引文数据库规范管理系统

需求分析说明书

1. 项目介绍
   1. 项目背景

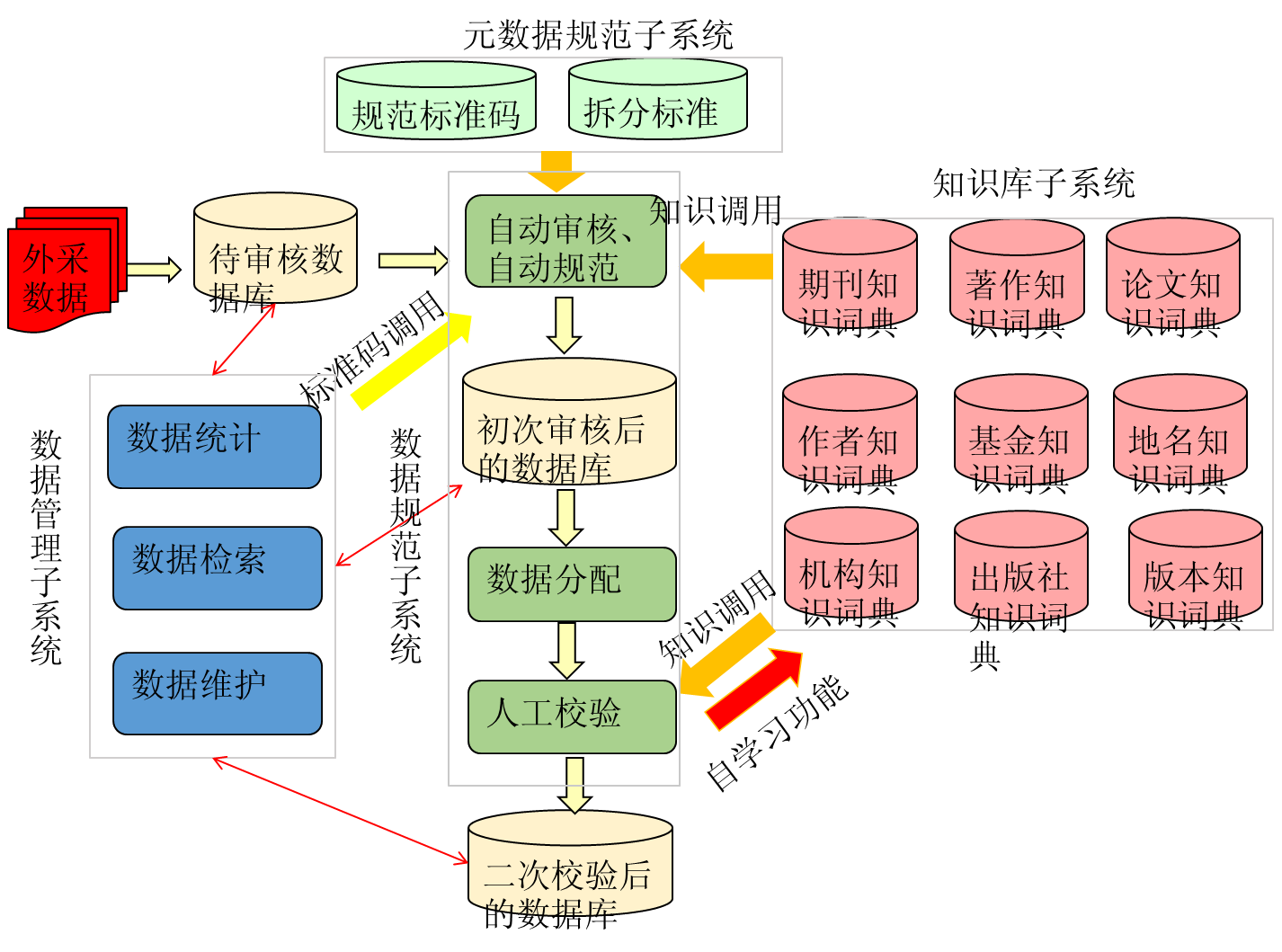
引文数据库规范管理系统的建设目标是为了优化引文数据规范模式，提升数据规范效率，进一步提高引文数据统计结果的准确性和可靠性，为完善哲学社会科学学术评价体系和评价标准，掌握学术评价话语权提供有力的数据支撑。

“中国人文社会科学引文数据库（CHSSCD）”始建于1999年，通过十余年的发展，目前CHSSCD已经成型并具有一定的规模，拥有1999年至今的近1300万条数据，并在此数据库基础上开展了多项文献计量学理论与科学评价应用研究，同时为科研工作提供定向服务。引文数据库规范管理系统的建立是CHSSCD的重要组成部分，随着引文数据的不断扩充和对数据质量标准的提高，原有数据加工规范系统已不能满足需要，本项目将在前期工作的基础上，对原结构设计上做出更为优化的设计和必要的更新，其工作流程的设计围绕着保证质量、降低成本和人机互动的高效原则，系统合理地明确了各层级任务。

引文数据库规范管理系统要求综合数据存储、自动审核、自动纠正、自学习、信息检索以及信息安全等高新技术手段，制定规范的数据标准体系，建设期刊数据资源规范加工平台，深入开发和应用期刊数据，促进以专著、学术论文、研究报告、论文集、学术资料、古籍、译著、译文、工具书、教材、普及读物、一般文章以及网络资源等引文信息的整合与利用，为科研成果评价提供有力支撑。

* 1. 项目内容

引文数据库规范系统的建设目标是利用系统的自学习能力进行引文数据信息的存储、管理、校对与修正及统计分析功能，通过对采购的原始数据的导入、索引、匹配校对、任务分配、审核入库等环节，规范数据的存储格式、确保内容的完整准确，以便于统计分析和挖掘。因此，引文数据库规范管理系统项目的建设内容包括五方面：元数据标准规范子系统建设、社科类期刊知识库管理子系统建设、数据规范处理子系统建设、数据资源管理子系统建设、运维管理服务子系统建设。总体设计思路如下图所示：



（一）元数据标准规范子系统建设

引文数据库元数据标准规范子系统的建设主要是对期刊论文的参考文献、基金、作者等信息制订的标引、分类和数据切分等标准，目的是通过数据标准规范体系向数据提供商进行明确的数据说明，规范引文数据库中的数据属性，解决数据注入不统一而导致数据的不准确等普遍存在的现实问题，并以此为基础对采购、加工的引文数据在数据审核时进行规范化的处理。

引文数据库元数据标准规范的制订要求遵循新闻出版总署关于印发《引文数据规范标准码》和《引文数据库著录规范手册》（评价中心研制）的相关要求。

（二）社科类期刊知识库管理子系统建设

建立社科期刊知识库及相关的知识管理维护系统，将数据规范、学术研究工作中形成的知识按类别进行集中的存储和管理，同时建立知识间的关联。为数据的自动审核、处理和后期人工校验提供后台的知识支撑。该知识库包括一系列知识词典，例如：期刊信息词典，著作知识词典、论文知识词典、作者词典、基金词典、地名词典、机构词典、出版社词典、版本词典等。

通过知识管理维护工具对知识库中的知识进行管理和维护，同时提供知识的自学习功能，在数据审核和学术研究中，有新的知识点被确认和使用时，系统自动将新的知识点纳入到对应的知识库中进行管理。

（三）数据规范处理子系统建设

基于社科期刊知识库，采用自动处理+人工审核校验的方式，实现引文数据的分类、审核、纠正等加工处理，将外部获取的社科期刊数据（参考文献著录规则不一的数据）不同转化为符合CHSSCD要求的数据。同时提供数据加工任务分配、数据校验、数据审核入库、任务统计及数据完成质量的评价等功能。

（四）数据资源管理系统建设

数字资源管理系统要求提供基于统一、安全和方便的数据管理入口，实现数据录入、校验和存储，为将来引入机构自建数据库资源与现有引文数据库整合提供可能，能实现跨库检索和数据统计功能。要求提供面向不同检索需要、功能先进的多种跨库检索方式，要求保证系统查全率和查准率的同时，有效地支持文献调研与情报分析等科研决策活动。此系统的数据统计功能还支持通过汇总、分析、数据挖掘，提供各类统计报表。

（五）运维管理服务系统建设

要求支持身份认证和资源使用权限管理等功能，并且详细记录用户的使用情况，形成完整的日志数据；要求支持通过汇总、分析、挖掘日志数据，提供各类日志统计和查询服务。

（六）平台支撑软件的采购及系统集成

本项目还需采购服务器操作系统软件、数据库平台软件等平台支撑软件，并完成系统的软件、硬件和科研成果数据库的集成（服务器等硬件的采购内容参见本项目另一标段）。

## 1.3建设周期

需求分析、系统设计、软件开发、平台采购、系统集成、试运行时间总计6个月。其中要求：标准建设、技术规范和详细的需求说明书要求在公司中标之日起2个月内完成，系统建设与平台采购4个月完成并进入试运行。

## 1.4建设原则

按照建设要求，在引文数据库规范管理系统设计上，应该遵循以下几方面的原则：

**（一）实用性原则**

引文数据库规范管理系统应建成功能实用、结构合理，可以为中国社会科学院各学术机构、研究人员和公众用户提供科研成果数字资源服务，为推动我国哲学社会科学信息资源共享服务开拓新型模式，同时规范科研成果资源共享行为，有效避免重复投资，提高资源的利用率。

**（二）可扩充性和开放性原则**

引文数据库规范管理系统的开放性是保证长效机制的具体要求，只有开放的引文数据库规范管理系统才能够兼容和不断发展，才能保证前期投资持续有效，保证引文数据库规范管理系统可分期逐步发展和日益完善。引文数据库规范管理系统在运行环境的软、硬件平台选择上要符合国家、部委与行业标准，具有良好的兼容性和可扩充性，能够较为容易地实现软硬件系统的升级和扩充，从而达到保护初期阶段投资的目的。

要求在设计上充分考虑引文数据库规范管理系统架构的扩展性和开放性，必须能方便地实现与其他信息的扩展对接，要求支持标准接口，方便中国社会科学院各学术机构获取科研成果数字资源，并快速应用于各自的专业系统中。

**（三）规范化原则**

引文数据库规范管理系统建设要求严格按照软件工程的一系列基本步骤（可行性论证、用户需求、初步设计、详细设计、项目实施计划、系统测试、系统试运行、系统验收）合理规范工程的实施过程。对每一阶段，要求提供相应阶段的书面报告。同时要求投标人予以配合，及时提供相关资料并认真分析需求，招标人将对每一阶段的成果进行及时的检验。

**（四）标准化原则**

标准化是引文数据库规范管理系统建设的基础，也是引文数据库规范管理系统与其他系统兼容和进一步扩充的根本保证。因此，对于引文数据库规范管理系统来说，系统设计和数据的标准化工作是极其重要的，这是各模块正常运行的保证，是系统开放性和数据共享的需要。

引文数据库规范管理系统建设符合国家、部委和行业标准，可直接调用实现资源共享，还可以根据需要制定自己的数据标准，保证系统整体的协调性和兼容性，便于规范化处理。引文数据库规范管理系统建设应在参照已有的图书文献行业标准的基础上，设计建立自己的数据标准。

**（五）安全性和保密性原则**

引文数据库规范管理系统的系统部署应充分考虑数据的保密与安全。要求引文数据库规范管理系统的运行及数据处理应用符合国家及部级主管部门的相关保密规定。

**（六）可靠性和稳定性原则**

稳定性是指引文数据库规范管理系统软硬件系统的正确性、健壮性两个方面。一方面，引文数据库规范管理系统在提交前应该经过反复测试，保证系统长期的正常运转；另一方面，引文数据库规范管理系统必须有足够的健壮性，在发生意外的软、硬件故障等情况下，能够很好地处理并给出错误报告，并且能够得到及时的恢复，减少不必要的损失。要求软硬件系统设计结构合理，系统运行稳定可靠。

**（七）经济性和实效性****原则**

要求在保证功能、性能指标的前提下，应尽可能降低成本，尽可能利用现有的资源，按计划在规定时间内实现项目建设目标。

## 1.5项目特点

引文数据库规范管理系统建设应当满足如下特点：

**（一）要求具有准确、快速的自动审核、纠错功能**

要求通过知识库和引文规范标准利用科学的算法，实现引文数据的初步分类、审核、纠正等加工处理，将外部采购的社科期刊数据快速的转换为符合要求的引文数据。

**（二）要求具有科学、高效的数据加工处理流程**

要求通过建立数据审核加工任务的分配、处理、审核、评价机制，优化评价中心的数据审核加工管理业务。

**（三）要求建立具有自学习能力的知识库**

要求通过建立具有自学习的社科期刊知识库及相关的知识管理维护系统，将数据审核、学术研究工作中形成的知识按类别进行集中的存储和管理，同时建立知识间的关联。为数据的自动审核和处理提供后台的知识支撑。

**（四）要求建立完备的社科期刊引文数据处理规范标准**

要求对已有的社科期刊引文数据处理规范标准进行完善和补充，建立满足学术评价的引文处理规范标准。

**（五）要求具有资源整合能力**

要求通过内容整合、检索整合等资源整合方式，实现中国社会科学院各研究机构已建数据库资源、外购数据库以及互联网资源与引文数据库规范管理系统的统一跨库检索。

**（六）要求具备数据检索及统计分析能力**

要求从多方面、多角度提供各种增值的知识服务，充分满足对信息、情报等方面的应用需求。并可以依据数据分析需求，以引文数据库规范管理系统所有数据为基础，完成相关的数据统计分析。

**（七）要求支持灵活的模块化结构**

要求从模板设置到功能服务，均支持自主配置，要求搭建灵活的、能够适应。

1. 技术需求

## 2.1 架构需求

引文数据库规范管理系统将以信息准确为中心，以服务学术评价，提高科研成果资源价值为目的，涵盖了科研成果资源数字化、分类、标引、检索、数据统计等各个环节。其总体构架如下图所示：

**科研成果评价**

**应用服务层**：科研成果评价，国家科研项目申报服务、科研人员成果评价服务，数据统计，数据报表发布。

**CHSSCD数据库**

已有CHSSCD数据库

数据管理层：数据管理维护、数据接口服务

引文数据库：作者、机构、基金…

数据人工审核模块

数据自动处理模块

供应商数据A

供应商数据B

供应商数据C

数据加工层

（一）数据加工层

要求完成各种的数据信息的标准化和元数据标引分类等工作。数字化资源是引文数据库规范管理系统的核心，引文数据库前期由来源库、引文库、作者库、基金库四部分构成，后续可根据使用过程中的实际需求进行库表和数据字段的扩充。通过采用自动处理+人工审核校验的方式，实现引文数据的分类、审核、纠正等加工处理，将外部采购的社科期刊数据转化为可以直接使用的社科期刊数据。

（二）数据管理层

要求提供对入库的数据进行增加、修改、删除、导入/导出等管理维护操作，依托存储备份机制对数据进行备份管理。

具有检索功能，方便数据共享与访问。数据管理层对上层应用服务层透明，提供基于标准协议的数据通信接口，要求提供链接整合、具有权限认证的检索整合和统一导航功能。要求可以支持原本分散在机构内部数据库内的数据资源充分发挥整体效益，进一步用于增值性的应用。

（三）应用服务层

应用服务层的主要目标是将引文数据资源发挥作用，提供高效的知识管理以及增值性服务。要求具有良好的扩展性设计，可以支持应用模块的新增、删除和升级等。要求支持与第三方应用系统通过标准接口实现对接。

## 2.2 设计需求

1) 要求采用开放的体系架构，系统必须支持以下标准与协议：

* + - OpenURL
    - ANSI/NISO Z39.50－1995. Information Retrieval(Z39.50)
    - MARC/CNMARC 机读目录格式
    - Dublin Core
    - Resource Description Framework (RDF) Model and Syntax Specification
    - Resource Description Framework(RDF) Schema Specification 1.0
    - Extensible Markup Language (XML) 1.0
    - OAI元数据获取协议(OAI Protocol for Metadata Harvesting)

2）要求支持DOI管理，通过DOI能获取记录的元数据信息。要求支持学科的主题词管理，满足学科主题词的特殊要求。

3) ▲要求必须采用成熟开放的大型关系数据库系统，能够满足引文数据库规范管理系统发展的需求，数据库结构定义和接口对用户开放。

4) ▲要求采用纯B/S架构兼容多种浏览器，包括IE10以上、Firfox 30.0以上、谷歌浏览器、360浏览器等。

5) 要求具有良好的可扩展性，要求完全支持Unicode，要求支持与中国社会科院已有应用软件系统的数据交换。要求支持二次开发，通过二次开发的各类应用程序可以与系统实现高度集成。

6) 要求具有较强的容错能力，使得系统在运行环境出现问题时能够最大限度保证数据安全。要求支持在业务不间断的情况下完成新旧系统的数据迁移。

7）要求支持DRM集成，实现对DRM产品的集成。

8) 要求提供数据库和库中表的详细结构及互相的关系，表及表中每一字段的意义和用途；提供一整套完整的API接口以及详细说明。评价中心的开发人员可调用这些API接口实现特定的业务功能。

## 2.3 功能需求

基于社科期刊知识库，采用自动处理+人工审核校验的方式，实现引文数据的分类、审核、纠正等加工处理，将外部采购的社科期刊数据转化为可以直接使用的社科期刊数据。

### 2.3.1元数据标准规范功能需求

要求投标人协助评价中心在引文数据库规范管理系统建设中，结合国家、行业、部委标准及中国社会科学院实际情况，在原有《引文数据规范标准码》和《中国人文社会科学引文数据库著录规范手册》（评价中心研制）的基础上制定元数据规范标准。该标准在系统进行数据自动处理时完成纠正错误信息及剔除无用信息、乱码。

例①：原数据：第1-5页；规范为：1-5

例②：原数据：第120-121页，转第200页

规范为：120-121，200

“中国人文社会科学引文数据库”为一种关系型数据库，共有4个数据子库，分别为：来源数据子库、参考文献数据子库、作者数据子库和基金数据子库，各数据子库的数据结构、字段设置参见附件1。可能形成的若干编码规范列表如下，具体实施时将根据情况有所增删。包括但不仅限于：

1、来源数据子库：

1. 来源刊名元数据标准
2. 来源年、期、页、卷元数据标准
3. 主办单位元数据标准
4. 主办单位地址元数据标准
5. 论文分类号元数据标准
6. 论文题名元数据标准
7. 关键词元数据标准
8. 语种元数据标准
9. CN与ISSN号元数据标准
10. 文献类型元数据标准

2、参考文献数据子库包括：

1. 母体元数据标准
2. 析出元数据标准
3. 母体属性元数据标准
4. 析出属性元数据标准
5. 编者、作者元数据标准
6. 出版社元数据标准
7. 年、卷、期、页元数据标准
8. 外文标识元数据标准
9. 出版地元数据标准
10. 译者元数据标准
11. 作者国别朝代元数据标准
12. 版次元数据标准
13. 编责元数据标准

3、作者数据子库包括

1. 作者名元数据标准
2. 作者单位元数据标准
3. 机构类型元数据标准
4. 省、市元数据标准
5. 邮编元数据标准
6. 学历元数据标准
7. 学历类型元数据标准
8. 职称元数据标准
9. 职称类型元数据标准
10. 专业元数据标准
11. 职务元数据标准
12. 籍贯元数据标准

4、基金数据子库包括

1. 项目名称元数据标准
2. 基金号元数据标准
3. 课题名称元数据标准

要求投标人协助招标人制定共享交换服务数据规范，如：目录和元数据规范、论文PDF图层数据规范、分类方法规范等。

要求投标人协助招标人制定符合ITIL标准的运维平台管理规范，如运维人员管理规定、数据加工服务流程规范、安全应急响应机制以及数据更新流程规范等等。

### 2.3.2知识库的自学习功能

要求建立社科期刊知识库，将数据审核、学术研究工作中形成的知识按类别自动记录，同时建立知识间的关联。同时提供知识的自学习功能，在数据审核和学术研究中，有新的知识点被确认和使用时，系统自动将新的知识点纳入到对应的知识库中进行管理，为数据的自动审核和处理提供后台的知识支撑。

要求平台应内置丰富的词典与词表信息，作为学术分析的基础；

要求该知识库包括期刊信息词典，著作知识词典、论文知识词典、作者词典、基金词典、省市词典、机构词典、出版社词典、版本词典等多种对象化信息；

要求词典词表可实时更新，人工干预。

#### 2.3.2.1期刊信息词典

字段设置：期刊名称、常见错误名称（期刊）、曾用名、学科分类、外文刊名、编辑出版单位、主办单位、地址、刊期、创刊年、出版规格、电话、电子信箱、网址等字段。

要求有良好的数据库记录维护界面，实现数据库的浏览、增加、删除、修改等数据库维护工作。

要求字段可以扩展，为后续可根据使用过程中的实际需求进行数据字段的扩充。

要求词典词表可实时更新，人工干预。

#### 2.3.2.2著作知识词典

字段设置：著作名称、常见错误名称（著作），作者、编者、译者、出版社、出版地、版本、出版年、卷等字段。

要求有良好的数据库记录维护界面，实现数据库的浏览、增加、删除、修改等数据库维护工作。

要求字段可以扩展，为后续可根据使用过程中的实际需求进行数据字段的扩充。

要求词典词表可实时更新，人工干预。

#### 2.3.2.3论文知识词典

字段设置：论文题名、常见错误名称（论文），作者、译者、所载刊物、年、期等字段。

要求有良好的数据库记录维护界面，实现数据库的浏览、增加、删除、修改等数据库维护工作。

要求字段可以扩展，为后续可根据使用过程中的实际需求进行数据字段的扩充。

要求词典词表可实时更新，人工干预。

#### 2.3.2.4作者知识词典

字段设置：作者姓名（中文、西文）、笔名、学科（研究方向）、学历、学历类型、职称、职称类型、所属机构、机构类型、性别、出生年、民族等字段。

要求有良好的数据库记录维护界面，实现数据库的浏览、增加、删除、修改等数据库维护工作。

要求字段可以扩展，为后续可根据使用过程中的实际需求进行数据字段的扩充。

要求词典词表可实时更新，人工干预。

#### 2.3.2.5基金知识词典

字段设置：基金名称、基金项目、项目号、基金类别等字段。

要求有良好的数据库记录维护界面，实现数据库的浏览、增加、删除、修改等数据库维护工作。

要求字段可以扩展，为后续可根据使用过程中的实际需求进行数据字段的扩充。

要求词典词表可实时更新，人工干预。

#### 2.3.2.6地名知识词典

字段设置：国别（中文、西文），地名（中文、西文）等字段。

要求有良好的数据库记录维护界面，实现数据库的浏览、增加、删除、修改等数据库维护工作。

要求字段可以扩展，为后续可根据使用过程中的实际需求进行数据字段的扩充。

要求词典词表可实时更新，人工干预。

#### 2.3.2.7机构词典

字段设置：机构名称、常见错误名称（机构）、机构曾用名、机构类型、地址、邮编、电话等字段。

要求有良好的数据库记录维护界面，实现数据库的浏览、增加、删除、修改等数据库维护工作。

要求字段可以扩展，为后续可根据使用过程中的实际需求进行数据字段的扩充。

要求词典词表可实时更新，人工干预。

#### 2.3.2.8出版社词典

字段设置：出版社名称、常见错误名称（出版社）、出版社曾用名、地址、邮编、电话等字段。

要求有良好的数据库记录维护界面，实现数据库的浏览、增加、删除、修改等数据库维护工作。

要求字段可以扩展，为后续可根据使用过程中的实际需求进行数据字段的扩充。

要求词典词表可实时更新，人工干预。

#### 2.3.2.9职称、学历词典

字段设置：职称、职称类别、职称级别（中、高、初）、学历、学历类型、专业、学位等字段。

要求有良好的数据库记录维护界面，实现数据库的浏览、增加、删除、修改等数据库维护工作。

要求字段可以扩展，为后续可根据使用过程中的实际需求进行数据字段的扩充。

要求词典词表可实时更新，人工干预。

### 2.3.3数据审核规范的人工智能化功能

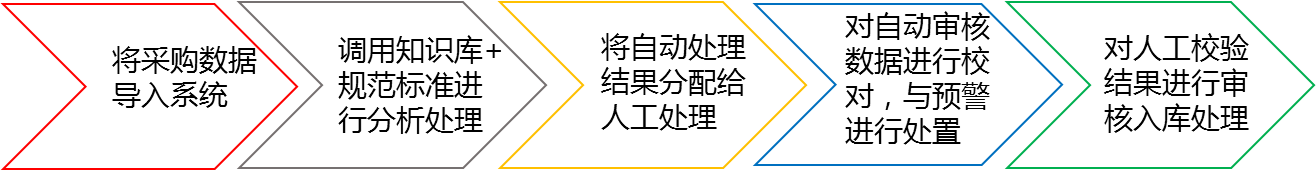
要求投标人实现基于社科期刊知识库和元数据规范标准的人工智能化数据审核规范，采用自动处理+人工审核校验的方式，实现引文数据的分类、审核、纠正等加工处理，将外部获取的社科期刊数据（参考文献著录规则不一的数据）不同转化为符合CHSSCD要求的数据。同时可对外采数据的质量进行评估，并能完成数据加工任务分配、数据校验、数据审核入库、任务统计及评价数据完成质量等功能。

### 2.3.4数据检索及统计功能

### 2.3.5数据管理的科学化

#### 数据审核加工流程的可追溯机制

建立数据审核加工任务的分配、处理、审核、评价机制，规范评价中心的数据审核加工管理业务。



#### 运维管理

要求支持角色管理和用户管理，可以方便地定义不同用户的权限，使得数据库操作同时具备方便性和安全性等。

要求实现对各子模块的权限分配，以及用户的特殊权限分配需求。要求支持对各个子模块和分类层级的赋权（如上传、编辑、下载、阅读、浏览等）；要求支持将权限映射到不同的角色（组）中，形成灵活的权限管理机制。

#### 日志管理

要求实现日志统计明细化：更加详细的日志分析、包括按数据库、用户和访问IP三种方式进行统计，每种方式都可细化到年月日，并都生成报表。

#### 数据检索及统计分析

要求实现从多方面、多角度提供各种增值的知识服务，充分满足对信息、情报等方面的应用需求。

要求实现数据库实现可以依据数据分析需求，以引文数据库规范管理系统所有数据为基础，从数据库中采集所需数据，进行各项数据对比分析，形成可视化报表，并设置有打印、导出功能。

要求支持对外部关联数据库或整合数据库采集数据信息；

要求支持在可视化图谱上进行分析操作；

要求支持报表的打印和导出功能。

### 2.3.6性能需求

#### 2.3.6.1整体性能需求

门户系统要求数据传输网络畅通、快捷、安全、可扩展；人机界面友好，输出、输入方便，图表生成美观，检索、查询简单快捷；

系统平台要求采用通用性好的计算机系统、安全可靠的操作系统以及大型数据库系统，保证系统良好的性能；

数据支撑平台要求应用支撑平台为业务应用系统的开发和运行提供技术支撑，并具有灵活的可扩充性和高度的可配置管理性；

安全性能应按照信息密级，在不同的信息安全域实施相应的安全等级保护，对不同安全等级的信息，通过身份认证和访问控制，实现授权访问，同时，整个系统具备数据备份、容灾恢复和应急响应等功能。

数据应及时、准确、完整，能够满足汇总统计、制表制图、分析计算等要求。

#### 2.3.6.2系统运行性能指标需求

引文数据库规范管理系统须具有支持海量数据、高性能、高可用、高可扩展、负载均衡以及良好的容错容灾能力。具体要求如下：

1) 海量数据支持

要求系统支持不小于100T规模数据的验证、存储、检索、分析的能力，同时支持数据随硬件性能提升的横向扩充。

2) 高性能

要求具有高响应速度和支持高并发特征。在正常网络速度情况下，支持响应各种用户请求的平均延迟不超过1秒。支持海量用户的大规模并发访问，单机支持不少于上百用户并发请求。整个服务器集群在不影响响应速度的前提下，要求支持1万用户同时在线，要求支持10万级别以上并发访问。

3) 高可用

要求在集群管理节点或者数据节点出现宕机的情况下，消灭“单点故障”，可以正常响应用户请求且响应速度不受影响。

4) 负载均衡

要求实现服务器工作节点间的负载均衡，根据服务器硬件、服务器网络、用户访问地区等因素设计负载均衡算法，将用户请求合理均衡到不同工作节点，充分发挥服务器性能。

5) 异地容灾

要求搭建主、备两套服务器集群环境，当主服务器集群因各种无法抗拒的因素出现整体宕机时，系统自动启用备用服务器集群，要求可以持续不间断地处理用户请求。

6) 高可扩展性

要求支持服务器集群系统的动态扩展：在集群系统不断电且能正常提供服务的情况下，支持工作节点的扩充；新添加的工作节点能自动响应服务请求；系统架构至少支持多套服务器集群以及每个集群内上百台服务器的扩展。

7) 集群容错

要求在集群有工作节点出现故障的情况下，故障节点的移除以及修复后的重新启用，不能影响集群的正常工作。

8) 数据量支持

文本转换性能要求：处理PDF文献资料的速度不低于200MB/秒

词典建立与更新性能要求：速度不低于1600汉词/秒，基于词典的字符串匹配分词效率不低于2400000(汉字/秒)

索引创建性能要求：索引创建速度不低于100MB/秒，索引文件的大小不高于0.19%，即基于100G的PDF数字资源索引不大于200M

查询流量性能要求：200G数据量，单词检索时间不低于0.003秒

查询延迟性能要求：平均响应时间不高于1秒

全文检索并发性能要求：不低于4000用户并发数

9)基本性能指标

全文检索速度：1秒/500GB

相似检索速度：20毫秒/100万文献量

单表记录个数：40亿

单表支持最大容量：8 TB (分区表最大容量 2PB)

可同时跨库检索最大数：255个

拥有400万数量级的概念关系词典

采用全切分分词引擎，切词准确率高达98%，速度达1M/S

10) 系统稳定性

系统有效工作时间（99.9%）

#### 2.3.6.3用户操作需求

系统要求软件界面友好，简单易操作，符合用户的操作习惯。能够简单方便地处理相应需求，用户不需要专业技术背景即可以方便地操作引文数据库规范管理系统。

## 2.4 系统安全保密需求

系统在实施过程中须在通讯链路、数据输入/输出、应用操作以及管理制度等诸多方面保证安全，对系统的所有用户进行分级管理，设置不同的角色，每个角色分配不同的操作权限和数据权限，符合中国社会科学院图书馆（调查与数据信息中心）数据保密性的要求。系统安全特性、访问控制须到页面级，具有较强的系统安全性和灾难恢复能力。同时，要求引文数据库规范管理系统的运行及数据处理应用符合国家及部级主管部门的相关保密规定。

## 2.5故障处理需求

引文数据库规范管理系统在设计时应充分考虑稳定性及对用户误操作的容错功能，保证在正常情况下系统能保持长时间无故障运行。但对于由于不可预见的原因导致的系统故障，应提供各种故障处理恢复机制使系统在尽可能短的时间内恢复运行，保证数据的安全性和完整性。

# 3其他需求

## 3.1项目具体需求要点

1. 为了保证系统的服务和质量，投标人需提供系统原厂商针对本项目的专项授权函原件，以及原厂商出具的所投产品能满足本性能及参数要求的确认函原件。
2. 投标人须根据项目要求结合本身从业经验，做出方案和说明，详细阐述该方案设计原则、依据、功能、性能、数据共享及所设计系统的技术先进性、科学性、合理性和操作的便捷性，所使用的开发工具、平台及与各业务关联技术实现。
3. 招标文件只是提出了本次引文数据库规范管理系统项目的一些原则性规定和思路，并不是详尽的要求，无论某些需求是否被明细列在招标文件中，投标人均有责任完善整个方案设计。
4. 投标人须对由招标人提供的所有数据信息实行严格的保密，并保证不伤害招标人及关联单位的利益，须与招标人签订有关保密协议。
5. 投标人须提供所开发软件系统的所有源代码，设计说明文档（系统设计文档、详细设计文档等）以及预留二次开发接口（包括二次开发接口说明文档）。
6. 为保证该系统今后的扩容升级建设需求，投标人在总体方案制定中应留有扩容升级余地，应用软件系统不设置许可证限制。
7. 投标人需说明采用何种软件开发环境，并详细说明其优缺点、所开发的软件在与其他子系统集成时可能出现的问题以及对其他子系统在选择软件开发环境方面的要求。
8. 本项目签订合同后，投标人需再次进行详细需求调研，并由项目主管单位、项目监理单位、建设单位三方对调研成果（即《软件需求说明书》）进行确认，作为项目开发过程中的重要范围依据。
9. 报价要求

投标人应根据自身经验，充分考虑在招标人提供的现有条件下，建立引文数据库规范管理系统项目的工作难易程度、工作量大小及对评估人员的技能水平、估价质量和时间要求等情况，合理报价。

投标报价应以人民币为结算货币，须包括服务价款、税费以及与该服务有关的一切费用。

## 3.2服务要求

1、投标人自项目最终验收通过之日起，为用户方提供三年的免费服务。

2、投标人保证提供7×24小时服务响应，系统出现故障时，维护工程师应在2小时内到现场，并在4小时内解决问题。

3、投标人必须提出服务期内的维护内容和范围（产品、技术、模块）。

4、在开发期和试运行期内，仍有可能不断完善需求，投标人须承诺在采购需求和政策法规范围内，随着用户需求的变动随时做出响应，修改应用软件。正式验收通过后，若有需求变动，在维护期内，仍应按用户需求对整个系统做出相应修改，以满足用户的需求。

## 3.3培训要求

本项目实施过程中，投标人应根据项目实施的计划、进度和需要与招标人的合理要求，及时安排相关人员进行培训。对于系统开发和维护人员的培训目标是能够独立、熟练地完成系统操作，并能够保障系统的日常维护及在基础平台上进行简单的二次开发；对于系统使用人员应使其能够独立地完成系统操作，并能进行科研成果发布及科研成果管理工作。

## 3.4知识产权要求

成果版权及使用，本项目的最终用户为招标人，项目的知识产权归招标人。