面向智慧国检的智能审核与异常预警  
项目计划书

# 一、项目简介

本项目将针对出入境新政带来的新需求、新问题，包括货物出入境通关时间大幅减少、抽检比例大幅降低、跨境贸易纠纷增加、监管科学化智能化的出入境检验检疫，采用大数据、区块链、物联网等先进技术，研究开发基于区块链的可信大数据服务、基于大数据的异常预警服务，在此基础上设计开发支持异常预警的出入境检验检疫智能审核系统，形成成体系的信息规范、基础设施、关键技术、核心系统与公共服务，打造基于大数据的可信、可追溯智慧国检平台，提升出入境检验检疫的服务能力与服务质量，促进口岸贸易物流快速发展。

# 二、项目背景与意义

上海是全球第三个年航空货量400万吨以上的城市，上海两机场是中国首个货运功能区纳入综合保税区和自贸试验区的机场，浦东机场是世界上唯一全球三大国际物流集成商进驻的机场；上海机场拥有三个货邮国际（地区）转运中心，电子运单使用率达52%，居全球机场第二、中国第一，航空货运信息集成系统日趋成熟，智慧物流园区建设正积极推进。

上海机场正在进一步构建连接“一带一路”沿线主要空港的航线网络，服务于“一带一路”沿线的产业转移和经贸合作，提升“一带一路”沿线国家的国际竞争力。为了对接国家战略，促进口岸贸易物流快速发展，国家质检总局于2017.11.1发布了出入境检验检疫的新政策，对通关效率、服务、质量提出了新的要求：

* 通关时间大幅减少、抽检比例大幅降低
* 跨境贸易纠纷增加、监管科学化智能化
* 基于大数据、区块链、物联网的可追溯智慧国检

为了提升出入境检验检疫的服务能力与服务质量，形成成体系的信息规范、基础设施、关键技术、核心系统与公共服务，打造基于大数据的可信、可追溯智慧国检平台。

# 三、研究目标与内容

## 3.1 基于区块链的可信大数据服务

### 3.1.1 智慧国检区块链应用方案与技术框架

【请咨询任博士】

应用方案：

1. 数据的存证
2. 商品信息的追踪溯源

技术框架：

电子商务交易技术国际实验室的众链

### 3.1.2 智慧国检查验数据分析与采集

【请杨明补充“查验数据”“数据来源”“数据性质”“数据类型”“采集方法”“相关技术”】

查验数据

调研智慧国检项目中哪些字段是业务所需要的，这些字段的所在的位置，这些字段的采集办法以及使用的相关技术

数据来源

1. 单证类数据
2. 数据库中存储数据
3. 系统通信过程中未录入的数据

采集办法

1. 通过ocr相关技术获取格式化数据
2. 调用数据库API获取数据库中的存储数据
3. 通过调用相关系统API设置http服务来存储系统数据

相关技术

1.ocr等相关文字图像相关识别技术

2.数据库编程

3.web编程与接口开发相关技术

### 3.1.3 面向智能查验的图像文本化处理

针对审核智能化的需要，分析智能审核涉及的图像类型、图像内容，选择高效、可靠、精准的开源图像处理软件，在其基础上设计图像处理模型与算法，开发针对性的图像处理子系统（模块），实现图像文字化、结构化、语义化。

OCR技术简介

OCR技术是光学字符识别的缩写(Optical Character Recognition)，是通过扫描等光学输入方式将各种票据、报刊、书籍、文稿及其它印刷品的文字转化为图像信息，再利用文字识别技术将图像信息转化为可以使用的计算机输入技术。可应用于银行票据、大量文字资料、档案卷宗、文案的录入和处理领域。适合于银行、税务等行业大量票据表格的自动扫描识别及长期存储。相对一般文本，通常以最终识别率、识别速度、版面理解正确率及版面还原满意度4个方面作为OCR技术的评测依据；而相对于表格及票据， 通常以识别率或整张通过率及识别速度为测定OCR技术的实用标准。

OCR技术工作流程

影像输入:欲经过OCR处理的标的物须透过光学仪器，如影像扫描仪、传真机或任何摄影器材，将影像转入计算机。科技的进步，扫描仪等的输入装置已制作的愈来愈精致，轻薄短小、品质也高，对OCR有相当大的帮助，扫描仪的分辨率使影像更清晰、扫除速度更增进OCR处理的效率。

影像预处理：影像预处理是OCR系统中，须解决问题最多的一个模块。影像须先将图片、表格及文字区域分离出来，甚至可将文章的编排方向、文章的提纲及内容主体区分开，而文字的大小及文字的字体亦可如原始文件一样的判断出来。

对待识别图像进行如下预处理，可以降低特征提取算法的难度，并能提高识别的精度。

二值化：由于彩色图像所含信息量过于巨大，在对图像中印刷体字符进行识别处理前，需要对图像进行二值化处理，使图像只包含黑色的前景信息和白色的背景信息，提升识别处理的效率和精确度。

图像降噪：由于待识别图像的品质受限于输入设备、环境、以及文档的印刷质量，在对图像中印刷体字符进行识别处理前，需要根据噪声的特征对待识别图像进行去噪处理，提升识别处理的精确度。

倾斜校正：由于扫描和拍摄过程涉及人工操作，输入计算机的待识别图像或多或少都会存在一些倾斜，在对图像中印刷体字符进行识别处理前，就需要进行图像方向检测，并校正图像方向。

文字特征抽取：单以识别率而言，特征抽取可说是 OCR的核心，用什么特征、怎么抽取，直接影响识别的好坏，也所以在OCR研究初期，特征抽取的研究报告特别的多。而特征可说是识别的筹码，简易的区分可分为两类：一为统计的特征，如文字区域内的黑/白点数比，当文字区分成好几个区域时，这一个个区域黑/白点数比之联合，就成了空间的一个数值向量，在比对时，基本的数学理论就足以应付了。而另一类特征为结构的特征，如文字影像细线化后，取得字的笔划端点、交叉点之数量及位置，或以笔划段为特征，配合特殊的比对方法，进行比对，市面上的线上手写输入软件的识别方法多以此种结构的方法为主。

对比数据库：当输入文字算完特征后，不管是用统计或结构的特征，都须有一比对数据库或特征数据库来进行比对，数据库的内容应包含所有欲识别的字集文字，根据与输入文字一样的特征抽取方法所得的特征群组。

对比识别: 这是可充分发挥数学运算理论的一个模块，根据不同的特征特性，选用不同的数学距离函数，较有名的比对方法有，欧式空间的比对方法、松弛比对法（Relaxation）、动态程序比对法（Dynamic Programming，DP），以及类神经网络的数据库建立及比对、HMM（Hidden Markov Model）…等著名的方法，为了使识别的结果更稳定，也有所谓的专家系统（Experts System）被提出，利用各种特征比对方法的相异互补性，使识别出的结果更加理想。

字词后处理：由于OCR的识别率并无法达到百分之百，或想加强比对的正确性及信心值，一些除错或甚至帮忙更正的功能，也成为OCR系统中必要的一个模块。字词后处理就是一例，利用比对后的识别文字与其可能的相似候选字群中，根据前后的识别文字找出最合乎逻辑的词，做更正的功能。

字词数据库：为字词后处理所建立的词库

结果处理与输出

经过处理后，在该阶段进行输出结果的处理使其满足一些特殊需求

OCR 主流框架

openocr

Google tesseract ocr

### 3.1.4 面向智慧国检的区块链应用开发

【请咨询任博士补充 “智慧国检数据存取方案” “区块链接口”】

### 3.1.5 基于区块链大数据的可信云服务

【由李银胜补充“智慧查验追溯体系”“智慧国检可信云开发” “智慧国检可信云运维”】

## 3.2 基于大数据的异常预警服务

### 3.2.1 智能查验异常预警场景分析

【请补充含场景分析与异常判断、布控策略】

（1） 冷链监管项目

货检课题组、圆立课题组、中检溯源课题组跟进

（2） 爱马仕智能查验项目

货检课题组、圆立课题组跟进

（3）鲜活货项目

复旦课题组跟进

（4）危化品项目

复旦课题组跟进

（5） 电子证书项目

复旦课题组跟进

### 3.2.2 质量管控异常预警场景分析

【请补充含场景分析与异常判断、布控策略】

### 3.2.3 大数据背景下的异常预警

【课题进行过程中完善】

（1）不能满足的预警需求

（2）无法解决的预测问题

（3）预警大数据分析模型

## 3.3 支持异常预警的智能审核系统

### 3.3.1 智能审核系统基本功能

（1）预录

* 实现代理或者收货人数据库对接；
* 实现数据自动导入，替换货物清单上传。

（2）预约

* 空港版进行对接，实现监管仓库放行、大众仓库查验预约、大众派车提货功能；
* 增加功能设定，如果存在不查验货物的货物清单没有上传的情况，不能进行预约；
* 条形码打印机调试。

（3）查验

* 增加标签合格批量选择功能；
* 实现抽采样信息的自动导入；
* 增加查验要求选项；
* 增加大众查验场站放行、整改、结算功能；
* 增加服装预检测报告核对功能；
* 增加中文标签仓库和备案信息的比对功能。

（4）采样

* 完善采样结果判定。

（5）登记

* 和ECIQ关联，在ECIQ没有完成核签之前，系统不能提交登记；
* 自动关联和生成电子版检验检疫证书。包括合格和不合格证书；
* 自动生成电子版原始记录单。

（6）查询

* 增加周报、月报、年度分析报告以及日常统计数据的生成功能；
* 增加货物的流程查询功能；
* 增加查询电子版相关单证的功能。

（7）增加管理界面

* 增加系统流程的回退功能；
* 增加中文标签整改仓库后续监管和标签整改承诺书核签功能；
* 增加企业诚信扣分功能，并与出入境企业信用管理系统对接；
* 增加无效数据的删除功能。

（8）需完善的问题

* 部分企业无法申请无纸化账号；
* 部分报检号的数据无法查询；
* 同一个收货人有多家代理公司，信息安全问题；
* 部分报检号检测结果反馈不正确；
* 当前电子版原始记录单登记结果不全，需进一步完善。

### 3.3.2 智慧国检之智能预审与数据统计

主干系统业务覆盖整个货检处。目前ICIQ主干系统主要以轻纺服装业务为主。在冷链、鲜活货等项目完成前主干系统需要将货检所有业务覆盖到位，确保主干系统与各子项目顺利对接。

主干系统查询统计功能需求：

* 数据库整理。使用现有的eciq数据库，筛选出货检处的数据（含审单放行）。而后对数据进一步切分细化，使数据能被定位到哪个科室具体哪项业务；
* 数据统计查询。调研统计查询需求，务必能使上级领导在电脑前能准确查询或统计到其想要的业务数据；
* 数据展示。设计数据图表样式，清晰明了美观；
* 报表生成。调研收集货检处所有报表，设置一键生成。

### 3.3.3 智慧国检之异常预警与布控

（1）主干系统由查验货物范围扩充涵盖所有审单放行货物

ICIQ覆盖范围是查验货物，随着11.1新政诸多货物被审单放行后系统无法覆盖整个空港货运物流情况。可以此为契机进一步深化智慧国检系统作用，将系统前置到审单环节，通过系统的智能筛查使真正安全可靠的货物被审单放行，而对于首次进口的、曾经有不合格、企业诚信出现问题等情况的货物在审单环节被提示风险。另一方面，所有货物的进口数据也会被完整纳入到系统中。

（2）加入货物首次入境识别功能

根据上级文件要求，首次入境的产品不实施审单放行。目前现有的系统无法识别首次入境，需进一步研究智能验证的方式由ICIQ在审单环节识别产品是否为首次入境。

（3）加入历次入境及检验记录翻查功能

进入查验环节后，查验人员可以在ICIQ中翻查该产品历次入境及检验记录，一方面作为风险提示支撑检验工作，另方面使各检验员执法尺度更统一避免出现相同产品不同检验结果的情况。

（4）进口计划审查制度

对于企业信用等级较高的，产品多次检验合格品质稳定的情况下，允许企业提交进口计划，经审查通过后续计划内进口的产品均列入审单放行。

（5）后续监管制度

对于审单放行的货物，由主干系统根据风险等级智能筛选出需要后续监管的企业和产品，经查不符合要求一方面可反馈到系统中降低企业信用等级、提高产品风险程度，另方面按智慧国检查验子项目要求对其后续远程监管完成整改。

### 3.3.4 智慧国检之基于虚拟现实的远程查验

（1）MR智能硬件

（2）MR专业软件

圆立课题组跟进

# 四、关键技术与创新

## 关键技术一：出入境检验检疫数据采集

本项目将结合出入境通关新政对抽检比例降低、审单时间减少的新要求，通过问卷、交流与用户面谈，调查出入境检验检疫涉及的风险及其相关信息来源，研究开发基于数据存取、图像处理、结构化信息处理技术；综合考虑区块链数据存取需求，开发多源数据采集与存取接口，在实现传统风险信息包括首次进口的、曾经有不合格、企业诚信出现问题等情况的货物信息追溯与采集基础上，研究通关新政背景下新出现的风险及其信息采集。

## 关键技术二：出入境货物查验异常预警

本项目将调查检验检疫动态监控对异常预警的需求，在已有的风险信息基础上，分析各数据源的数据类型与性质，研究开发针对性的数据统计、数据挖掘、大数据分析算法，特别是传统数据分析方法难以发现、处理的业务异常，研究开发针对性的大数据采集与处理算法，选择适当的大数据分析平台与工具，开发基于大数据的风险异常预警及其可视化工具，支持多维度、多性质数据可视化，实现异常预警模块与智慧国检相关系统有机集成。

## 关键技术三：出入境货物查验关联数据保全

本项目将调查货物出入境涉及的供应链与申报环节，分析基于物联网技术的货物生产、流通、市场可采集数据，研究制定货物可追溯数据标准规范，研究开发基于联盟区块链的可追溯数据存取机制与接口，实现基于区块链的可追溯服务“可信云”。

## 关键技术四：支持自动风险布控的智能预审

本项目将调查检验检疫监管部门的决策需求，结合现有审核流程与问题，分析大数据分析条件下的智能审核服务、流程与决策依据，实现货物出入境智能审单、业务统计、交互审核、异常监测。

## 关键技术五：支持货物风险处理的智能决策

本项目将调查检验检疫监管部门的货物风险处理政策与数字化需求，研究开发基于异常预警大数据的智能决策支持系统，实现相关风险信息自动提取、自动分析、自动风险布控、智能风险处理决策支持。

# 五、年度计划与指标

## 第一年度：

完成异常预警场景、需求分析，完成智慧国检区块链实验环境；图像处理算法与子系统基本满足使用要求；实现部分大数据分析的可视化；可演示基本概念的大数据分析算法框架与原型；实现智能审核的主要功能。

## 第二年度：

丰富异常预警场景；实现一个以上大数据分析需求与算法；实现智慧国检区块链业务环境；图像处理算法与子系统达到实用要求；实现大数据分析的可视化；提升智能审核系统的用户体验与功能。

## 第三年度：

形成基本完善的异常预警场景、大数据分析模型、预警功能；完成可实用的实现智慧国检区块链业务环境；完成大数据分析的可视化；形成用户体验良好、稳定、可靠的智能审核系统。

# 六、预期成果及考核

## 6.1 基于区块链的可信大数据服务

成果交付形式：（软件）[代码、文档]

主要技术指标：

## 6.2 基于大数据的异常预警服务（软件）

成果交付形式：（软件）[代码、文档]

主要技术指标：设计初步的大数据分析框架；可演示大数据分析涉及的基本概念与效果；实现1个以上典型的大数据预警功能；大数据分析模块与ICIQ预审核系统有机集成、独立更新。

## 6.3 支持异常预警的出入境检验检疫智能审核系统

成果交付形式：（软件）[代码、文档]

主要技术指标：

## 6.4 面向智能审核的大数据可视化子系统

成果交付形式：（软件模块）[代码、文档]

主要技术指标：实现业务统计、审核过程、异常监测；支持多维度、多性质数据展现；直观、形象；突出数据之间的关联性；支持一定的交互性；可视化模块与ICIQ预审核系统有机集成、独立更新。

## 6.5 面向智能审核的图像处理子系统

成果交付形式：（软件模块）[代码、文档]

主要技术指标：支持2两种以上类型的图像处理；文字正确率>95%；结构准确率>95%；图片处理速度不慢于2秒/图像；图像处理模块与ICIQ预审核系统有机集成、独立更新。

## 6.6 面向远程查验的MR子系统

成果交付形式：MR硬件、软件APP [代码、文档]

主要技术指标：

## 6.7 智慧国检区块链接口信息规范 [文档]

成果交付形式：文档

主要技术指标：

## 6.8 大数据、区块链领域的关键技术

成果交付形式：[代码、文档、论文]

主要考核指标：