策への応用》。按照 NACSIS-CAT 的规则,卷 1、卷 2 以“中心地論”为正题名进行集中著录(以下称记录 A);而卷 3 则以其分辑题名作为正题名分散著录(以下称记录 B)，“中心地論”则作为丛编题名。此时记录 A 虽然携带卷 1、卷 2 的卷册信息,但按照 NACSIS-CAT 的规则仍会被作为该丛编的高层记录与记录 B 进行连接。如此,卷 1、卷 2 被著录于作为高层记录的记录 A,而卷 3 则被著录于作为分辑记录的记录 B,同一丛编的三个分辑实际上处于不同的书目记录层级,从而形成非均衡书目记录结构。而 CAT2020 规定带有卷册信息的记录不作为高层记录使用。此时记录 A 不能作为丛编高层记录与记录 B 连接,但如有需要,可以另外编制不携带卷册信息的书目记录作为高层记录使用。如此,在 CAT2020 下《中心地論》的所有分辑将处于相同的书目记录层级。废除非均衡书目记录结构的概念将高层记录与分辑书目记录彻底进行了区分,使书目记录层级结构的边界更加清晰。

上述 2.2 中的调整是对书目记录结构的一种扁平化设计,以便于计算机的识别与处理。同时这种设计也细化了书目记录的信息粒度,从而实现对资源实体更为详细和专指的描述。这与当下以 RDA、NCR2018 等为代表的,基于 FRBR 模型的新一代编目规则的发展趋势也是相适应的。CAT2020 对于书目记录结构的革新对于我国的编目工作也具有一定的参考价值。

2.3 新增并立书目记录机制

CAT2020 新增了并立书目记录机制,允许在一定条件下对同一文献资源采用不同的书目记录以不同的方式进行描述。在这种情况下产生的多个书目记录称为并立书目记录。描述同一资源的多个并立书目记录将不被视为重复记录,因而不会被 CAT2020 系统合并。例如以下情况中新编记录与既有记录属于并立书目记录的关系:

- (1) CAT2020 系统切换前已经按集中著录方式编制了多卷书书目记录,在切换后允许按照出版物理单位重新编制分辑的书目记录。
- (2) 原先按照单行著作著录的某本书,由于出版了续卷,其题名信息的一部分(如其他题名信息)如今成为了丛编题名,此时允许将此书以丛编分辑的形式重新编制一个书目记录。
- (3) 某成员馆发现其在编文献已有对应书目记录,但是书目记录中的描述与在编文献

存在不同,而按照CAT2020规则又不允对相关字段直接进行修改时(如某本书同一版次不同印次的页数存在差异),允许重新编制一个书目记录以准确描述在编文献。

CAT2020还设置了一个专用的关系记录表(relation table)存储并立书目记录的控制号(书誌ID),从而为相关记录建立联系。在NACSIS-ILL等书目信息服务界面下,相关的并立书目记录可由系统通过关系记录表进行识别,并以聚合成组或相互链接的形式呈现给读者。

并立书目记录机制使NACSIS-CAT/ILL切换到CAT2020后根据新旧规则编制的书目记录能够并存,在大幅减少了系统切换所带来的维护工作量的同时也提高了编目工作的灵活性。

2.4 将书目记录与其对应的著者名称规范数据自动进行连接

CAT2020针对拟前置导入的外部书目记录中一部分具备条件的记录,可由系统利用虚拟国际规范文档(Virtual International Authority File, VIAF)自动形成与NACSIS-CAT著者名称规范数据的连接。这些条件包括:①外部书目记录中包含有著者名称的规范记录编号,且该编号已经登记到VIAF;②NACSIS-CAT中存在对应的著者名称规范数据,且该记录编号也已经登记到VIAF。在导入符合上述条件的外部书目记录时,系统可以利用记录中的著者名称所指向的VIAF数据匹配到NACSIS-CAT中对应的著者名称规范数据,从而自动进行连接。

当下,RDA以及NCR2018等新一代资源组织规则陆续开始实施,在新规则中名称规范控制对于准确识别资源及在此基础上的资源组织至关重要,可以说是今后编目工作中必不可少的一个环节。CAT2020应用VIAF对外部书目记录自动连接对应的本地著者名称规范数据,提升了规范控制工作的效率,减轻了编目员的负担。这对于我国编目界也是一个很好的提示,应该加速推动我国的规范控制工作纳入以VIAF为代表的国际规范数据共建体系中去,以使世界范围的规范数据能够更加广泛地应用于我国的编目工作。

2.5 废除书目记录的馆际协调制度

CAT2020废除了NACSIS-CAT/ILL规则下成员馆对非本馆创建的书目记录进行某些修改时必须需要与该记录的创建馆协商操作的馆际协调制度。允许成员馆在遇到上述情况时根据在编资源的实际情况另行新建书目记录加以描述。

随着规范数据的广泛应用及其带来的计算机识别与自动处理书目记录能力的提升,CAT2020得以通过并立书目记录机制使对同一资源的不同描述能够并存。从而将本就紧缺的图书馆人力资源从费时费力的馆际协调工作中解放出来,投入到其他工作中去。

3 小结

在设计NACSIS-CAT/ILL系统之时,学术信息的载体以传统的纸质文献为主,读者的检索习惯还是基于卡片目录时代的方式。而且受当时的技术条件所限,计算机系统无法对书目记录进行深度处理,文献资源的组织和揭示工作主要还是依赖人工进行。在NACSIS-CAT/ILL中,书目记录的制作与服务是一体的,读者所见即为编目员所编,这种设计有其时代合理性。而时至今日,随着学术信息资源电子化不断加速,信息粒度不断细化,读者的检索习惯也随之逐渐由查找具体的资源实体向查找关心的内容转变。面对这种转变,CAT2020采取了将书目记录的制作与服务分离的设计理念,使其从一种提供给读者的最终信息产品转变为一种提供给学术信息检索系统进一步

处理的中间信息产品,读者所见不再局限于编目员所编书目记录的简单呈现。通过前述的扁平化结构设计,CAT2020 赋予了书目记录更高的结构弹性与信息容量,在此基础上充分发挥计算机的处理能力,使其深度参与书目记录的制作、服务与维护工作,为计算机处理条件下的高水平学术信息服务奠定了基础。

NACSIS-CAT/ILL 是我国高校日文文献编目工作所使用的主要外部数据源^[11],因此从 NACSIS-CAT/ILL 到 CAT2020 的系统升级将对我国今后的日文编目工作产生深远的影响。2020 年 8 月 3 日,CAT2020 已经正式投入使用,国内编目界应尽快对其编目规则的细则展开更为全面、深入的研究与讨论,以便使我国的日文编目工作实践能够与其顺利对接。同时我们也应该了解到,面对不断发展变化的学术信息流通环境,CAT2020 也仅仅是迈向日本未来学术信息流通体系的一个过渡性系统,需要继续关注其后续发展动向,汲取其中的有益经验,研究我国的应对之策。

参考文献:

- [1] 学術審議会. 今後における学術情報システムの在り方について(答申) [EB/OL].
https://www.jstage.jst.go.jp/article/jcul/16/0/16_638/_pdf/-char/ja
- [2] 学術図書館研究委員会. 学術情報の利用に関する調査(SCREAL) 2014 基本集計 [EB/OL].
http://www.screal.jp/2014/SCREAL2014_summary.pdf
- [3] 三角太郎. NACSIS-CAT 検討作業部会における検討状況 [EB/OL].
https://www.nii.ac.jp/content/korekara/archive/jla_misumi_20161016.pdf
- [4] 佐藤義則. NACSIS-CAT の再構築 [EB/OL].
https://www.nii.ac.jp/csi/openforum2016/track/pdf/20160526_CAT_1_sato_y.pdf
- [5] これからの学術情報システム構築検討委員会. 「NACSIS-CAT/ILL の軽量化・合理化について(基本方針案の要点)」 [EB/OL].
https://www.nii.ac.jp/content/korekara/archive/korekara_doc20151027.pdf
- [6] これからの学術情報システム構築検討委員会. 「NACSIS-CAT/ILL の軽量化・合理化について(基本方針)」 [EB/OL].
https://www.nii.ac.jp/content/korekara/archive/korekara_doc20160629.pdf
- [7] これからの学術情報システム構築検討委員会. 「NACSIS-CAT/ILL の軽量化・合理化について(実施方針)」 [EB/OL].
https://www.nii.ac.jp/content/korekara/archive/korekara_doc20170208.pdf
- [8] これからの学術情報システム構築検討委員会. 「『NACSIS-CAT/ILL の軽量化・合理化について(実施方針)』からの変更について」 [EB/OL].
https://www.nii.ac.jp/content/korekara/archive/korekara_doc20180125.pdf
- [9] これからの学術情報システム構築検討委員会. NACSIS-CAT/ILL の軽量化・合理化について(最終まとめ) [EB/OL].
https://www.nii.ac.jp/content/korekara/archive/korekara_doc20181019.pdf
- [10] 国立情報学研究所. 目録情報の基準 第4版 [EB/OL].
<https://www.nii.ac.jp/CAT-ILL/archive/pdf/kijun4.pdf>
- [11] 李晨英, 韩明杰, 谢琴芳. 我国高校图书馆日文文献书目数据库建设研究现状 [J]. 图书情报工作, 2002(3): 74-77