在iUAP DI中利用NC数据字典进行数据输入

在数据仓库项目实施过程中，通常需要从NC生产库中抽取数据在数据仓库中，从NC数据库输入数据时，表的数量庞大且每个表的字段众多，通常无法直观的知道表和字段的确切含义。在实际应用中通常数据仓库的UAP Home和NC的UAP Home分别独立部署，使用语义元数据也无法解决该问题。在iUAP DI 1.2.4以后的版本中可以使用NC的数据字典进行数据输入，本文下面将详细介绍。

NC的数据字典的结构原理如下图：



其中NC数据字典服务在NC636以后的版本中自然包含该服务，在NC636以前的版本中需要打一个数据字典服务的补丁，才能使用该服务。

打开iUAP DI Designer，在左下角的资源区域，可以看到数据字典选项卡，如下图：



在空白处右键，在菜单中选择“新建”，弹出如下图的输入对话框：



输入名称和地址，地址格式是：*<协议>://<IP|机器名|域名>:<端口>*，如果是使用的http协议可以省略http://，如果使用的协议默认端口也可以省略端口。填写好地址后点击加载按钮，可以在帐套下拉列表中看到设置的帐套，如果出现错误，可能在指定的地址上没有数据字典服务，需要确认地址是否正确和数据字典服务是否可用。选择一个帐套后点击确定即可建立一个数据字典服务链接。新建一个数据字典的同时，iUAP DI还会创建一个数据连接，D把新创建数据连接和数据字典关联起来。

新建完成后，展开该数据字典可看到里面的内容，如下图所示：



可以把一张数据字典表直接拖拽到转换面板上，这样会在该转换中会自动新建一个关于该表数据字典输入，也可以在工具箱中的输入抽屉中选择数据字典，这样会创建一个空的数据字典输入，双击打开该转换步骤设置界面如下图所示：



数据连接是NC数据存储的数据库连接，拖入输入字典会自动选择，也可以选择其他的数据连接。点击显示名后浏览按钮，弹出一个字典选择对话框，可以选择其他的数据字典表，表名是选择数据字典关联的表的名称，随显示名带出，不可编辑。选择列中默认是所有的的列，也可以删除不需要的列，如果误删点击重新获取，可再次获取所有的列。过滤条件可对数据进行简单过滤，多个过滤条件是与的关系。排序是数据按照指定的字段排序。设置好后，即可点击右上角的预览进行数据预览。

如果需要多个表连接，可以通过”记录连接”对两个输入流进行连接，在下面的例子中是对两个数据字典的连接，总览图如下：



分析主题定义输入设置如下：



分析主题分类设置如下：



记录连接设置如下：



这样两个数据流通过一个字段和多个字段进行数据的连接，与SQL的join有同等的作用。

总之，iUAP DI的数据字典输入提供了中文语义，方便了数据从NC中载入数据，多个表的连接可以使用记录连接实现。目前iUAP DI仅支持NC的数据字典，在iUAP DI后续版本中，会根据需要会加入U8等其他产品的数据字典。