去中心化数据集市（Decentralization Dataset Fair）产品文档

背景

大数据与人工智能技术的深度发展与应用，深刻的影响了我们生活的方方面面。而目前大数据与人工智能的一个主要痛点就在于数据的问题。而数据的价值，也逐渐被人们所重视。一方面数据可以作为重要资源，具备资产的价值。另外一方面，从单个个体来看小范围的数据资源又没有使用价值（这里不涉及隐私数据）。但从可用数据资源来看，这些数据资源都是集中在巨头企业之中。而去中心化数据集市的意义，就在于数据的共享与数据的资产化。很多节点（或者用户）组成的区块链网络，共享自己所有的数据资源能够形成巨大的数据集合，可以供给AI领域使用，而每个节点（或者用户）又可以从中获得共享数据的收益，自己也能利用这些数据来供自己的使用或者研究。

设计

该系统会设计这些方面：1、用户管理。2、问题域设计。3、数据上传。4、数据检查。5、数据下载。6、数据审查。7、系统自治。

1、用户管理。所有的用户都是基于Blockstack的DID身份管理系统来做。

2、问题域设计。所有的数据都是围绕着问题域来做。可以创建一些具体的应用领域数据，如电商领域，社交领域，计算机视觉领域，自然语言处理领域。只有加入这个问题域，才可以上传这个领域的数据。

3、数据上传。暂时只是把数据上传进去，不需要做数据清洗跟数据初步处理。

4、数据检查。这部分待定，可以运用智能合约，即在问题域中创建智能合约，对于不符合规范或者设计用户隐私的数据不让其上传。由于Blockstack暂时不支持智能合约，这部分可以以后在做。

5、数据下载。对于某问题域的数据，该问题域的成员可以下载这些数据集。

6、数据审查。任何一个数据集，都可以被该问题域的成员所审查。对于有异议的数据集，可以在系统中动议（如删除这个数据集），当投票超过半数即系统自动执行该动议。审查记录可以保存在该数据集下。

7、系统自治。这部分可能等Blockstack发行token(代币)之后，才能去做。即创建问题域，需要获得一笔投资。然后该笔投资在该问题域中运行，上传数据集可以获得代币，数据审查动议提出者可以获得代币，被执行动议者需要惩罚代币。下载数据集如何在该问题域中以代币的形势付款。以及进入该问题域的成员，也需要递交会费跟用户行为规范保证金。如果该用户违规或者违法，则在系统中对该用户进行惩罚。

创建问题域

