**雅培项目调研**

## 概述

雅培项目使用工序化平台进行数据处理，满足客户获得渠道数据的需求。

## 服务范围

1. 数据采集
2. DDI采集：渠道内商业采购、销售与库存数据，以数据直连（DDI）的方式获取。
3. 人工采集：无法采取DDI方式获取数据的商业，倍通以人工采集方式收集其采购、销售与库存数据。
4. 格式整理

对于人工采集的数据，将数据文件格式按照标准模板进行整理。

1. 数据标准化处理

标准化后的商业采购、销售及库存数据，标准化处理内容包括：商业名称、产品名称、终端名称、批号、数量单位。

1. 主数据服务
2. 处理客户发起的新增、变更，包括更名、合并的核实、确认与修正。
3. 流向新增终端，倍通新建，并清洗出唯一编码、标准名称（去特殊字符）、省份、城市、属性字段。
4. 日交付

每日DDI标准化处理后采购、销售、库存数据的提供（目前已取消）。

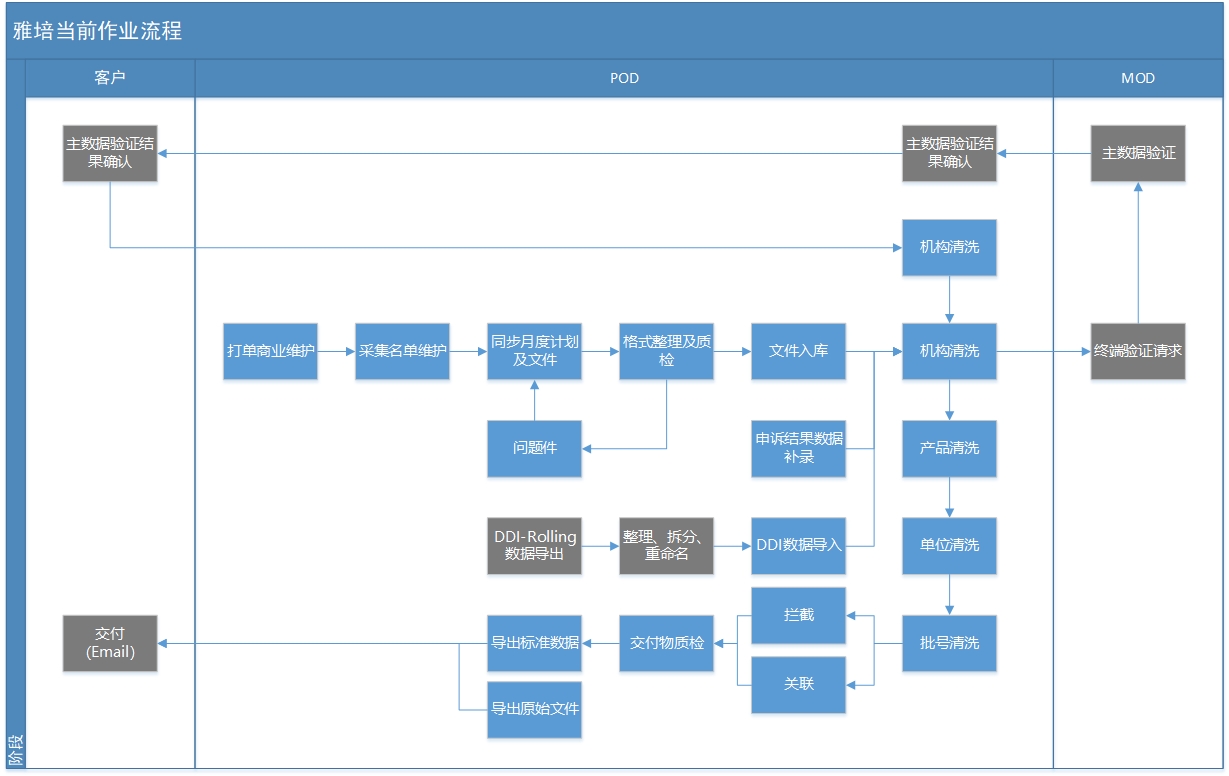
1. 月交付

每月W13，交付上月全量商业的进销存数据，同时交付月原始流向文件。

1. 申诉

对于来自商务或销售团队，对于数据的疑问，进行与商业的核实与修正。

## 项目流程



#### 3.1项目流程概述

1. 初始化数据
2. 数据采集
3. 流向入库
4. 自动清洗数据
5. 手动清洗数据
6. 数据交付
7. 数据申诉

## 初始化数据

1. 配置基础数据
2. 地理信息
3. 一二级属性
4. 经销商信息
5. 导入客户提供数据
6. 机构主数据
7. 产品主数据
8. 机构映射关系
9. 产品映射关系

### 数据采集

1. 打单商业
   1. 数据内容：商业信息，并配置商业是否返回DDI日数据。
   2. 数据维护：

每月从委托方得到经销商主数据，其中新增的经销商由主数据部门清洗出PHcode和PHname，根据系统现有的机构信息，补充相关信息，如省市区或名称。使用【打单商业】导入系统。

对导入数据进行规则校验，规则如下：

1. 校验重复数据：在进行“全量，增量，合并，删除，导入改码”时，“客户经销商编码”不能存在两条相同值的数据。
2. 校验重复数据：在进行“全量，增量，导入改码”时“经销商名称”不能存在两条相同值的数据。
3. 校验重复数据：在进行“增量”时，本次导入的“经销商名称”不能与系统中已存在的经销商名称相同。
4. 校验地理信息错误：商业所属的省份、城市信息不能为空，区县可以为空。
5. 校验地理信息错误：省份、城市信息必须为标准地理信息中的值。比如广东必须为“广东省”不能为“广东”。
6. 校验地理信息错误：所属城市必须是所在省份内。
7. 经销商级别校验：级别不能为空，并且需要是这几项其中之一：('T1','T1B','T2','其他')。
8. 经销商名称不存在校验：当进行“导入改码”操作时，导入数据的经销商名称必须已存在系统中。
9. 采集名单
10. 数据内容：商业信息+采集方式（包括人工、DDI、补录）+执行年月。
11. 数据维护：

使用【客户数据管理】导入采集名单。雅培每个月倒数第二第三个工作日给出采集名单，由主数据部门清洗出PHcode和PHname。

项目组判断哪些是“人工”哪些是“DDI”采集，哪些是“补录”数据。

导入数据校验：

1. 导入数据中，不能出现重复的列名。
2. 导入sheet的名字要与ClientDataManage.DataItem和dbo.ExecuteLine中的名字一致。
3. 导入数据的列名要与[DBO].[ExecuteLine]的列名完全一致。
4. 导入数据的记录条数要>0条。
5. 导入数据中的“项目编码,经销商编码,执行年份,执行月份,采集类型”五列不能同时具有相同值。
6. 导入数据中的“经销商编码”与“经销商PHCode”需要与系统中的经销商信息相符。
7. “采集类型”只能为（'人工','补录','DDI'）其中之一。
8. 采集计划
9. 生成采集计划

将采集名单中“补录”与“DDI”的数据生成对应的采集计划。

1. 同步采集计划

判断采集方式为月度采集的计划从月度系统中同步到工序化平台。

系统校验与异常处理：

1. InputItem表中是否已经完成了对DataItem的引用配置。
2. [dbo].[CollectingPlan]采集计划表中完成对DataItem的引用配置。
3. 月度回传的数据中的列必须与[DBO].[CollectingPlan]采集计划的列配置一致。
4. 导入数据的记录条数要>0条
5. 导入数据中的经销商PHCode于“打单商业”中的值相符。
6. 导入数据中的经销商PHCode于“采集名单”中的值相符。
7. 如果不存在则计划与流向都无法进入。需要线下沟通再次调整“打单商业”和“采集名单”。
8. 如果月度系统流向已上传，但采集名单中没有维护，则需要补充维护采集名单后联系IT协助解决。

### 6. 流向入库

#### 6.1日流向入库

*选择开始、截止日期从DDI数据源获取数据（当前使用Rolling数据源）。（目前已取消）*

#### 6.2 月流向入库

月流向包含人工、DDI、补录三种类型。

1. 项目流向模板设定
2. DDI数据导入模板

根据委托方的交付物信息，映射关系和流向采集信息，制定导入模板文件。并根据业务数据含义，指定系统中的对应标准列名。进销存流向模板分三个不同的EXCEL,但后续下载时会合并在一个文件中。

1. 手工数据导入模板

因为数据来源不同，模板与DDI模板不同，制作过程类似。

1. 人工采集数据通过格式整理方式入库
2. 从月度系统获取人工采集流向文件并存储。向入库队列中插入记录，作为格式调整任务。
3. 自动格式整理
4. 读取格式整理模板

第一优先级：该项目、该经销商历史模板，最近一个成功的月份。

第二优先级：其他项目该经销商历史模板。

1. 格式处理规则

最优先使用by经销商by月份级的定制规则。

第二优先级：使用项目级定制规则，在【项目信息】手工导入模板中配置。

第三优先级：使用系统级默认规则：先用上月的规则来校验（sheet名称一致），如果校验不通过，则使用更早积累的历史数据调整结果。

1. 数据校验

原始列与目标列一一对应。

所有原始列都匹配上目标列。

对应经销商级别交付模板中必要目标列都存在对应的原始列。

数据非空、长度、格式。

1. 记录状态

自动格式整理结果写入数据库。

1. 分配任务

分配策略：格式调整作业任务阈值、格式调整质检任务阀值、商业格式调整作业员、商业格式调整质检员。

1. 人工格式整理

先进行原始文件下载、进行格式判断，如果格式符合要求则进入格式调整环节；如果格式不符合要求，则线下调整（解决没有表头加表头，进销存一张表中，串行等）完毕后进行文件上传。选择该文件进入格式整理页面。

格式整理内容：

1. 创建标准列与原始列的对应关系
2. 设定起始行
3. 保存模板
4. 导入类型选择：全量、增量
5. 流向类型选择：原始流向作业、申诉作业
6. 下载文件
7. 打开原始文件
8. 状态设定：本月无量、无法返回、等待返回、忽略
9. 提交质检：系统验证（不漏字段、不可为空字段）
10. 纸质格式整理

下载文件，进行格式调整，上传文件后，进入【格式整理】环节。

1. 质检过程
2. 原始文件下载。
3. 按照用下载的原始文件与目前系统调整完的结果进行质检。人工质检规则，（是否窜列，条数，原始数量与调整后数量是否一致）。
4. 如果质检通过则“质检通过”。
5. 如果存在问题则“质检退回”。
6. 质检通过（入库）。
7. 问题件处理

问题件：如果发现原始数据存在，则由格式整理人员标记为问题件。

采集人员与经销商确认问题原因，如果规则类问题以经销商反馈后进行格式调整；如果是原始数据问题，需经销商提供二版流向。再次进入格式调整环节。

1. DDI月数据通过手动导入方式入库
2. 下载模板

选择模板文件下载后，by经销商by业务类型填入业务数据并导入上传。

1. 导入文件

规则：每次只能传一种业务类型,且导入文件名必须是|经销商编码|下划线|采销存缩写(PSI)，sheet页必须是中文：销售、采购、库存，且只支持单sheet页批量上传。

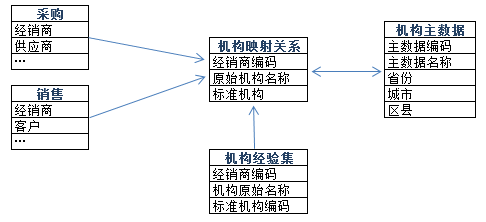
1. 入库自动质检项设定
2. 销售数量：不为空，忽略“0”为数据
3. 客户名称：不为空
4. 产品名称：不为空
5. 批号：通过(;:)字符分割提取
6. 销售日期：不为空，日期范围校验12个月以内。
7. 补录数据通过手动导入方式入库

导入文件类型选择补录，其他过程同DDI月数据入库。

## 7.自动清洗数据

### 7.1机构清洗

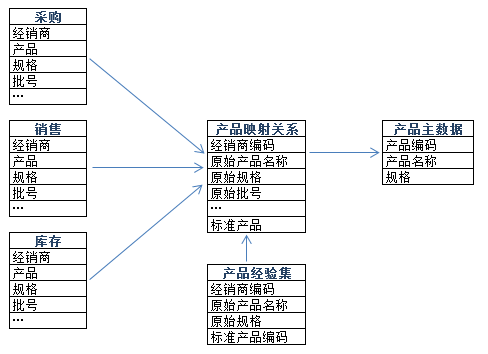
1. 机构Mapping关系



1. 机构Mapping步骤
2. 根据流向文件，提取映射关系关键字段MatchCode（经销商编码+原始名称）。
3. 查询历史映射关系表，如MatchCode存在，则复用该映射关系，使用该映射关系中关联的主数据
4. 如MatchCode不存在，则新增映射关系，查询经验集，使用经验集自动匹配
5. 如经验集不存在，则新增映射关系，手动Mapping，根据主数据建立映射关系
6. 规则逻辑
7. 用原始名称查询主数据是否存在，如不存在，则按业务流程新建主数据
8. 主数据变更，触发机构映射关系更新
9. 经验集变更，触发机构映射关系更新

### 7.2产品清洗

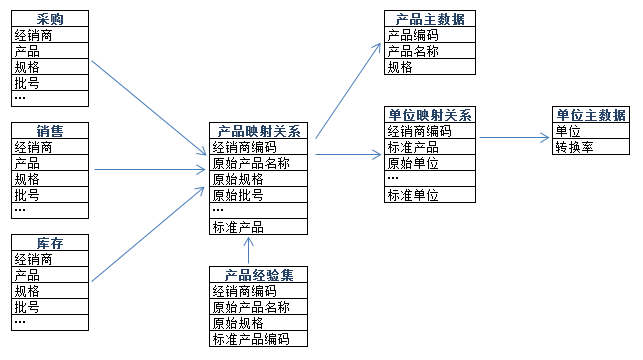
1. 产品Mapping关系



1. 产品Mapping步骤
2. 根据流向文件中经销商编码+产品名称+规格+生产厂家，提取映射关系关键字段MatchCode
3. 查询历史映射关系表，如MatchCode存在，则复用该映射关系，使用该映射关系中关联的主数据
4. 如MatchCode不存在，则新增映射关系，查询经验集，使用经验集自动匹配
5. 如经验集不存在，则新增映射关系，手动Mapping，根据主数据建立映射关系
6. 规则逻辑
7. 用原始名称查询主数据是否存在，如不存在，则按流程新建主数据
8. 经验集变更，触发产品映射关系更新

### 7.3单位清洗

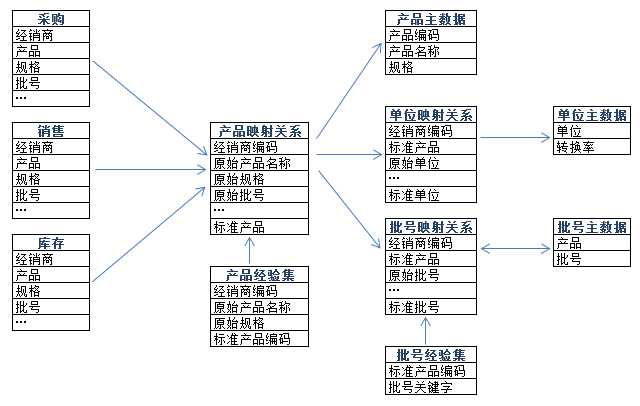
1. 单位Mapping关系



1. 单位Mapping步骤
2. 根据新建产品映射关系生成新的单位映射关系（经销商编码+产品编码+原始单位）
3. 根据主数据建立映射关系
4. 规则逻辑
5. 用原始名称查询主数据是否存在，如不存在，则按流程新建主数据
6. 包含单位倍率转换
7. 产品匹配更新，触发单位映射关系更新

### 7.4批号清洗

1. 批号Mapping关系



1. 批号Mapping步骤
2. 根据新建产品映射关系生成新的批号映射关系（经销商编码+产品编码+原始批号）
3. 查询经验集，使用经验集自动匹配
4. 如经验集不存在，则根据主数据建立映射关系
5. 规则逻辑
6. 用原始名称查询主数据是否存在，如不存在，则按流程新建主数据
7. 产品匹配更新，触发批号映射关系更新
8. 批号经验集变更，触发批号映射关系更新

## 8.手动清洗数据

1. 建立映射关系

更新该映射关系中的标准机构编码，为人工选择的机构编码。

1. 设置黑名单

跟此映射关系相关流向不会交付。

1. 设置问题

提交管理人员核对，对应的经销商数据不会交付。

1. 设置忽略

无法找到目标，作为待处理状态。

## 9.数据交付

1. 项目数据加工（关联表）
2. 商业渠道表
3. 医务架构表
4. 医务架构特殊关联表
5. 零售架构关联表
6. 数据发布
7. 发布月标准流向
8. 数据校验
9. 问题流向文件不交付
10. 非正常状态不交付
11. 被删除流向不交付
12. 非当月或返回品不交付
13. 有问题整个文件不交付
14. 文件打包

生成文件（命名、格式）

1. 文件交付

人工交付

## 10.数据申诉

从月度系统下载申诉数据，进行格式整理，再导入SDP进行后续作业。

备注：

雅培项目现在不需交付日标准流向数据，只需交付DDI原始流向数据，通过从rolling系统导出交付。

将rolling系统导出的DDI月原始流向数据手动导入工序化平台，经数据标准化处理后包含在月标准流向中交付。