

〔美〕F.W.兰开斯特等 著



情报系统的兼容性

科学技术文献出版社

情报系统的兼容性

[美] F.W.兰开斯特 著
L.C.史密斯

姚维范 译 刘昭东 校

科学技术文献出版社

内 容 简 介

情报系统的兼容性对于情报服务机 是决定两个或两个以上机构之间的交流范围和数据交换范围的一种特性。这些机构的工作越是兼容，交流和交换也就越容易进行。因此实现兼容性十分重要，它是有效合作、互相连接和资源共享的先决条件。

本书适用于各类情报系统、图书馆系统，以及各种计算机系统；也可作为大专院校计算机专业、图书情报专业师生的参考书。

COMPATIBILITY ISSUES AFFECTING INFORMATION SYSTEMS AND SERVICES

Prepared by
F. Wilfrid Lancaster
and

Linda C. Smith
Unesco, Paris, 1983

情报系统的兼容性

[美] F. W. 兰开斯特 著
L. C. 史密斯

姚维范 译 刘昭东 校

科学技术文献出版社出版

北京昌平星城印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

787×1092毫米 32开本 9.5印张 204千字

1989年10月 北京第一版 第一次印刷

印数：1—3300册

科技新书目：202—119

ISBN 7-5023-0866-0/Z·123

定价：3.80元

序 言

对于情报服务机构来说，兼容性是决定两个或两个以上机构之间的交流范围和数据交换范围的一种特性。这些机构的工作越是兼容，交流和交换也就越容易进行。实现兼容性十分重要，因为它是有效的合作、互相连接和资源共享的先决条件。

在某项工作中，使两个机构获得兼容性的一种显而易见的方法，是使双方在进行工作时遵守相同的标准。这样，对一个领域来说，制订标准的机构及其工作是至关重要的，而情报服务界对国际标准的认可和采纳也是十分重要的。

然而并不是所有的标准都能促进兼容性，而只是那些促进机构之间某种形式的交换的标准才是这样的。此外，在采用不同标准的机构之间，通常也是可以获得兼容性的，虽然这样做不很容易，也许费用更大一些。一种方法是直接将一个中心(A)的输出转换成适当的形式，用以作为另一个中心(B)的输入。比如，编写一个计算机程序，将A的记录格式转换成能被B的计算机执行的格式。另一种获得机构间兼容性的方法是建立某种形式的中间装置，它能起到一种转换机构的作用，可将任何一个有关中心的输出转换成被另一个有关中心可用的形式。因此，采用一种“通用交换格式”可将A所用的处理格式转换成适用于B（以及C、D等等）的处理格式，反之亦然。“中介词典”对于某些机构的检索词汇

起到同样的作用；而转换语言，对于这些机构的计算机识别命令语言来说，也可获得相同的结果。所以，转换性可以看作是建立兼容性的一种重要手段。

计算机和通讯技术极大地促进了机构之间的信息交流。正是由于这一原因，情报服务的自动化对兼容性的要求比以往任何时候都更加迫切。另一方面，因为计算机允许某些转换自动进行（如记录格式），所以计算机本身也能促进兼容性。

情报服务极大部分都与处理代表物理项（图书、期刊、期刊文章和其他信息载体）的记录有关。而兼容性的重要性是在于这些记录之中和对这些记录的处理，而不在于所代表的物理项或对物理项的处理，尤其是在机器可读形式记录的情况下。

两个情报中心 A 和 B 假定都采用计算机处理，它们能够交换数据的程度取决于对某些问题的回答，诸如：程序是否允许 B 的计算机读出 A 的记录并以有效的方式处理 A 的记录；A 和 B 的记录能否合并；A 的程序能否在 B 的计算机上操作；A 的计算机能否和 B 的计算机通讯；A 的内部终端能否与 B 的计算机通讯？对这些问题做出肯定回答的程度，主要取决于这两个机构之间在数字编码、记录格式、硬件、软件和通讯协议上存在多少兼容性。

各种兼容性问题对情报服务的用户有直接的影响。假定一个用户希望通过一个计算机中心对几个可用的数据库进行联机检索，则相关的兼容性问题有：用户终端能否和计算机连通（包含通讯协议的兼容性）；所有数据库是否能用同一方法进行检索（包含记录格式和单一的命令语言之间的兼容

性)；检索策略能否从一个数据库转换到另一个数据库(包含数据库之间词汇表的兼容性)；所有的检索结果是否都能被用户理解(包含打印或显示的目录记录之间的兼容性)；多个输出是否能被合并，重复部分能否被删除(包含可识别重复条目的充分的兼容性)；检索输出能否由用户以机器可读形式存储起来(包含输出介质和用户存储介质的兼容性)？

假定用户希望检索3个不同计算中心的数据库，则还会引起这样的问题，即用户终端是否与这3台计算机都能通讯，以及能否使用单一的命令语言(要求3个中心采用同一标准或有某种转换语言)？兼容性问题也影响两个用户互相通信的能力(如用电子方法交换意见和交换信函)。

过去的20年中，在保证目录记录之间的兼容性方面取得了许多成就。允许目录记录交换的一个国际标准格式(ISO 2709)已被广泛采用。在这个标准的基础上，世界科学技术情报系统(UNISIST)编订的参考手册为文摘和索引服务机构及其他单位制订机读目录著录的形式和内容，提供了一套完整的国际标准化的工作手册。MARC互换格式和它的变型，包括UNIMARC，已被图书馆界广泛应用，在图书馆之间建立起水平很高的兼容性。新的通用交换格式(CCF)试图满足图书馆界和情报界的需要，它的推广采用一定是有希望的。

IFLA为各种不同类型的资料制订的国际标准目录著录规则，为在这个领域中达到兼容性做出了重大贡献，同时国际标准图书编号(ISBN)和国际标准连续出版物编号(ISSN)，使这些书目项能够单一识别，因而促进了不同数据库的互联。而字符转换的标准化条例，或使这些条例的更加兼容方面，

以及人名典据的管理，特别是在国际等级上的管理，则需要做出更大的努力。

对著作组成部分的加工，在保证兼容性方面所取得的成就是少得多的，特别是在处理期刊中的文章方面，这是非常遗憾的，因为期刊文章构成了现在出版的文献的大部分。目前甚至连单一的识别符也没有，这样就使识别机读数据库之间的重复引文大大地复杂了。在二次出版者之间和一次出版者之间，书目引文的习惯作法仍然有很大的差别，这样就给读者造成了许许多多的问题。而新的关于著作组成部分的国际标准目录著录能否对这一情况提供改善，还有待观察。

对保证标引语言兼容性的许多可行的方法都做过研究，包括词汇表的合并，建立中介词典，以及采用宏观词表和微观词表等。尽管按照 ISO 的词表结构标准可以达到结构上的兼容性，而除基本结构以外，词表之间的兼容性能否取得重大进展是值得怀疑的。术语管理在一般领域内保证兼容性已经取得了较大的进展(如INFOTERM的工作所证明的)，通过机器翻译技术进行多语种处理也取得了较大的进展。

情报系统的设计者和用户可以从其他应用领域内取得的软件和硬件兼容性的发展获得益处。为程序语言、命令语言、输入/输出装置、存储装置、字符集、通讯以及机对机连接所制订的标准，对于建立比较统一的组成部分是有帮助的，而这些组成部分构成了情报系统。不过，技术上的新发展往往比标准的发展快。例如，在情报系统中所引用的新技术——如录像磁盘，视频数据系统，以及地区网络——可能成为不相兼容的新原因。同时技术上的新发展，在很多情况下，使转换代替了标准化，成为达到兼容性的一种手段。当

没有标准存在或感到标准束缚性太大而不能满足本身要求时，则转换机能可使某一给定系统与另一系统达到兼容性，以便资源共享。

现在仍然存在许多问题。对取得进一步的成就来说，重大的障碍可归因于情报服务界各个部门的要求不同，造成这些不同的原因是由于民族和文化的不同，研究工作需求和大众需求的差异，以及所要求的目录详尽情况的变化，如一次出版者、二次出版者和图书馆提出的要求等。

标准化活动的“分阶段进行”也给标准化本身带来了问题。如：国际标准的出现，从某种意义上说，对比较先进的机构往往太晚，他们已经制订了自己的标准，如果完全采用新标准，则可能面临相当可观的转换费用。各种国际标准由于缺乏完整性，也造成了困难。

一般的说，在情报服务部门之间使兼容性进一步得到发展，将依靠情报界促进国际标准或协议的能力，使国际标准或协议有足够的权威性和灵活性，以取代目前指导情报界大量工作的许多国家标准、地区标准和部门标准。

情报服务部门采用国际标准的收益是很大的，发展中国家尤其是这样。他们非常需要共享他们的资源和有效地利用发达国家的资源。不过，情报服务部门必须警惕不要采用那种标准，即它们本来是为某些特殊用户制订的，而事实上却可能降低为这些用户服务的效果。例如，对某一专业情报中心用户提出的精确要求可能需要编订一套高度专业化的特定的词汇表，在这种情况下采用现有的比较通用的词表可能就不合适。

联合国教科文组织在促进情报传递循环的各部门——一次出版者、二次出版者、图书馆、各类情报中心——之间的

兼容性方面，在加紧地做更多的工作。因此，在整个传递循环中以互相配合和整体协调的方法，改善了文献加工的效率。此外，对保证电子出版领域的兼容性，现在就应给予注意。因为人类未来的信息沟通对电子形式的依赖要比对传统的印刷纸张为媒介的依赖似乎大得多。

兼容性并不总是依赖于标准，在某些情况下，通过工作规范和工作方法的协议也可达到兼容性。最后，联合国教科文组织是有能力在发达国家之间、发展中国家之间，甚至在仍然缺乏标准的那些活动中，促进兼容性。

目 录

序 言

第一章	前言	(1)
第一节	定义	(1)
第二节	情报传递渠道	(5)
第三节	情报服务	(9)
第四节	一次出版物	(20)
第五节	物理制品	(22)
第六节	本书讨论的范围	(22)
第二章	系统与标准	(24)
第三章	文献目录记录	(27)
第一节	文献著录规则	(28)
第二节	记录格式	(32)
第三节	向通用交换格式发展	(39)
第四节	数据交换的某些问题	(50)
第五节	参考文献著录规则	(56)
第六节	杂志名称	(61)
第七节	字符转换	(64)
第八节	典据文档	(67)
第九节	编档顺序	(75)
第十节	连续出版物的馆藏数据	(75)
第十一节	与目录记录有关的问题	(76)
第四章	检索语言	(78)
第一节	词汇表的兼容性	(81)

第二节	中介词典	(90)
第三节	综合词表	(95)
第四节	微观词表	(99)
第五节	宏观词表	(103)
第六节	算法研究	(109)
第七节	多语种处理	(111)
第八节	术语兼容性其他方面的问题	(117)
第九节	与检索语言有关的问题	(119)
第五章	软件	(125)
第一节	程序和程序设计语言	(126)
第二节	命令语言	(132)
第三节	数据库管理系统	(138)
第四节	机器辅助检索	(144)
第五节	用户系统接口	(150)
第六节	与软件有关的问题	(157)
第六章	硬件	(160)
第一节	计算机	(162)
第二节	输入/输出装置	(165)
第三节	信息存储装置	(169)
第四节	字符集	(174)
第五节	信息传输	(179)
第六节	计算机网络	(188)
第七节	机对机接口	(194)
第八节	与硬件有关的问题	(200)
第七章	数值数据库	(204)
第八章	管理工作的兼容问题	(212)
第九章	实例研究	(215)
第一节	前言	(215)

第二节	太平洋图书馆常务会议(SCOPAL)所属地区目 录中心.....	(215)
第三节	发展科学的国际情报系统 (DEV SIS)	(216)
第四节	毒物学联机情报数据库 (TOXLINE) 和人口 联机情报数据库(POPLINE): 关于数据库合 并的实例研究.....	(222)
第五节	目录数据的交换: 地质学方面的实例研究.....	(231)
第六节	CCL: 公共命令语言标准化的实例研究.....	(235)
第七节	iNet: 系统互联的实例研究.....	(242)
第八节	CSIN: 机器辅助检索的实例研究.....	(246)
第十章	结论	(255)
附 录	本书所用缩写词	(269)

第一章 前 言

第一节 定 义

“可兼容的”一词，如同英语通常的用法一样，涉及到人与人之间的一种关系。如果两个人可以“融洽地在一起生存”，他们就被认为是可以兼容的。兼容性与同一性并不是一回事。正如Shilling指出的，两个人可以是兼容的，但毫无疑问，他们并不是同一的。

在情报服务界，兼容性这个词的意义从它的更为普遍的用法上看，并没有很大改变。不同的系统、系统分支和系统活动，如果它们可以协调地在一起运行，（即以最小的工作量进行有效的交流和记录交换），那么它们就是兼容的。换句话说，兼容性的意思是指两个实体联合在一起工作的能力。俄文“Sovmestimost”（即Compatibility—兼容性）一词，实际上常常被定义为“可被结合的能力”。Henderson等人对兼容性曾有过更为正式的叙述：

当一个系统的处理结果可被另外具有相似的但不一定是等同系统的机构立即直接应用时，这些系统就被认为是可以兼容的。协作的机构，如果他们系统的产品可以互相交换使用，而不需要特殊的加工和转换，这些系统必定是兼容的。

世界科学技术情报系统(UNISIST)的可行性研究报告用略有不同的词句，给出了本质上相同的概念：

兼容性：是系统的一种特性，这类系统的产品可以互相交换使用，尽管在符号、结构、物理载体等方面有着差别，但不需要任何特殊的转换手段。

研究兼容性的主要原因是为了促进机构之间的合作与资源共享。Henderson等人对合作给出的定义是：

在为情报系统的资料进行的有效加工或在交换这种加工的产品方面，各个机构之间所做的共同努力或共享。

他们还提出了兼容性与转换性之间的区别。不同之点是达到变换或统一的难易程度有差别。如果两个机读数据库是按照同一规则进行数字编码并采用同一记录格式时，那么这两个数据库可认为是兼容的。也就是说，两者可用同一计算机软件操作，而不需要任何进一步的加工处理。如果每个数据库采用的是不同的记录格式，那么两者就不能用同一软件操作。这样就需要编写转换程序，以把一个数据库的格式转换为另一数据库的格式。这两个数据库是可以转换的，但它们并不是兼容的。可以认为转换性比兼容性更为普遍的一种性质。许多情报服务机构的输出可以从一种转换为另一种，也可以互相变换或混合，而实现这种操作的难易程度，则取决于它们输出的兼容的程度。

UNISIST的可行性研究对转换性给出的定义是：

系统的一种特性，即这类系统的产品通过“转换”程序能够达到互换。

Lebowitz专门谈到机读记录的格式问题，他对同一性、兼容性和等同性作了区别：

同一格式是这样的，即建立这种格式的规则是同一的，或虽然规则不同，但得到的文献目录记录没有区别。

等同格式的目录记录不是同一的，但以合适的计算机软件处

理后，可以制成同一的。

兼容格式的记录可以与其他可兼容格式的记录一同在几个系统中应用。

兼容性是允许不同交流格式的记录被统一成一个可用文档的特性。

从以上意义看，“等同的”与“可转换的”大体是同义的。但Lebowitz又指出等同性不一定可以进行逆转换。例如，一个机构采用专门词汇表所做的标引记录，可以自动地转换到第二个机构所采用的内容更为广泛的词表中去。但逆转换是不能自动进行的。举一个实际的例子，在词表A中的专指词“链霉素”，可以自动转换为词表B中的广义词“抗菌素”。但在词表B中的这一广义词下标引的文献，没有人的知识的介入（即再标引），是不能转换成词表A中所用的专指“抗菌素”的这个词的。

同一性是一种客观的和绝对的特性。两个实体或是同一的，或不是同一的。而兼容性和转换性则是比较主观的和相对的。我们可以认为它所指的是两个实体可以转换（或等同）的程度，或两个实体的兼容的程度。

显然兼容性与标准化是密切相关的。两个机构越是遵守共同的标准，它们的各种活动与输出也越可能达到兼容。如果两个情报中心所编订的词表都能严格地遵守国际标准化组织（ISO）的标准，那么这两种词表将比根据不同标准或根本不采用任何标准编订的词表，更加兼容，这就是一个例证。

本书所叙述的并非关于标准化本身的问题，而是关于影响情报传递工作效率的各种类型和等级的兼容性。然而，因为兼容性与标准化互相紧密地联系在一起，因此也不能完

全脱离标准化的问题。书中将涉及各种兼容性问题所用的标准，不过并不准备对所有标准做出说明*。

然而阐明兼容性和标准化之间的关系是完全必要的。几个机构采用共同的标准是使这些机构的活动达到兼容的最为明显的方法。但从兼容性的观点来看，并不是所有的标准都是有意义的，只是那些影响机构之间实物或数据交换的标准才有意义。举一个例子，缩微平片的尺寸标准是有兼容意义的。它影响到各个机构对缩微平片交换、存储和应用的效率。而与缩微平片保存期限有关的标准，则是各个机构自己权限范围的重要方面，与兼容性没什么关系，因为它并不影响机构之间交换缩微平片的能力。虽然如此，情况仍在迅速地变化，技术的发展不久会使原来兼容性并不重要的那些领域，变得重要起来。例如，在采用光学字符识别器自动读出全文之前，铅字字形的标准与兼容性的关系是很小的。

没有标准化也能达到兼容性。例如当一种记录格式用计算机程序转换为另一种格式时，可以通过某些过程或装置来协调不同的操作。这样看来，标准化是达到兼容性的明显的途径，但不是唯一的途径。

研究兼容性问题的主要原因是为了使系统或网络互相联接，根据Baker的说法，这种互联隐含着一种往复关系：一个系统既可以向其他系统输出，也可以由其他系统向自己输入（即可以给出记录，也可接受记录）。Baker认为兼容性对

* 情报传递过程的有关标准在《世界科技情报系统情报处理标准指南》(UNISIST Guide to Standards for Information Handling)一书中已有介绍，有关的ISO标准已单独出版。在制订文献目录方面实际应用的标准汇编也已由国际图书馆学会联合会(IFLA)出版。

互联来说是最为重要的，但仅兼容性本身并不能保证互联。

兼容性也与一致性有关，在本文中一致性所涉及的是数据格式的正规性和可判断性。某些类型的数据的一致性比其他数据更易于达到(也是更希望达到的)。因此希望图书馆或其他机构在确定作者姓名在记录中的表示形式时，应该保持一致性。如果所有机构都能采用同一标准(即编目条例)，这形式的一致性是不难达到的。而另一方面，在确定一种出版物的主题和选择表明主题事项的术语时，要达到一致性则困难的多。而且也可能不太希望达到这种一致性，因为对同一种出版物，不同的机构所感兴趣的原因也是很不相同的。标准对机构内部和机构之间的一致性都起促进作用。各个机构越是遵守标准，情报服务界也就越能达到一致性，一个系统和其他系统也就越能够兼容，而各种形式的资源共享也就更可能经济地、切实可行地存在下去。

第二节 情报传递渠道

研究与应用工作所产生的结果在科学和其他领域里传播的最重要的渠道，在图1中从宏观上做了描述。方框图的上部“用户界”包括两个部分：(1)研究和开发方面的工作人员；(2)应用和研制方面的工作人员。图1所表示的是情报的流通，即把研究、开发和应用活动产生的成果与经验，迅速有效地向需要情报并能从中获益的用户传播。

如图所示，用户界各方面的人员根据他们研究和开发的结果或他们在某些领域中的应用经验，提出报告。报告可以