

# 大数据与农业监测预警

吴建寨 副研究员

中国农业科学院农业信息研究所

# 内 容 提 纲

一、“大数据时代”的来临

二、农业大数据的发展

三、农业监测预警对大数据的需求

四、大数据背景下的农业监测预警

# 一、“大数据时代”的来临

# （一）如何认识“大数据时代”

## 1、数据度量角度的感性认识：

1Byte = 8 Bit

1KB = 1,024 Bytes

1MB = 1,024 KB = 1,048,576 Bytes

1GB = 1,024 MB = 1,048,576 KB = 1,073,741,824 Bytes

1TB = 1,024 GB = 1,048,576 MB = 1,099,511,627,776 Bytes

1PB = 1,024 TB = 1,048,576 GB = 1,125,899,906,842,624 Bytes

1EB = 1,024 PB = 1,048,576 TB = 1,152,921,504,606,846,976 Bytes

1ZB = 1,024 EB = 1,180,591,620,717,411,303,424 Bytes

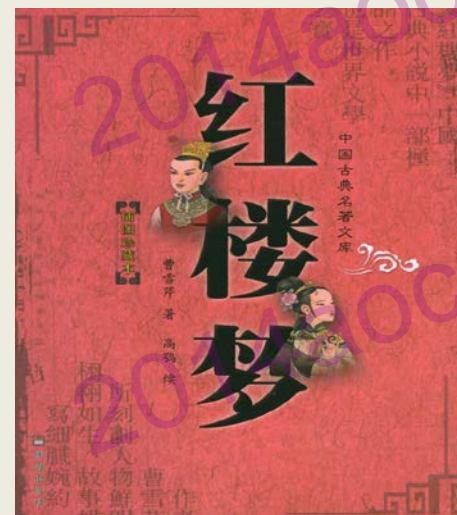
1YB = 1,024 ZB = 1,208,925,819,614,629,174,706,176 Bytes

《红楼梦》含标点87万字（不含标点853509字）

每个汉字占两个字节：1汉字=2bytes

1TB 约等于631,903 部

1PB 约等于647,068,911部



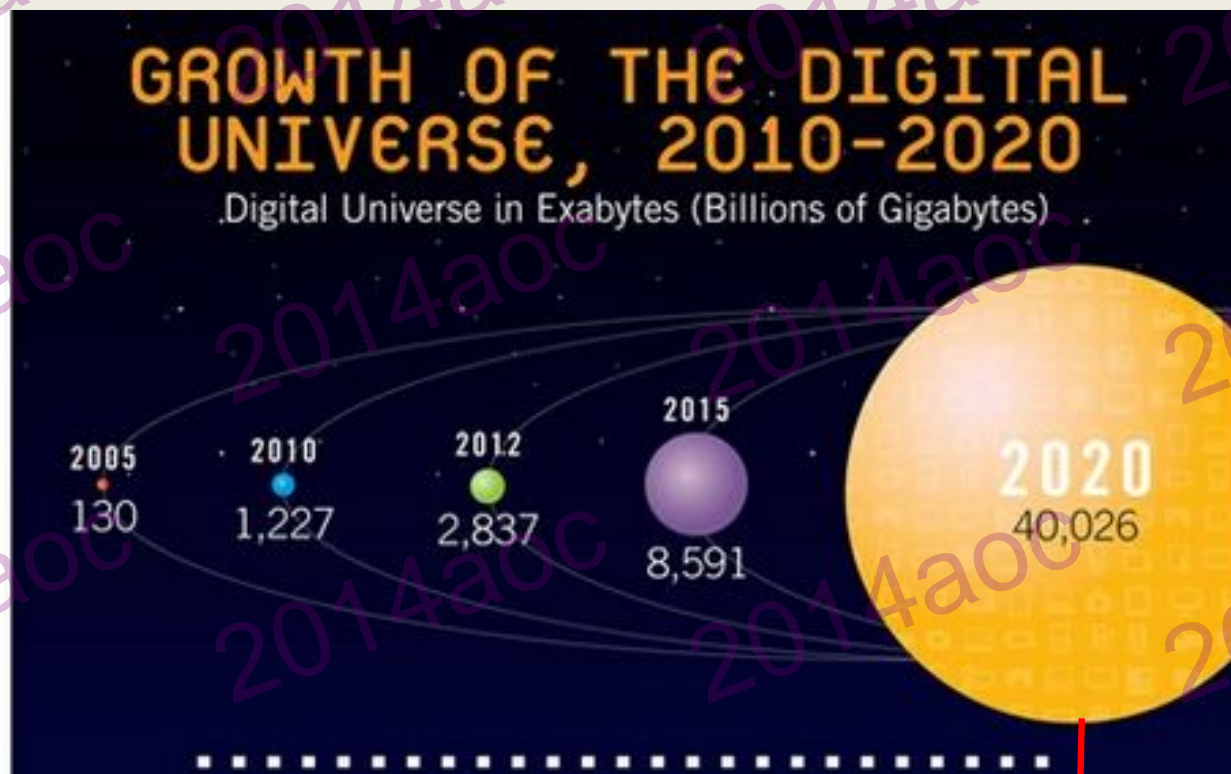
中国国家图书馆藏书3119万册(2012年底 )

百度每天处理的数据量将近100个PB,

百度1天≈5000个国家图书馆的信息量



IDC对未来全球数据存储量的调查：



比2012  
增14倍

中国占  
22%

约等于地球上沙滩上所有沙粒总和的57倍

信息技术的发展，推动数据变“大”！

## 2、“数据”与“信息”认识的层次递进：

- ◆ 目前社会大系统中的组织、个体都在通过信息发生着联系，涵盖到社会的各个领域。
- ◆ 信息社会的概念早已经被提出。



**数据 ≠ 信息**

数据处于信息的底层；

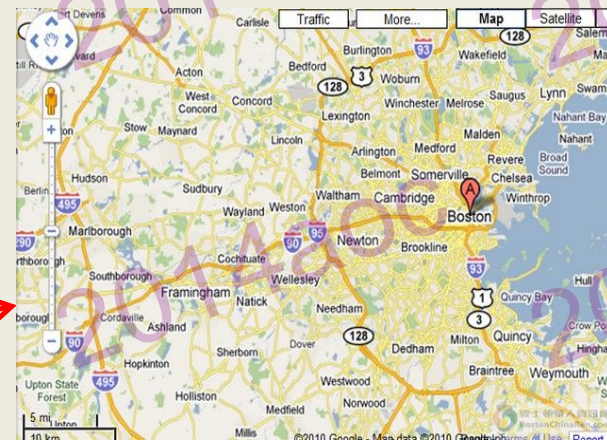
信息是对数据的提炼。



**“信息社会”到“大数据时代”，人类认识客观世界加深！**



## (二) “大数据” 行业应用



商品零售

电力、  
医疗...

交通管  
理

2014 “  
两会”

公共安  
全





### (三) “大数据”研究

Full text access provided to Chinese

# nature

International weekly journal of science

Search

Journal home > Archive > Letter > Full Text

**Journal content**

- Journal home
- Advance online publication
- Current issue
- Nature News
- Archive**
- Supplements
- Web focuses
- Podcasts
- Videos

## Letter

*Nature* **457**, 1012-1014 (19 February 2009) | doi:10.1038/nature07634; Received 14 August 2008; Accepted 13 November 2008; Published online 19 November 2008; [Corrected](#) 19 February 2009

### Detecting influenza epidemics using search engine query data

Jeremy Ginsberg<sup>1</sup>, Matthew H. Mohebbi<sup>1</sup>, Rajan S. Patel<sup>1</sup>, Lynnette Brammer<sup>2</sup>, Mark S. Smolinski<sup>1</sup> & Larry Brilliant<sup>1</sup>

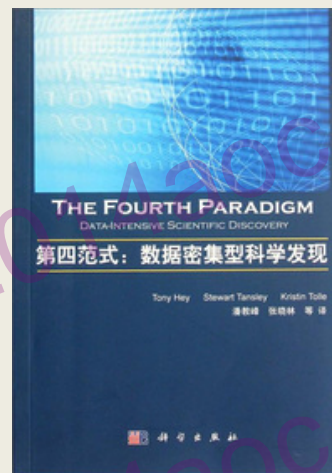
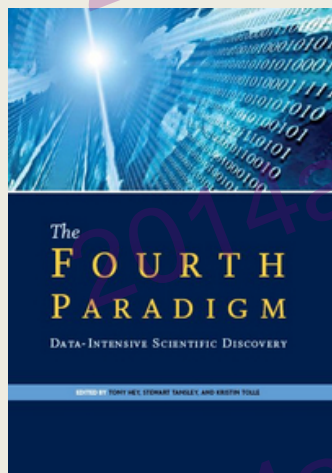
1. Google Inc., 1600 Amphitheatre Parkway, Mountain View, California 94043, USA
2. Centers for Disease Control and Prevention, 1600 Clifton Road, NE, Atlanta, Georgia 30333, USA

对于流感爆发的季节、全美流感的分布，可以到具体的州与地区。

<sup>1</sup>Google Inc., 1600 Amphitheatre Parkway, Mountain View, California 94043, USA.

大数据的理念与技术，推动研究领域拓宽！

# 吉姆·格雷（2007）科学研究第四范式：



- ◆ 2011年,《science》杂志推出了“Dealing with Data”专题。



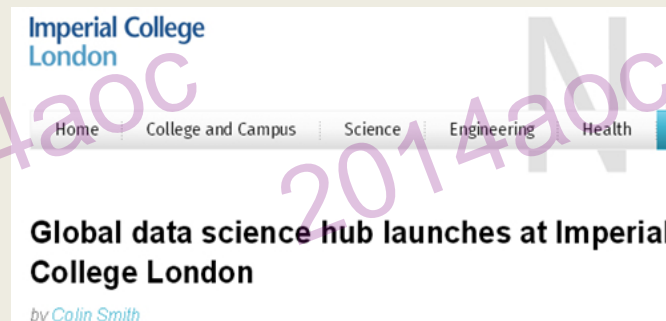
- ◆ 2012年3月美国联邦政府发布了大数据研发专项研究计划 (Big Data Initiative), 提升到国家战略。



- ◆ 2014年1月份斯坦福大学和牛津大学联手获得300万美元研究资金 (来自李嘉诚基金会)



- ◆ 2014年4月1日, 英国帝国理工学院宣布启动大数据科学研究中心





## ◆国内研究纷纷启动

2012、2013年多次会议主题。



研究专项启动。



专门研究机  
构成立。



大数据科学与工程研究中心在北航揭幕

【发布时间:2012年09月18日】 【来源:北京航空航天大学】 【字体:大 中 小】



## 二、农业大数据的发展



2011年5月，麦肯锡公司发布《大数据：创新、竞争和生产力的新前沿》报告，认为大数据已经渗透到各个行业领域，成为与物质资产和人力资本并论的生产要素。



传感器、遥感卫星、高性能计算机模拟、高通量仪器、科学影像等产生的电子化的科学数据，将大大提高科学数据的产生。

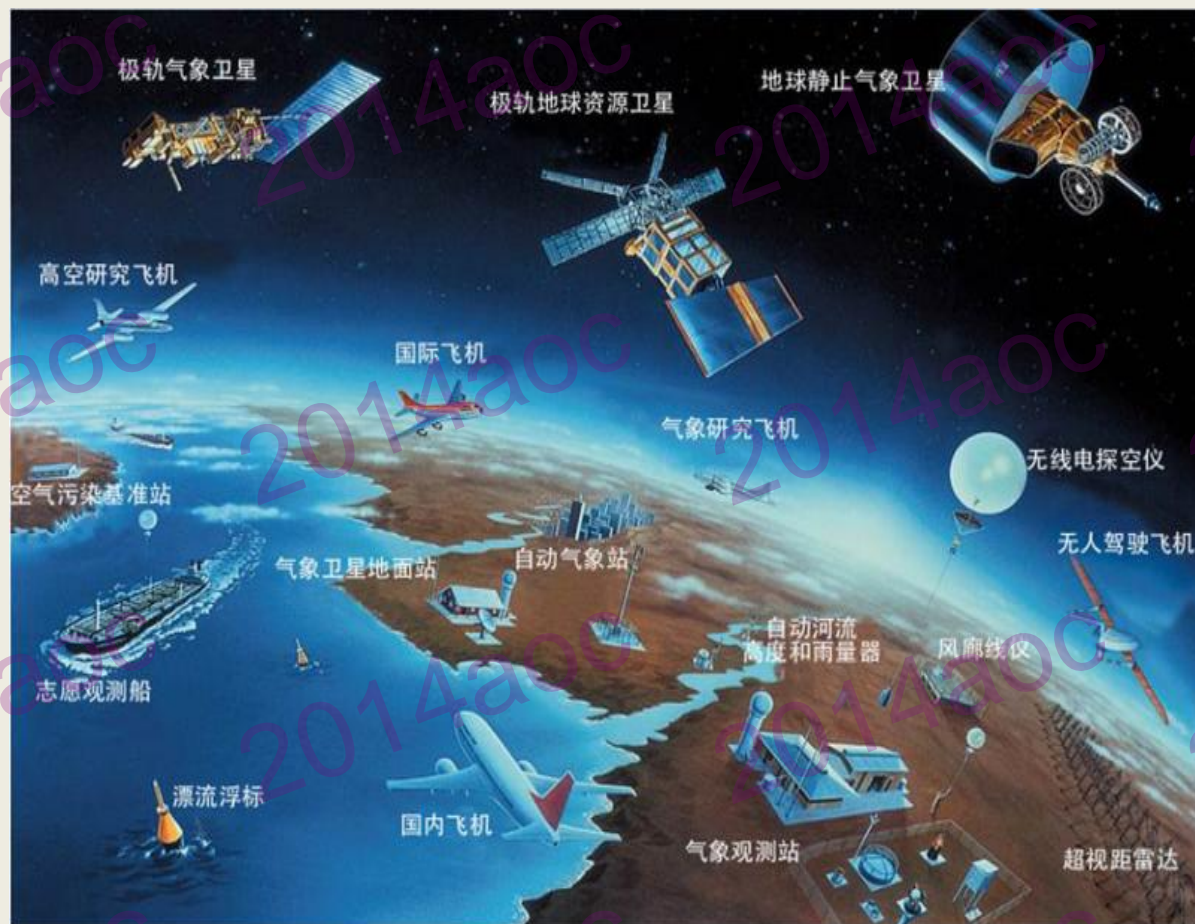


无线网络传输推动了农业信息的数据的产生和传播，由此产生的数据及增长速度将比历史上的任何时期都要多，都要快。

# 1、气象数据

目前，每年的气象数据已接近PB量级；政府间气候变化专门委员会（IPCC）第四次评估报告中采用了超过了29,000个观测资料序列，第五次报告中使用的数据量达到了2.3PB。

## 全球气候观测系统（Global Climate Observing System）

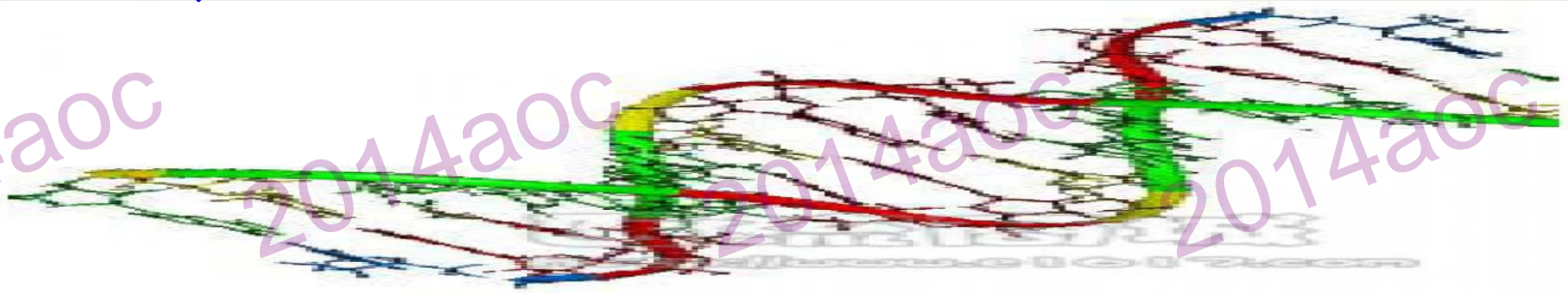


## 2、生物信息数据

1984年，由NIH创办的基因数据库（GenBank）；

截止2011年4月，在传统数据存储区，共有1.35亿条序列记录，包含1265亿碱基对；而在WGS（Whole Genome Shotgun）分区内包含1914亿个碱基对，6272万个基因组序列。

目前，每年全球产生的生物数据总量已达EB级。



### 3、资源环境数据

对地观测（遥感）技术农业资源研究提供空间数据，而且，遥感监测一直向高时空分辨率、高光谱、多频段方向发展，数据生产能力越来越强。

包括：农业小环境、土地利用/覆被、土壤物理/化学、数字高程模型等方面。





## 4、作物生长监测数据：

设备实现全天候实时监测、监测指标的变多、监测点指数级增长，所产生的数据快速增长。

仪器监测每10分钟采集1次数据（60多个指标\*52万次/年）+市场监测数据+统计监测数据+视频监测数据 → >15G/年



$$y_w = \sum_i^n y_{wi} (T_j, R_j, L_j, S_j, \dots)$$

## 5、农业统计数据:

### 农业统计

中国统计年鉴

中国农村统计年鉴

中国农村住户调查年鉴

全国农产品成本收益资料

中国饲料工业年鉴

中国农业发展报告

.....



953个品种，全国1000家市场，每日价格等11个指标的信息，每天超过1000万个数据.....

## 农业领域的“大数据”特征：

➤ **数据来源更多更广**：调查数据、仪器监测数据、传感器监测数据、模拟数据、分析数据……；

➤ **数据类型更多**：数值、文本、声音、图像、html……；

➤ **数据体量更大**： $b \rightarrow K(10^3) \rightarrow M(10^6) \rightarrow G(10^9)$   
 $T(10^{12}) \rightarrow P(10^{15}) \rightarrow E(10^{18}) \rightarrow Z(10^{21}) \rightarrow Y(10^{24}) \rightarrow ?$

➤ **数据价值密度更低**：数据呈几何级增长，有用的信息增长相对慢

➤ **数据更新更快**：年  $\rightarrow$  季  $\rightarrow$  月  $\rightarrow$  旬  $\rightarrow$  日  $\rightarrow$  时  $\rightarrow$  分

### 三、农业监测预警对大数据的需求

## (一) 从信息流理论视角看农业监测预警

农业信息流是指农业系统中反映农业诸多要素表征在各环节变化运动过程中形成的流动轨迹。



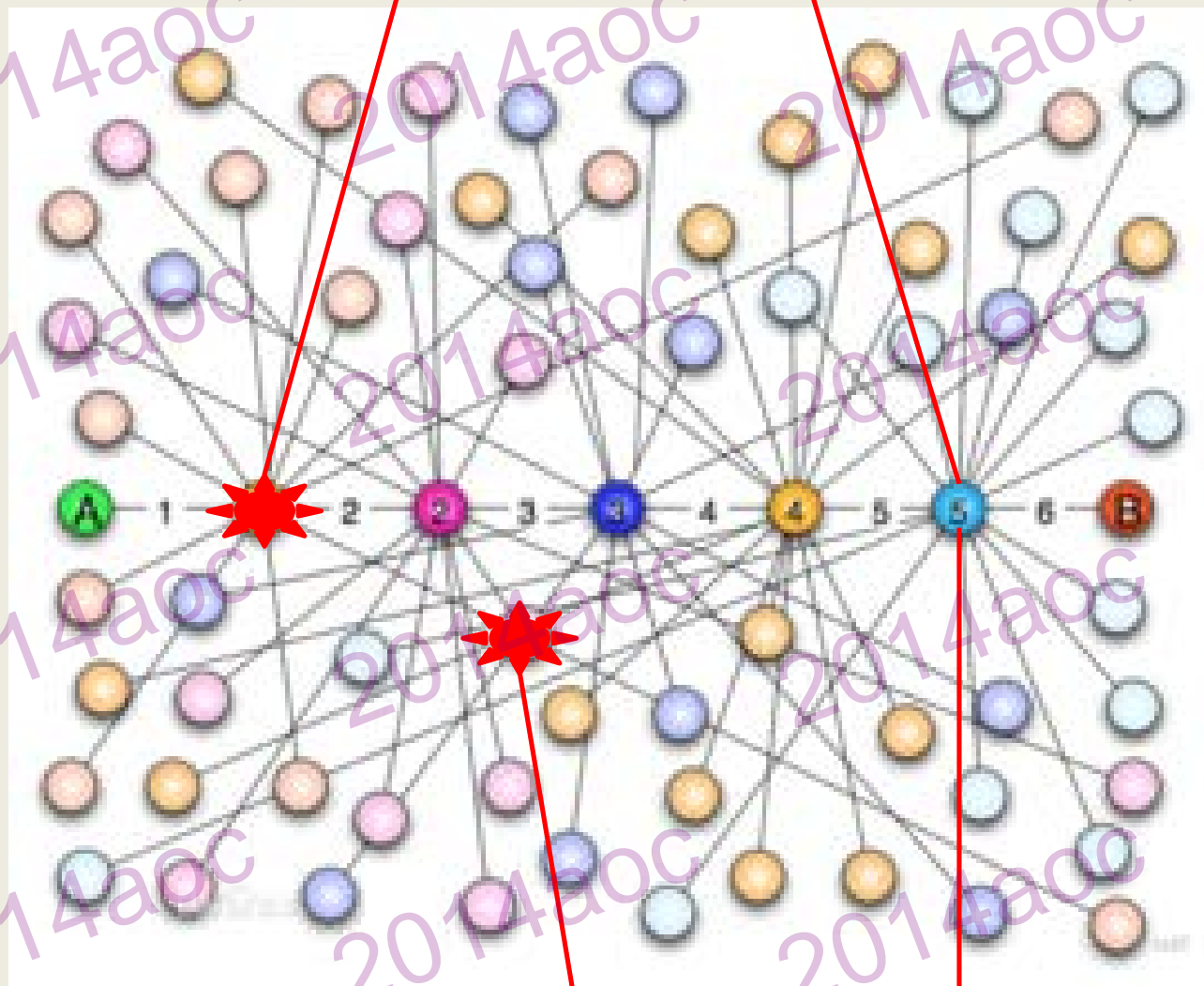


# 信息流与农业监测预警



- ◆ 农业信息流存在数量、方向、速度、目标，并控制着人和物的有目的、有规律的活动。
- ◆ 信息流有规律、有方向的流动，勾画出农业系统活动的基本流程与秩序，**为开展农业监测预警提供了理论可能与基本方法论。**

通过监测田头价格，预警零售价格



猪粮比

猪肉供需

农业监测预警

关键节点、关键  
信息的“关联”

强大的数据支撑

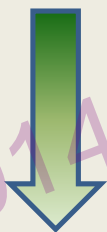
## 二、当前的不足

信息流  
三要素

信源

信道

信宿



存在  
问题

基础数据支撑  
不足

数据处理方法  
欠缺

信息服务不够  
精准

## 1、基础数据支撑不足

- **数据粒度不够：**时空尺度、品种分类等方面过于笼统，不能支撑精准调控；
- **数据匹配性不够：**农产品市场信息相对缺乏；
- **数据时效性不够：**生产、消费、贸易等信息发布滞后；
- **标准化不够：**不同部门的数据难以融合，联动性差。

## 2、数据处理方法欠缺

- 数据的融合技术需要突破

不同格式、解决语义冲突

- 数据处理、反应滞后

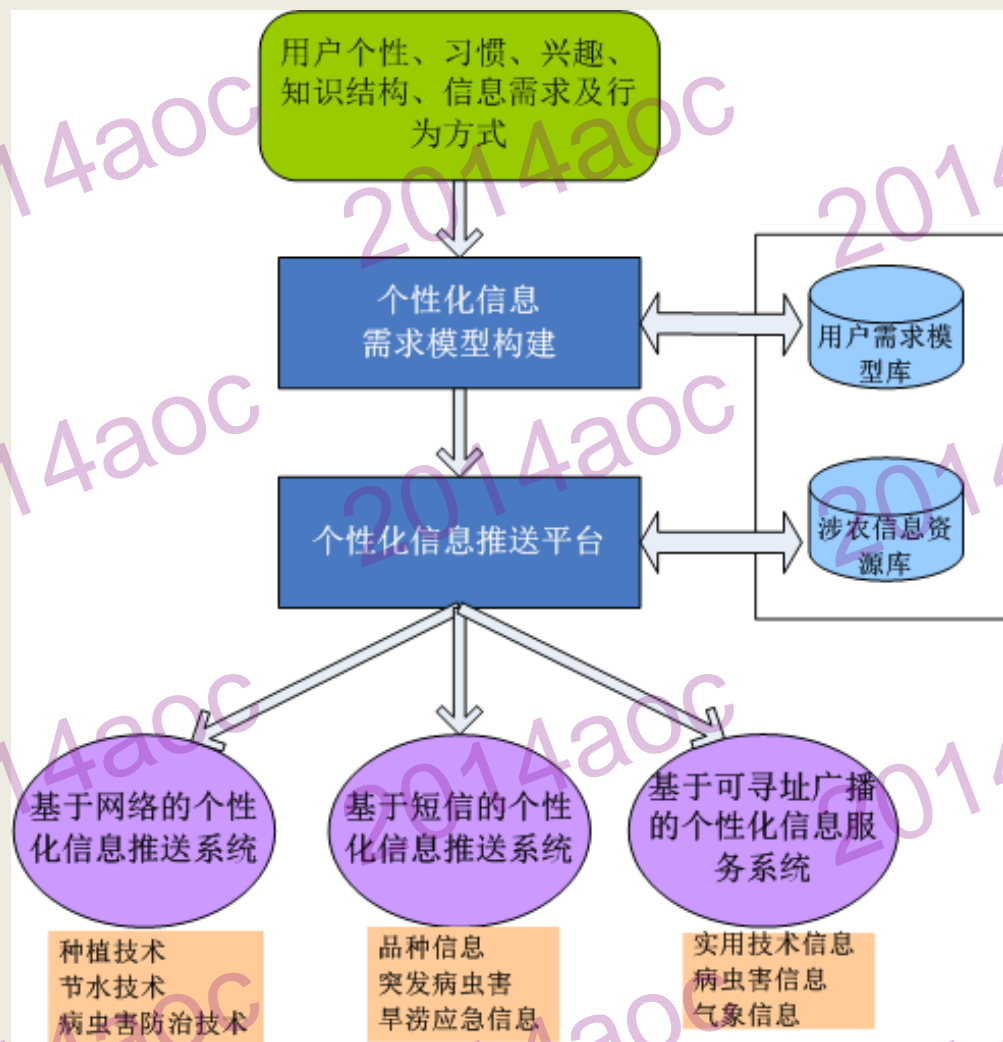
对警情的实时跟踪、同步反应

- 数据处理的智能化不足

数据庞大、用户众多



### 3、信息服务不够精准



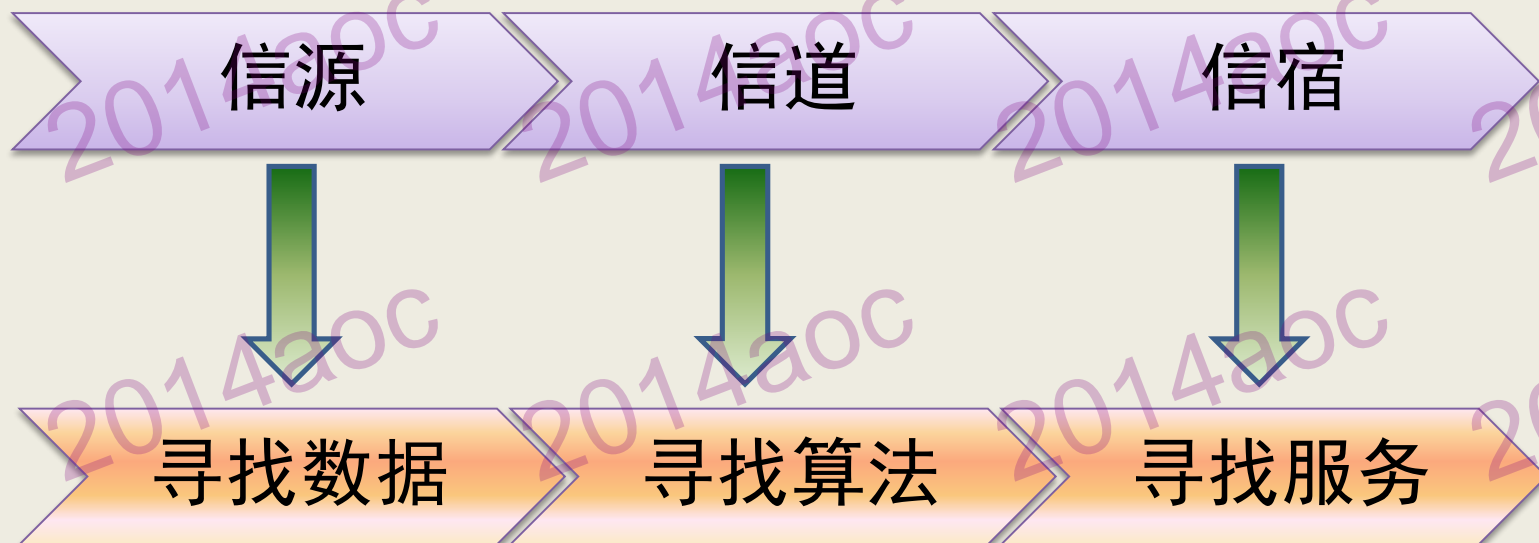
- 内容针对性不够
- 方式个性化差
- 平台服务能力相对有限

## 四、大数据背景下的 农业监测预警

# 主要做什么？

信息流  
三要素

关键  
工作



1. 寻找数据：建立粒度细化、标准统一、实时动态的农业基准数据库。

➤ 粒度细化

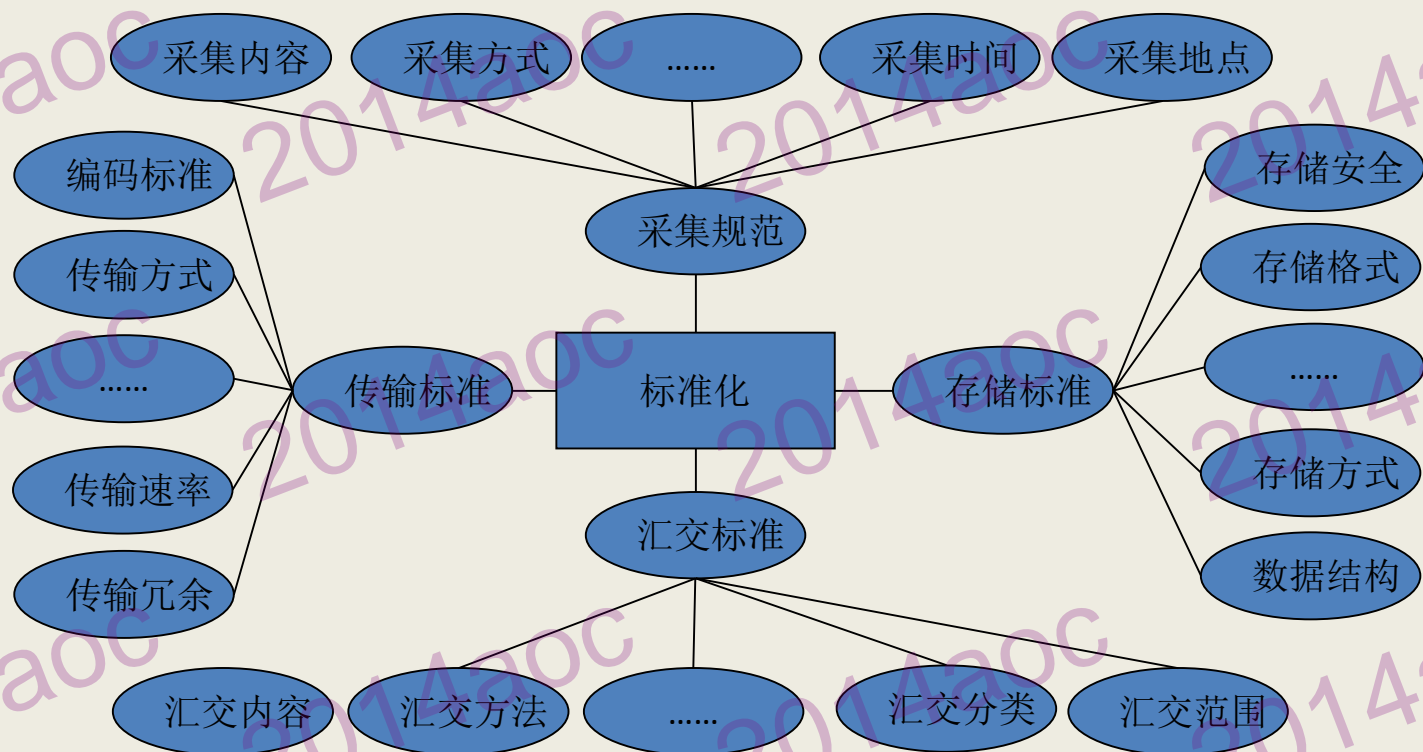
	原有	要求
工作形式	常规监测	常规监测＋热点跟踪
品种结构	“菜篮子”产品	大宗农产品＋小品种产品
内容要求	总体供求	产业链全过程监测
时间尺度	年度、季度	月、旬、周、日、时
空间尺度	全国	省域、市域、县域、田块

监测与预警对象不断“细化”：

分析品种细化、内容要求细化、预警区域细化、监测周期细化

## ► 标准统一

主要包括采集、传输、存储、汇交的标准规范。



涵盖农产品信息多个维度，多环节前后衔接的标准  
体系下的基准农业数据库



## ➤ 实时动态

- 传感器的数据采集精度、准确度和标准化将提高；
- 数据采集成本更低，采集范围（品种、领域）更大；
- 实时监控能力更强。



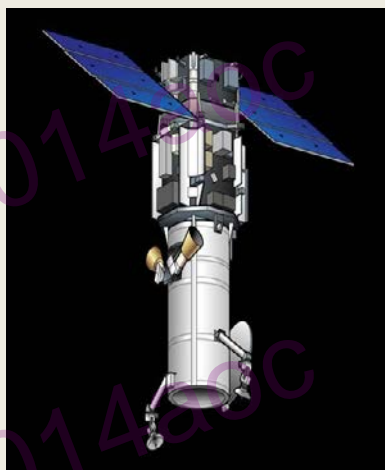
土壤盐分传感器



作物生长微纳传感器



温度传感器



卫星遥感传感器

- 传感器技术应用更加广泛，如气象信息采集中的温度传感器、光照传感器、水分传感器等；遥感监测中的光学传感器、远红外传感器、微波传感器等；
- 物联网技术的发展，将推动农业大数据信息的实时获取、传输，在农业小环境的温度、湿度、光照、降雨量，土壤养分含量、pH值等方面广泛应用。

## 自然资源类数据

IDC预测，  
2020年占40%



便携式市场信息采集终端，实现市场信息的即采即报，支持的农产品品种近100种，采集指标超过10个。

**社会经济类信息**

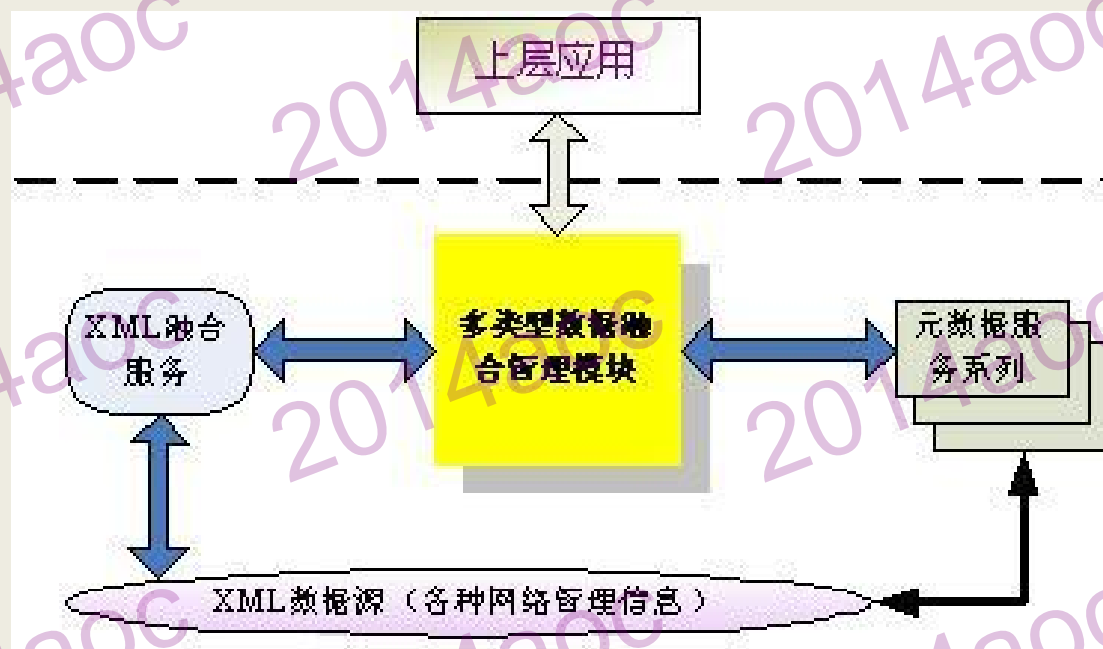
## 2. 寻找算法：建立支撑数据融合、关联、智能、实时、集成的技术体系与模型系统

### ➤ 融合

决策层

特征层

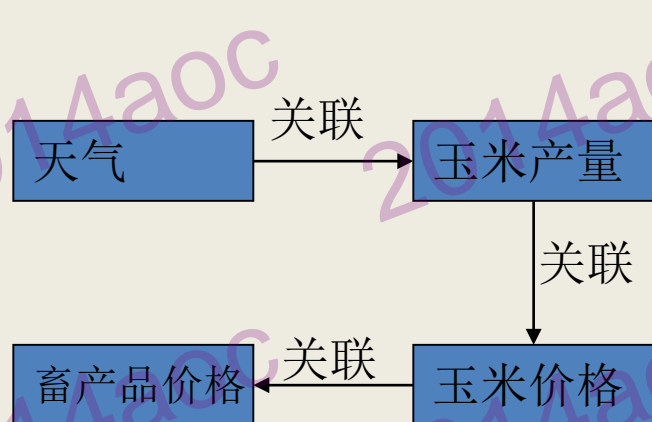
数据层



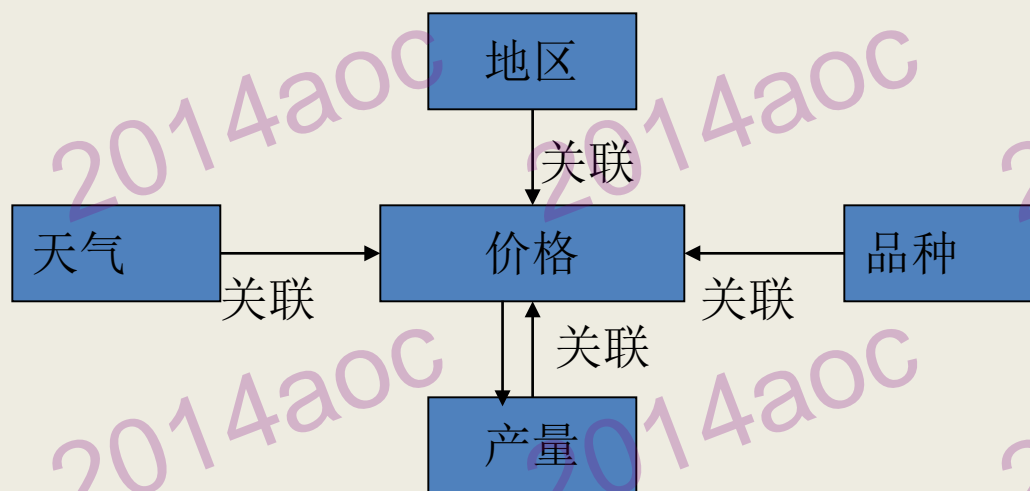
解决数据的语义冲突、建立数据之间关联，实现农业生产环境信息、农业社会经济信息的多元融合。

## ➤ 关联

数据关联是指搜索事务数据库中的所有细节或事务, 从中寻找重复出现概率很高的模式或规则, 其可分为一维数据关联和多维数据关联



一维数据关联



多维数据关联

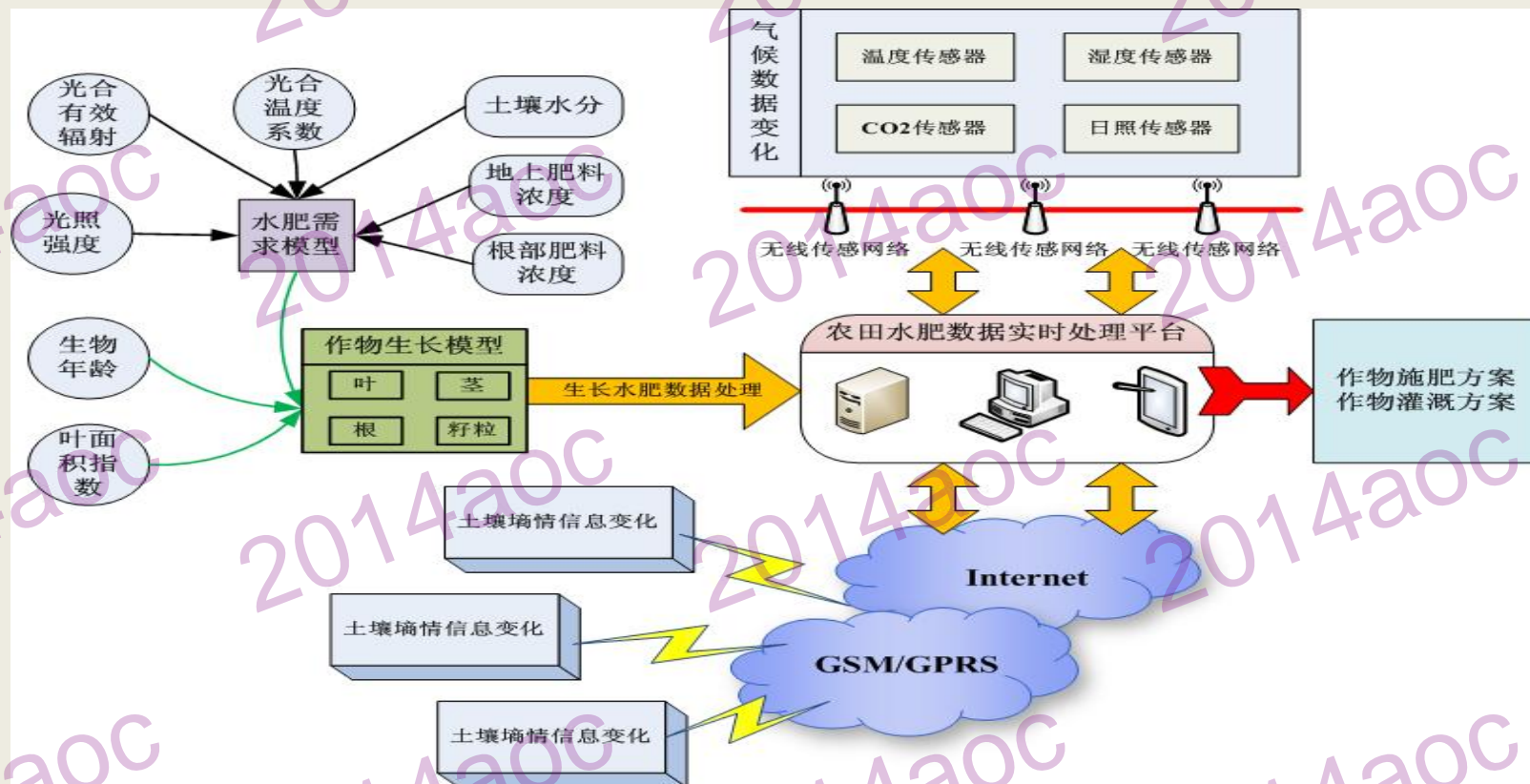
## ➤智能

农业信息智能处理，借助现代信息技术手段，分析方法由过去侧重专家经验判断为主向重视数据分析、模型分析以及计算机模拟与智能判断相结合。

	一般信息分析预警	智能信息分析预警
处理数据量	受人类本身限制	极大提高速度，有效弥补人工分析的不足
信息管理	人工管理与笔记备忘相结合	[通用/专用] 管理信息系统
分析方法	分析模型	分析模型与计算智能相结合
分析速度	正常，及时	实时，同步
结果判断	模型结果、专家判断为主	智能化模拟、智能判断
分析主题	较广泛	主题更明确



## 实时



农田水肥实时处理技术：为作物生产提供科学施肥和节水灌溉的实时智能处理方案，为亩产量节约肥料、水。

美国

- |    | A                    | B   | C            | D   | E        | F         | G         | H        | I        | J        | K        | L        | M        | N        | O        |
|----|----------------------|---|--------------|---|----------|-----------|-----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 1  |                      | CHINA   |              | UNITS   | kt 80/81 | mkt 81/82 | mkt 82/83 | kt 83/84 | kt 84/85 | kt 85/86 | kt 86/87 | kt 87/88 | kt 88/89 | kt 89/90 | kt 90/91 |
| 2  | Code                 | Variable                                      | Contribution |   | cal 1981 | cal 1982  | cal 1983  | cal 1984 | cal 1985 | cal 1986 | cal 1987 | cal 1988 | cal 1989 | cal 1990 | cal 1991 |
| 42 |                      |   | ----- If     | then price is endogenous; if U, then exogenous. |          |           |           |          |          |          |          |          |          |          |          |
| 43 | <b>MACRO</b>         | <b>MACRO</b>                                  |              |   |          |           |           |          |          |          |          |          |          |          |          |
| 44 | <b>Standard Glob</b> | Standard Global Variables (from Macro Team):  |              |   |          |           |           |          |          |          |          |          |          |          |          |
| 45 | ICGDP                | Industrial country GDP                        |              | Bill. Real US\$                                 | 17400    | 17407     | 17919     | 18724    | 19381    | 19969    | 20620    | 21541    | 22340    | 22927    | 23121    |
| 46 | ICGDPGR              | Growth rate                                   |              | Annual % char                                   | 1.6      | 0.0       | 2.9       | 4.5      | 3.5      | 3.0      | 3.3      | 4.5      | 3.7      | 2.6      | 0.1      |
| 47 | USDEF                | US GDP deflator                               |              | 2005=100  | 46.57    | 50.81     | 55.57     | 58.96    | 61.29    | 63.59    | 65.52    | 66.98    | 68.79    | 71.15    | 73.8     |
| 48 | USDEFGFR             | Growth rate                                   |              | Annual % char                                   | 8.3      | 9.1       | 9.4       | 6.1      | 3.9      | 3.8      | 3.0      | 2.2      | 2.7      | 3.4      | 3.1      |
| 49 | USCPI                | US CPI  |              | 2005=100  | 46.56    | 49.42     | 51.01     | 53.19    | 55.08    | 56.14    | 58.19    | 60.95    | 63.48    | 66.90    | 69.7     |
| 50 | USCPIGR              | Growth rate                                   |              | Annual % char                                   | 6.2      | 6.2       | 3.2       | 4.3      | 3.6      | 1.9      | 3.7      | 4.1      | 4.8      | 5.4      | 4.1      |
| 51 | PROIL                | Oil price (from ref. price file)              |              | CUS\$/bbl                                       | 70.87    | 60.49     | 50.82     | 48.28    | 43.79    | 22.22    | 27.93    | 21.73    | 26.01    | 30.13    | 25.0     |
| 52 | PROILGR              | Growth rate                                   |              | Annual % change                                 |          | -14.6     | -16.0     | -5.0     | -9.3     | -49.3    | 26.0     | -22.4    | 19.7     | 15.9     | -17.1    |
| 53 |                      |   |              |   |          |           |           |          |          |          |          |          |          |          |          |
| 54 |                      | Standard Country Variables (from Macro Team): |              |   |          |           |           |          |          |          |          |          |          |          |          |
| 55 | POP                  | Population                                    |              | Millions  | 997.0    | 1012.5    | 1028.4    | 1042.8   | 1058.0   | 1074.5   | 1093.7   | 1112.9   | 1130.7   | 1148.4   | 1163.    |
| 56 | POPGR                | Growth rate                                   |              | Annual % char                                   | 1.6      | 1.6       | 1.6       | 1.4      | 1.5      | 1.6      | 1.8      | 1.8      | 1.6      | 1.6      | 1.1      |
| 57 | DGDP                 | Domestic GDP                                  |              | Mil CLCU  | 233817   | 255103    | 282902    | 325911   | 369904   | 402465   | 449143   | 499892   | 520389   | 540168   | 58985    |
| 58 | DGDPGR               | Growth rate                                   |              | Annual % char                                   | 9.1      | 9.1       | 10.9      | 15.2     | 13.5     | 8.8      | 11.6     | 11.3     | 4.1      | 3.8      | 9.1      |
| 59 | PCD/GDP              | Per capita GDP, local                         |              | CLCU/Capit                                      | 235      | 252       | 275       | 313      | 350      | 375      | 411      | 449      | 480      | 470      | 50       |
| 60 | PCD/GDPGR            | Growth rate                                   |              | Annual % change                                 |          | 7.4       | 9.2       | 13.6     | 11.9     | 7.1      | 9.6      | 9.4      | 2.5      | 2.2      | 7.1      |
| 61 |                      |   |              |   |          |           |           |          |          |          |          |          |          |          |          |
| 62 | DOMDEF               | Dom GDP deflator                              |              | 2005=100  | 26.08    | 26.03     | 26.31     | 27.60    | 30.39    | 31.79    | 33.40    | 37.45    | 40.75    | 43.06    | 45.9     |
| 63 | DOMDEFGR             | Growth rate                                   |              | Annual % char                                   | -0.2     | -0.2      | 1.1       | 4.9      | 10.1     | 4.6      | 5.1      | 12.1     | 8.8      | 5.7      | 6        |
| 64 | DOCPI                | Domestic CPI                                  |              | 2005=100  | 24.19    | 24.72     | 25.15     | 25.69    | 27.22    | 29.48    | 31.57    | 35.20    | 43.24    | 46.57    | 47.9     |
| 65 | DOCPIGR              | Growth rate                                   |              | Annual % char                                   | 2.2      | 2.2       | 1.7       | 2.1      | 6.0      | 8.3      | 7.1      | 11.5     | 22.8     | 7.7      | 3        |
| 66 | NER                  | Nominal Exchange Rate                         |              | LCU/US\$  | 1.62     | 1.89      | 2.07      | 2.60     | 3.12     | 2.94     | 3.45     | 3.72     | 3.77     | 4.78     | 5.3      |
| 67 | NERGR                | Growth rate                                   |              | Annual % char                                   | 0.0      | 16.5      | 9.9       | 25.6     | 19.7     | -5.8     | 17.6     | 7.8      | 1.2      | 22.0     | 11.1     |
| 68 | RER                  | Real Exchange Rate                            |              | LCU/US\$  | 3.28     | 3.78      | 4.01      | 4.80     | 5.94     | 6.57     | 6.86     | 6.40     | 5.53     | 6.87     | 7.74     |
| 69 | RERGR                | Growth rate                                   |              | Annual % char                                   | 15.3     | 15.       |           |          |          |          |          |          |          |          |          |

# 中国农产品市场监测预警系统——CAMES

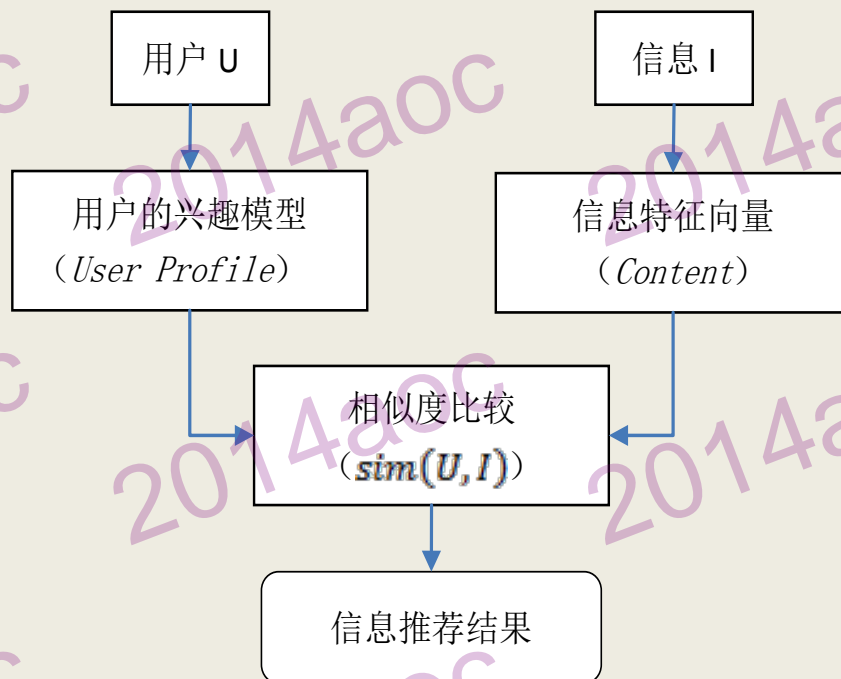
- 实现了作物机理模型与农经模型的有机统一；
- 监测预警的产品覆盖面广；
- 涵盖的区域可以到县级；
- 具有生产、消费、价格、贸易等监测预警多功能性。



**3. 寻找服务：**根据用户知识结构、信息接收方式等特点，确定适宜内容、选择合理方式开展有针对性信息服务。

➤ 寻找内容

通过分析用户需求与数据本身内容，建立匹配模型，选择适宜服务内容。



## ➤寻找方式

要通过分析用户的知识背景、需求目标等决定信息以何种方式实现精准服务。

### 信息推送方式

推送方式 用户	频道推送	邮件推送	短信推送	语音推送
农业管理部门用户	知识型、一般消息型	知识型、一般消息型	应急消息型	
农业科研教学相关人员	基础数据型、知识型和一般消息型	基础数据型、知识型和一般消息型	应急消息型	
农业技术推广人员	基础数据型、知识型信息、一般消息型	基础数据型、知识型信息、一般消息型	知识型、一般消息型、应急消息型	知识型、一般消息型、应急消息型
农业生产经营者			知识型、一般消息型、应急消息型	知识型、一般消息型、应急消息型



**谢谢！  
敬请指正！**

