

什么是数据管理系统

关于数据

数据是一个宽泛而复杂的概念。人们常常将数据简单定义为“信息的记录”。这种简化的定义实际上只是一次概念的转移。如果我们继续追问：信息又是什么？人们仍然很难给出清晰的答案。在本书中，我们并不试图给数据一个严格而准确的定义。这难以做到。我们暂且选择实用主义的立场，讨论数据到底是用来干什么的。我们发现，在人类世界里，数据通常有两种明确的功用。

数据的第一大功用是作为信息的载体。也就是说，它被用于记载正在发生的或曾经发生的事实，比如：张三身高175厘米体重75公斤，李四于2020年3月18日花费15元在淘宝上购买了一只高露洁全效牙膏，2020年1月20日14点35分在北纬37度东经15度的区域发生了6.5级地震，等等。这样的事实被记录下来，形成数据，便于我们传递信息、互通有无。信息的汇聚也有利于我们作出更好的决策。同时，被积累下来的数据成为历史记载，从中我们可以总结经验，发现新的知识。总之，数据被用于满足了人的信息需要，向人提供问题的答案。当今时代，数据的重要性越发凸显。依靠四通八达的IT基础设施，人们可以收集大量数据，又利用机器学习等技术从中发掘出各类规律，用于提升决策、生产和科学研究的效率。数据的加速汇聚导致了“大数据”产业的形成。数据成为价值连城的生产资料，需要得到有效管理。

数据的第二大功用是作为事实的凭据。例如，如何认定“张三的银行账户中有500万”这件事是否属实？稍微思考一下，我们会发现，这件事无法通过实物来判断。银行显然不会单独为张三准备一个保险箱，并且在里面装入500万现金。相反，我们只能依靠银行的账本或者张三的存折上的数据来判断这个事实。如今，账本和存折都已经数字化，数据被存放于计算机系统里的数据库中。再举一个例子，我们如何判断“霞飞路909号的洋房是属于李四的”？同理，所有权是抽象的概念，无法实物化。我们的判断仍然只能依靠数据。这项数据记录在李四的房产证上，同时也记录在房产管理机构的数据库中，并且通常以后者为准。除财产之外，一个人的众多属性通常都只能依靠数据来确定，比如他（她）的国籍、职位、资格等等。人类社会在自我管理中会用到很多形而上的概念，比如财产、组织、等级等。基于这些概念的事实认定通常只能依靠数据。这些事实数据是人类社会实现人、财、物管理的基础，需要被妥善保管。

对很多人而言，数据的第二种功用貌似只是第一种功用的特殊情况。比如，我们可以认为，作为第二种功用的数据也是为了记载信息，只不过记载的是权威信息。在笔者看来，没有必要在概念上做过多纠缠。在实际生活中，数据的这两种功用是明确不同的。如果作为第一种功用的数据丢失或损坏了，并不会影响人类世界的既有事实。相反，如果作为第二种功用的数据丢失或者损坏了，我们对事实的认定就变化了（比如财产就变更了）。之所以区分这两种功用，是因为它们对数据管理功能提出了不同的需求。

数据管理功能

无论作为信息的载体还是事实的凭据，数据都需要被好好管理起来。那么，如何才能做到数据的有效管理？我们可以将数据管理功能分为以下几个方面。

保护数据的持久性：首先，数据需要被妥善地存放起来。无论发生什么意外，都不应该让数据轻易丢失或损坏。事实上，没有任何一种存储介质或硬件设施能够做到万无一失，无论是纸张、硬盘、光盘。因此，我们常常需要对高价值的数据进行定期备份，减少数据因意外而丢失的概率。当数据有了多个备份，我们还需要让它们保持一致，避免发生错乱。

让数据可用：其次，数据需要被有效地组织起来，便于我们快速查找信息。以图书为例，当我们家里的藏书越来越多时，就不能再将它们随意地堆砌在书架上，而需要对它们进行分类或者索引，否则以后查找起来将会很

困难。对一座有百万藏书的图书馆而言，书籍的分类和索引就更加重要了，需要大量人力物力进行维护。数据必须组织起来才能供人使用。当数据越多、结构越复杂，人们对数据组织的要求就越高。

保证数据的真实性：当数据被用作事实的凭据时，我们需要保证数据如实地反映现实世界的变化。比如，当张三往银行里存入一笔钱后，我们必须保证他账户上的余额增加。当李四卖掉自己的房屋后，我们必须保证房产管理系统里的户主变更。让现实世界的变化分毫不差地落实在数据上，这是数据管理的另一项重要职责。

提供数据处理的平台：数据的使用过程是一个计算过程，这个过程可以在应用程序里进行，也可以在数据管理系统里进行。后者通常针对事先定义好的数据处理模式，比如对数据进行排序、累加等操作。将固定模式的数据处理程序移到数据管理系统中，并通过简单接口提供给应用程序，将大大减轻程序开发的负担。

保证信息安全：最后，数据中含有敏感信息。我们需要防止信息被泄露给没有权限的人，还需要防止数据被恶意损坏或篡改。这就需要一套有效的安全防护措施。

以上的各种数据管理功能并非独立存在。它们常常交织在一起。比如，信息安全需要数据组织的协助，以便区分什么人可以访问什么数据。又比如，数据的真实性和持久性往往是关联的，数据失真和数据损坏可以通过同样的机制来预防。这里的功能划分只是给读者一个直观印象，至于其中的具体定义和细节，本书后面再作详细介绍。

数据管理系统的目的

顾名思义，数据管理系统的目的当然是提供上述的数据管理功能。但这里涉及一个通用性的问题。可以想象，不同应用会使用不同的数据，对数据的组织方式和访问模式也都不同。如果要全方位满足应用对数据管理的需求，我们可能不得不为每一个应用单独构造一套数据管理系统。然而，如前文所述，数据管理系统是具备通用性的系统软件，不可能为某一个应用进行定制。相反，我们应该找到不同应用对数据管理功能的共同需求，将这些共同点抽象成为数据管理系统的功能，而将差异化的需求交给上层的应用程序去满足。也就是说，我们无法让数据管理系统包揽所有的数据管理功能，而需要在它和应用程序之间划一道界限。

事实证明，划出这道界限并不简单。如果我们将数据管理的功能尽可能推给数据管理系统去实现，我们可以减少应用程序开发的工作量。但这样做可能造成数据管理系统的臃肿，无法适应变化，也可能导致数据管理功能的接口过于复杂，以至于系统设计达不到模块化的标准。如果我们精简数据管理系统的功能，可以让系统变得更加通用、更加灵活，但可能会让应用程序开发的负担过重。这里似乎没有完美的选择，更多是需要一种折衷。我们后面会陆续介绍不同的系统都选择了什么样的折衷点。

总之，数据管理实际上是由应用程序和数据管理系统共同完成的。从实用的角度考虑，数据管理系统的目的是以最好的方式协助应用程序实现数据管理功能，而并不是承包整个数据管理的功能。某种意义上，数据管理系统只是一个工具。就像在制造家具时，我们并不指望有一台全能的机器，只要告诉它我们需要什么家具，它就能自动造出来。我们通常使用的是锯子、尺子、凿子这样的简单工具，辅以灵活的双手，就可以构造出任何我们需要的东西。这些工具经历了时间的洗礼，被证明是最有效的方式。同理，在构建应用的数据管理功能时，软件工程师并不指望一套全能的数据管理系统，他们更在乎这个系统用起来是否顺手。这可能是本书提出的一个独特观点。我们在后面的内容中会进一步阐述这个观点的真实性。

[上一页<< | >> 下一页](#)