

文章编号：1007—5399 (2019) 05—0033—03

数据集市推进邮储银行业务数据分析的思考

刘万里

(中国邮政储蓄银行, 北京 100808)

摘 要：阐述了邮储银行业务部门数据分析工作的现状，分析了业务部门数据分析存在困难的原因，提出了业务部门加快数据分析工作的建议。

关键词：数据集市；监管报送；内部决策；客户营销

中图分类号：F61 **文献标识码：**A

数据分析难、分析慢、分析不准确是中国邮政储蓄银行（以下简称“邮储银行”）各级业务部门数据分析工作面临的三大问题。而在《中华人民共和国网络安全法》《中华人民共和国电子商务法》付诸实施、“业务协同、客户共享”等邮银协同发展的背景下，如何满足监管要求、保障消费者权益，并平衡好大数据精准营销、敏捷型业务合作伙伴关系的建设，是银行、速递、保险等集团各下属单位面临的共同问题。本文从数据分析的应用场景出发，尝试分析、解答邮储银行业务部门数据分析工作合规拓展、邮银协同的可能框架，为建立邮储银行敏捷型数据分析工作机制、邮银互通的数据分析路径提供可行性方案。

1 邮储银行业务部门数据分析的开展情况

数据分析是指从既有的数据信息中提炼出有价值的的结果。基于邮储银行总行某业务部门2016年1月至2018年6月末的数据分析，整理得到了该条线数据分析工作的主要应用场景，见表1。从统计结果看，该业务部门数据分析需求共计75份，其中监管报送类数据分析4份、内部决策类19份、客户营销类37份，不便分类的需求有15份，业务部门客户营销类数据分析需求呈绝对优势，符合业务部门的职能定位。从数据分析的时长看：监管报送类数据分析需求以监管截止日期为准，一般每例数据分析的准备时间不超过1周；内部决策类数据分析需求，基本在部门领导提出后48小时内提交，如果数据结果可用，需求会固化成按月提取的报表继续推进；客户营销类数据分析需求，在没有数据集市或大数据平台数据实验室的情况下，从业务部门提出需求到获取数据结果，单例需求的平均完成周期最少是2周。

表1 数据分析分布情况统计表

类别	数量	举例	实现方式	准备时间(例)	备注
监管报送	4	农村电子银行	数据提取	1周	周期
内部决策	19	快捷支付转化率	数据分析	48小时；月	急切
客户营销	37	交叉营销项目评估	数据分析系统对接	2周	闭环使用
其他	15	系统建设	数据提取	——	数据库

2 业务部门数据分析工作存在困难的原因

2.1 监管报送类数据分析困难原因

监管报送类数据分析，一般都是口径相对固定的长期数据报送需求，重点和难点都在报送口径的一致性。尽管是长期性报送，但此类数据在缺乏全行型数据集市的情况下，更难保证数据的准确性、时序数据的可比性。数据分析难、慢和不准确的原因大致包括两个方面。

一是报送数据往往因前台业务系统的变动而调整。在完整的“数据仓库-数据集市-监管报送系统”体系下，业务源系统在新增、变更或者关闭相关产品或服务后，后线系统通过数据抽取、清洗和加载等数据生产的完整过程，能够记录、传递业务变更对应的数据调整内容。但是由于缺乏数据集市这一中间环节，后台报送部门无法知悉以上产品或服务的变动事实，同时中台数据提取部门也无法掌握数据调整的具体逻辑，这样就不得不“按件”重复数据抽取、清洗和入库。据此逻辑，在全行性数据集市缺失的情况下，越是交易创建早、业务发展快的部门，报送数据的准确性就越差、可比性就越低。

二是发生数据错误后补救成本太高。根据经验，业务调整后两个季度内的报送数据误差不大，但在业务系统正式运营半年后，年度同比数据的偏差就会凸显。而此时，限于向监管重新报送、替换数据的可能性基本不存在，相关部门不得不“调整”数据，再根据业务发展趋势择机解决数据报送不准确的问题。然而，在数据集市基础上建立的监管报送系统，往往能及时捕捉、修正源系统变更造成的指标变动，最终提供准确、可比较的基础数据指标，满足监管报送需求。

2.2 内部决策型数据分析困难原因

内部决策型数据分析，往往是部门领导关注的方向性、探索性的数据分析需求。特点是需求不明确、数据要求频次不固定，难点是分析相近部门关联指标的可比性高低。从业务部门看，数据分析的难、慢和不准确问题主要集中在内部决策型的数据分析需求工作上。

一是现有的内部决策型数据分析模式不具备可持续性。

一方面,内部决策型数据分析自身要求敏捷开发的分析模式。相较于监管报送、客户营销类的数据分析需求可以通过报表系统、系统对接得以固化的特点,部门领导内部决策型数据提取需求不仅单次数据分析耗时较长,还会随业务发展而长期存在。另一方面,全行业务部门普遍缺乏数据分析专岗,业务部门与软件研发中心设想的“业务需求—数据提取”前后线合作模式并不流畅。在行方技术人员缺乏的情况下,前后线两方合作往往演变成“行方业务人员—业务系统厂商—软件研发行方技术人员—软件研发中心外包技术人员”的四方邀约会谈机制。其中,业务系统厂商、软件研发中心外包技术人员更换频繁,降低了四方会谈的稳定性。总体来看,业务部门必须自设数据分析专员,掌握系统间的数据映射关系,支撑部门内部决策数据分析工作。

二是数据集市的缺失延长了数据分析时长,降低了数据准确性和可比性。一方面,邮储银行目前各业务部门普遍缺乏一个数据加工分析的中间层——数据集市,在源系统更新后的大数据平台及后线系统缺乏实时动态更新的连锁机制。在缺乏中间层的情况下,一次生产系统升级等同于一次数据“伤害”。而在数据管理部门缺位的情况下,要想实现源系统变更后的大数据平台和后线系统的数据同步,基本不具备可行性。另一方面,从数据分析人员的操作来看,在缺乏数据集市作为数据调整中间层的情况下,为了尽可能保证数据的准确性,每一次数据分析都被迫拉长为“数据采集—数据清洗—数据分析—数据监测”的全业务流程。尽管完整的数据分析工作在数据集市完备时同样如上执行,但是不存在跨业务部门数据集市的情况,每一次数据分析意味着和关联系统的运维或厂商重复一遍,浪费了前后线系统过多的精力和时间。最后,考虑到银行类金融机构科技转型的大背景,业务发展转型阶段的更新上线频次将不断增多,在缺乏数据集市的情况下,历史数据的绝对可比较基本不存在。这在一定程度上挫伤了部门领导对内部决策型数据分析的信心,增加了数据核校的工作量。

2.3 客户营销类数据分析困难原因

客户营销类数据分析,多服务于条线业务拓展、客户关系维系,特点是周期性强,难点是数据实时性高,底线是不引起投诉。当前邮储银行此类数据分析困难的原因主要包括以下两个方面。

一是交易信息的不记录导致客户行为分析困难。相较于逻辑集中等核心系统对客户动账类交易信息的记录,手机银行、个人网银等电子渠道对账务类交易的记录是不完整的,而嵌入在以上电子渠道的业务源系统,往往对非动账类交易数据的记录缺失,线上营销客户的行为分析只能从各个源系统追踪客户交易渠道的偏好、产品持有结果的时序变化,然后归纳出客户持有产品的行为偏好,而无法探究客户交易结果的产生逻辑。而后者,才是营销活动场景建设的源泉。此外,一旦嵌入在电子渠道的源系统不记录交易发生时的所属渠道,客户的线上交易行为就会漏记,客户交易偏好的度量就有偏差,最终在线上资源的配置上就会遗漏目标客群数

量,增加客户流失的可能性。

二是缺乏综合性客户营销类数据集市,导致精准营销数据不准确。目前,邮储银行线下网点普遍配备了综合客户各渠道联系方式的个人营销系统,基本保证了客户联系方式的准确性。而线上渠道尚未投产以客户为中心的全行性综合营销数据集市,相关业务部门如果不能通过全渠道整合营销数据,往往只能提取到客户首次和邮储银行某个部门建立合作关系时的预留信息,一旦客户变更了联系方式,客户失联的可能性较大。而在手机号码自然流转的背景下,一次“脱靶”的营销往往会带来一笔投诉。这在《中华人民共和国网络安全法》执行的背景下,又可能会引发公司治理有效性的监管处罚。

3 业务部门加快数据分析建议

3.1 建立事业部型的数据集市

建成大数据平台后,为何还要建设数据集市,数据集市的建设是否为重复性建设,数据仓库、数据集市的逻辑关系有助于解答这两个问题。数据仓库和数据集市都普遍采用ETL技术,并且综合考虑了数据来源、交易时间和存储性能的数据集合。二者最大的不同在于建设目的的差异,数据集市是以迅速响应业务需求为目的,而数据仓库的目的是尽可能完整、正确地存储业务数据。具体到邮储银行,从系统架构看,数据集市是连接生产源系统和业务系统的中间层系统,数据仓库是接收生产源系统的终端系统。从支撑业务发展看,数据集市可以直连监管报送、统计报表和业务营销系统,为业务发展提供准确、实时的清洗后数据,而数据仓库则可以直连内容管理平台、运维系统,为还原客户交易场景、解答客户查询需求提供在线支持。从部门职责看,数据集市的管理单位是软件研发中心,而数据仓库的管理单位则是数据中心。

未来数据集市的建设建议考虑跨部门的事事业部型数据集市,即包含业务源系统、间接系统(大数据平台)的复杂来源型数据集市。只有更多业务系统的高频数据实时同步,事业部型数据集市才能获取最新、更高频的客户联系方式和交易偏好,才能有利于精准营销数据的产生,服务于以客户为中心的零售银行建设。

3.2 在业务部门设立数据分析专岗

根据数据分析需求分类,监管报送类以最近一次报送口径为准,客户营销类以目标客户的即时联络为主,这两项数据分析需求需要固化成报表系统或实时调用的系统间接口文档。从数据分析的供给侧看,提高数据分析人员稳定性和业务经验,是解决此类数据分析需求的重要抓手。在总行层面设置具有绩效考核标准的数据治理管理一级部,加强数据分析、数据治理工作的管理;增加各级机构技术人员的职务层级,降低数据分析人员的流失率;在各级机构各业务部门增设数据分析专岗,提高数据分析专岗人员的绝对数和相对比例;在现有软件研发中心业务对接处室实行“业务—科技部门”人员双向流动机制或业务条线特派员机制,盘活存量数

据分析人力资本，提高业务科技双方的数据分析能力。

3.3 保证业务部门数据分析人员的数据可达性

只有触达数据、了解数据和使用数据，才能掌握数据、运用数据和管理数据。目前，软件研发中心在总行、部分分行试点了大数据平台数据实验室，尽管存在相关系统的数据没有接入、部分数据分析软件配置尚不到位等问题，但在数据可达性方面，大数据平台数据实验室减少了近八成的数据导入工作量，为各级机构业务部门数据分析的开展提供了入口，为数据分析工作提供了便利。建议总行、一级分行相关业务部门尽快申请数据分析权限，从统计报表、绩效考核等部门核心指标的计算开始，掌握数据分析的基础知识。

3.4 构建数据全生命周期的管理模式

如果各级机构业务部门数据分析专员熟悉数据、开展数据分析，是自上而下拆解任务指标，那么，通过系统对接下发营销数据、跟踪数据营销结果，则是打造数据分析闭环、支撑分行发展的又一项重要内容。具体到总行、一级分行的业务部门，可以项目建设为驱动，搭建“大数据平台数据实验室—大数据平台—数据集市—业务系统—大数据平台”的数据采集、数据模型训练、模型布设、数据整合、数据/标签输出、营销结果回传的“前店后场+营销全流程”合作模式，即数据实验室抽取关联系统基础数据、训练基础模型，在大数据平台布设成熟模型，通过接口文件输出给相关数据集市或直连业务系统，最后从业务系统同步营销结果至大数据平台，进而更新数据实验室的样本集，为下一轮数据营销、数据治理提供基础数据。

3.5 数据脱敏、随机抽样保证内部决策型数据分析的安全

转化成数据分析的要点就是，尽可能减少可追踪的个人真实数据的暴露。尽管内外部数据分析人员签署保密协议、尽职承诺，但是数据分析人员使用原始生产数据分析数据，难免有数据泄漏、丢失的可能，加上金融科技的发展，行内业务人员掌握数据分析、业务经验之后，内外勾结等合规风险就会增加。打破恶性循环的一个切入点就是建立大数据平台数据脱敏入库+随机抽样代表性样本库。具体而言，一是监管报送、客户营销类数据分析需求和内部决策型数据分析分割，内部决策型数据分析使用的基础数据是脱敏后的源系统数据，避免数据分析人员在数据分析过程中从事非法获利行为，监管报送、客户营销类数据分析直接使用未脱敏的源表数据。二是内部决策类数据分析不能使用全库数据，只能使用随机抽取的代表性样本数据。在提供全样本的情况下，数据分析人员仍可通过异常值检验等数据分析手段反向推定或获取脱敏前的数据样本真实信息。而随机抽样可以保证对整体数据（客群）的代表性，并且避免了脱敏失效的隐患。

3.6 总分行共用数据实验室，加强数据共享工作

数据分析不能缺少总行的战略引导，但也不能忽略区域性客户的结构性差异。在有限人力的基础上，数据分析工作大致可以分为总行、一级分行、二级分行及以下三个层级。其中，数据分析的智力供给重点是总行和一级分行两级机构，二

芬兰邮政关闭自营网点 拓展合作服务网点

芬兰邮政计划今明两年新增700多个合作服务点，同时关闭18家自营网点，预计裁员108人。

随着客户需求的变化，芬兰邮政对零售网络进行现代化改造和扩张，使公司在人口稀少的芬兰拥有2 800个服务点和包裹柜。

客户一直希望邮政服务点能够延长营业时间，并设在其他邮政服务设施附近。芬兰邮政对此解释道，邮政服务点和包裹柜是由其他合作伙伴运营的，因此都设在服务供应商的营业场所内，这些服务供应商在晚上与周末都正常营业。

经谈判双方协商，芬兰邮政会为18个计划关闭的邮政服务点寻找外部合作伙伴或是采取另一种服务模式，同时在2021年3月前逐步关闭这些自营网点，新的本地合作伙伴将提供与邮政服务点相同的服务。

未来，芬兰邮政自营的6个网点将设在赫尔辛基、埃斯波、于韦斯屈莱和罗瓦涅米，其余97%的零售网点将由合作伙伴运营。

芬兰邮政零售网络总监表示，越来越多的客户倾向于使用包裹柜，促使芬兰邮政的服务点数量翻番，芬兰邮政与客户的距离也更加紧密。

除了客户需求变化和营业时间诉求外，此次谈判的目的还包括节约成本。由于信件和出版物业务量急剧下降，芬兰邮政需要在未来三年内至少削减1.5亿~2亿欧元的成本。另外，处于中心位置的自营网点成本过高，而延长营业时间也不符合成本效益原则，因此公司决定关闭这些自营网点。

（王蒙 译）

级分行及以下机构则以数据采集、数据反馈和客户联系为主。根据线上客户交易习惯，总分行机构数据分析工作的开展应以全渠道交易行为分析为主，二者职责边界只是客户归属的区域差别，在数据使用、模型优化方面应共用共享。

参 考 文 献

- 1 叶剑陶. 农村商业银行绩效数据集市的构建及挖掘应用. 华南理工大学, 2017
- 2 曹培, 马瑞全. “数据集市”助邮政大数据应用能力升级. 中国邮政, 2017, 8

收稿日期: 2019-01-31

作者简介: 刘万里(1988~), 男, 河南唐河人, 博士, 主要从事数据分析、风险管理研究。