### 调研现状分析报告

#### 一、现有系统的业务模式

1. **数据管理业务模式**
   * **数据共享**：基于三清单（需求/责任/负面清单）机制，通过平台实现目录系统与交换系统对接，但负面清单未线上化。
   * **数据开放**：通过上海市开放平台统一开放文件数据，计划迁移至华为OBS存储。
   * **数据归集**：采用直连委办业务系统或前置机接入，实时采集以API解析为主，部分场景涉及Kafka/Flink。
   * **数据加工**：按需求方时间要求完成主题数据加工（10个工作日内），分层存储（ODS/DWD/DWS）。
   * **数据服务**：以库表、API共享为主，文件共享占比较小，需对接API服务管理工具。
2. **安全与合规模式**
   * 采用国密SM4加密，安全日志通过华为数据安全中心统一管理。
   * 数据销毁需安全团队（谋乐）全程监控，分类分级流程需与谋乐团队协同。
3. **运维与运营模式**
   * 前置机运维采用Zabbix监控，大数据平台依赖阿里云运维工具。
   * 运营团队角色包括中心、服务商、运维方，未严格细分内部角色。
4. **衍生业务模式**
   * 国家数据直达基层：通过普元系统对接国家平台API。
   * 社区云数据回流：以API方式实现加密数据采集与校验。

#### 二、现有系统的主要业务处理流程

1. **核心业务流程**
   * **数据共享流程**：需求对接→归集→元数据采集→编目→订阅→交换。
   * **数据加工流程**：加工规则制定→数据清洗→交付。
   * **工单处理流程**：建任务单→分解工单→审批→验收→任务单关闭。
   * **数据销毁流程**：一网协同平台下线状态触发→目录系统下线→安全团队监控销毁。
2. **辅助流程**
   * **数据安全分级**：预分级（上链前）→普元编目→谋乐分级检查。
   * **接口采集流程**：签名生成→分页请求（1000条/次）→解密校验→API上报对账。

#### 三、现有系统的业务运行情况

| **模块** | **现状描述** |
| --- | --- |
| 数据共享 | 负面清单未线上化，自动化授权机制缺失；市级代理接口需人工封装转发。 |
| 数据归集 | 直连委办系统存在风险，仅卫健委有前置机；实时采集以API为主，流式场景不明确。 |
| 数据服务 | API Key线下管理，调用日志需人工返还市级；文件共享占比低，需迁移至sFTP/OBS。 |
| 数据安全 | 加密与审计日志分散处理，完整性校验依赖手动操作。 |
| 工单系统 | 交付物以统计数据为主，评审记录未实际使用；角色间无数据隔离。 |
| 社区云数据回流 | 接口采集存在条数限制，需多次请求；签名生成依赖外部程序。 |

#### 四、工作负荷

| **业务类型** | **工作负荷描述** |
| --- | --- |
| 数据归集任务 | 调度异常频发，需定期测试；日均处理10+个委办局数据接入任务。 |
| 工单处理 | 2024年工单类型超20种，月均处理量约500单，需人工分类统计。 |
| 前置机运维 | 仅1台卫健委前置机，Zabbix监控覆盖服务器状态，日均巡检2次。 |
| 数据质量检测 | 单表字段超100时需逐条配置规则，无批量导入能力，耗时增加50%。 |
| 医保数据迁移 | 需迁移200G结构化数据+数十TB影像文件，涉及12类业务数据。 |

#### 五、局限性

1. **功能缺陷**
   * 数据目录缺少”访客模式”，未登录用户无法查看。
   * 数据集成工具不支持自动添加调度时间字段，预览功能缺失。
   * 工单系统无佐证模块，交付件需人工转换。
   * 数据水印能力未明确，对外供数安全策略不完善。
2. **处理能力不足**
   * 社区云接口单次仅支持1000条数据采集，全量采集效率低。
   * 数据质量规则配置无法批量导入，影响大规模数据处理。
3. **响应问题**
   * 告警通知依赖平台消息，无法通过邮件/短信推送（受网闸策略限制）。
   * API调用日志需人工对接市级表结构，响应延迟达1-3天。

#### 六、对现有系统的其他说明

1. **技术依赖**
   * 市级实时数据归集依赖阿里云DataHub，区级需适配对接。
   * 组织数据与认证依赖一网协同平台接口，服务商用户需独立管理。
2. **架构特点**
   * 采用混合云架构：政务外网与互联网区隔离，数据迁移需跨网闸。
   * 分层数据湖设计（ODS/DWD/DWS），但六层模型暂未启用。
3. **外部协同**
   * 需对接一网协同平台获取系统上下线状态，实现目录动态管理。
   * 与市开放平台、国家数据平台存在级联关系。

#### 七、用户改善性建议

| **模块** | **优化建议** |
| --- | --- |
| 数据共享 | 实现三清单全线上化，建立目录到共享的自动化授权机制。 |
| 数据服务 | 对接普元系统实现API Key线上管理，支持敏感数据动态加密。 |
| 工单系统 | 开发佐证模块，实现交付件自动转换与约束关系分析。 |
| 数据安全 | 集成智能分类分级工具，实现预分级与编目联动。 |
| 用户体验 | 增加数据目录访客模式、DeepSeek问答搜索、用户水印（组织+姓名）。 |
| 运维监控 | 替换Zabbix为国产化监控工具，支持国产服务器环境。 |
| 数据开放 | 建立高频数据治理机制，提升开放数据容量与密度。 |
| 技术能力 | 评估数据底座平台对大文件（影像/视频）的采集存储能力，优化流数据处理效率。 |

**备注**：后续需重点解决自动化能力缺失、跨系统协同效率低、大规模数据处理瓶颈等问题，同时关注市级标准动态调整对区级系统的影响。