## **实验03、数据治理相关技术**

实验目的

1. 学会区块链与智能合约的相关代码的编写与部署
2. 了解数字对象架构组成及数据模型
3. 能运用所学知识对具体的数字对象进行封装
4. 能利用智能合约来完成数据治理中的相关任务

场景描述

为满足医疗商保理赔场景下各方对数据的需求，市政府及数据中心决定牵头建立一个数据交易平台，该平台能够支持各方进行数据上传、数据交易、数据溯源等功能。

由于早年间信息系统不规范，某医院档案中心存在大量纸质发票。如今为了进行数字化管理，并便于数据价值挖掘，该医院希望能将纸质发票进行数字化处理，并作为在进行数据脱敏等隐私保护技术处理后，将其作为数据资源提供到数据交易平台中。

实验数据

本实验的发票信息节选[飞桨医疗发票数据集](https://aistudio.baidu.com/datasetdetail/168112/0)中的部分信息，见文件，医疗发票示例如下：



实验要求

1. 根据目前医院的发票图片的信息，设计门诊发票数字对象模型，并选择至少五个发票图片，按照自己定义的格式进行封装。具体封装要求如下：

* 标识符分配：为每张发票分配一个符合DO-IRP协议要求的数字对象标识符（DOI），并解释其组成部分。发票的对象标识符可由发票代码构成。
* 元数据收集：列出发票元数据字段，并为每个字段提供适当的值。
* 实体封装：抓取发票的对应信息并转变成json格式保存。

【拓展实验】

人工进行发票图片信息识别费时费力且容易出错，而现在文字识别（OCR）的技术已经完善，能够直接从图片中提取出文字信息.同时，大语言模型有着强大的逻辑推理能力，能够将识别出的信息并进行组织，按照定义好的发票数据对象模型，生成可用的发票数字对象。

请自行寻找可用的模型或平台，并基于此设计一套门诊发票自动识别AI

1. 为实现一个可信的数据交易平台，市数据中心决定采用区块链相关技术。请你基于H3C-Gaea区块链平台，设计实现并部署相关智能合约。要求实现以下核心功能：

* 数据资源（以上文的发票数据资源）的创建、修改、销毁、查询等
* 数据资源交易，某方支付某一数据资源的获取价格后，可得到该数据资源的使用权

【拓展实验】

思考数据交易场景下可能出现的其他需求，并在智能合约中实现

1. 按照任务要求，进行系统展示与实验报告书写。