



北京城市实验室  
Beijing City Lab

ID of the slides

5



## Slides of BCL

[www.beijingscitylab.com](http://www.beijingscitylab.com)

### How to cite

Author(s), Year, Title, Slides at Beijing City Lab, <http://www.beijingscitylab.com>

E.g. Long Y, 2014, Automated identification and characterization of parcels (AICP) with OpenStreetMap and Points of Interest, Slides at Beijing City Lab, <http://www.beijingscitylab.com>



# 多源社会化数据 智能获取与规划应用

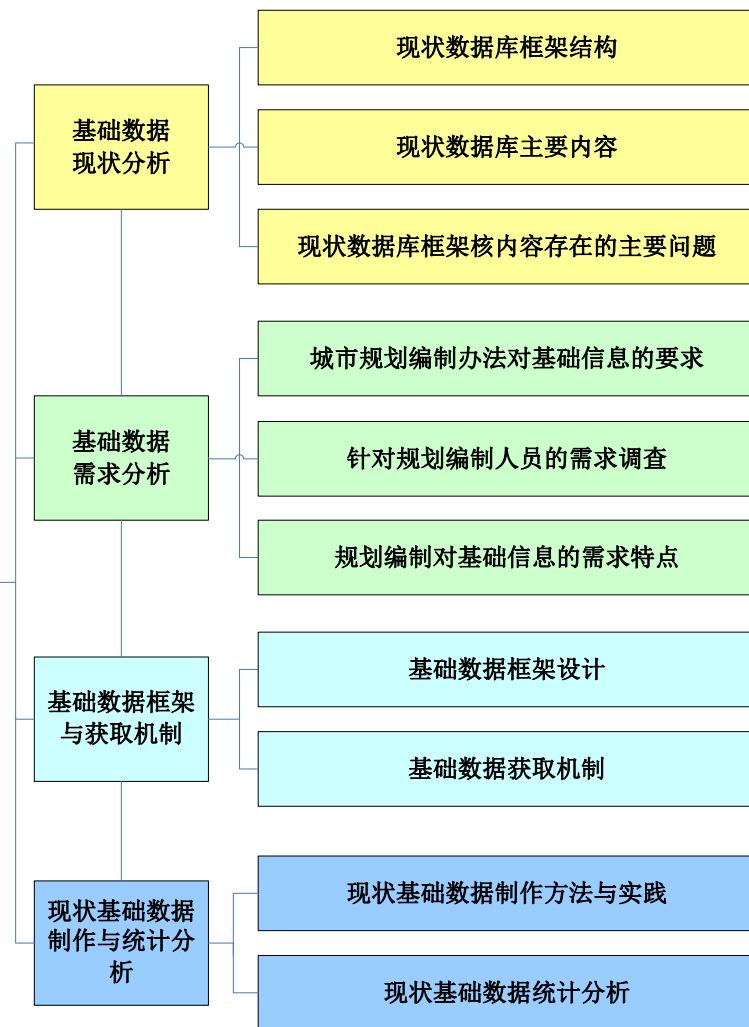
北京市城市规划设计研究院 规划信息中心

# 课题的背景1：传统数据获取途径不畅

- 统计资料
- 测绘资料
- 各委办局协调
- 小规模调查

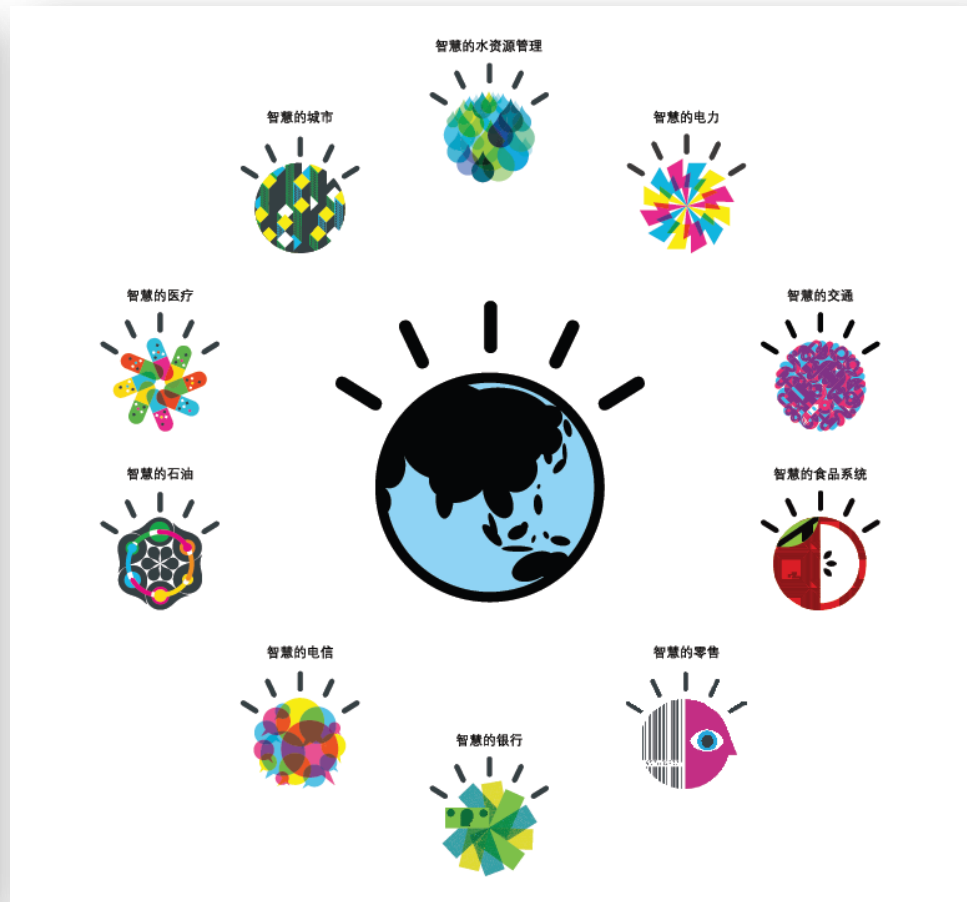


城市规划编制  
基础数据获取  
与整合机制  
研究



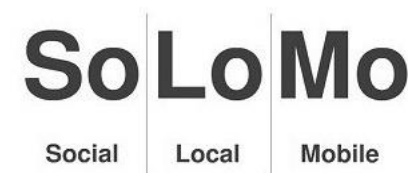


# 课题的背景2：大数据与智慧地球



# 课题的背景3：开源与众包精神

- 开源：OPEN SOURCE
  - 开源城市
  - 开源政府
  - 信息公开
- 众包：Crowdsourcing
  - 自媒体
  - VGI
  - SoLoMo





# 研究的目的

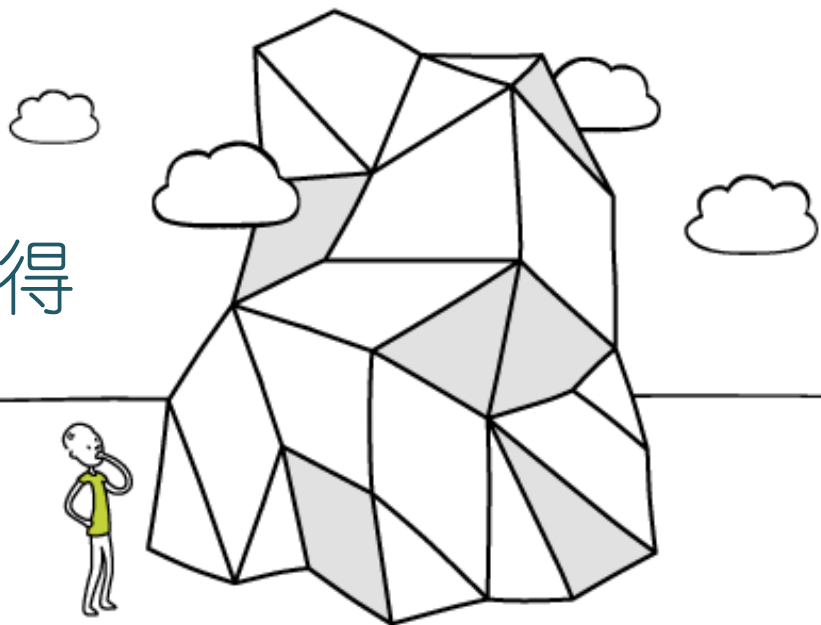
- 拓源：拓展城乡规划的数据获取渠道
- 探路：大数据应用于城乡规划的可行之路
- 播种：开拓规划从业者的眼界



Is Big Data Applicable?

# 课题的内容

- 探矿：社会化数据有什么
- 采掘：社会化数据怎么获得
- 加工：获得数据怎么处理
- 应用：社会化数据怎么用



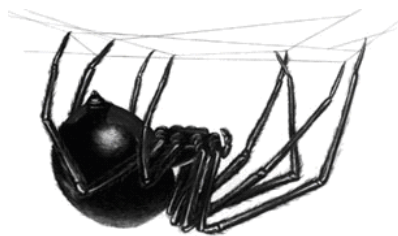


# 探矿：

专题类型	典型网站
空间	Baidu、Google、MapABC、MapBar、Sogou、OSM
房产	搜房、安居客、链家、我爱我家、sina、焦点
公司	阿里巴巴、sina等各类公司黄页
社区	北京社区网、房产网
教育	教育黄页、北京教委
医疗	医院黄页、北京卫生局
文物	北京文物局
公共设施	Aibang、dianping等网站
就业	51job, 智联招聘
社会活动	dianping、weibo、街旁
...	...



# 采掘：



Browsers



Your Websites



Source Websites



基本信息

物业类别 住宅, 公寓、公寓

建筑类别 板楼 塔楼 小高层

环线位置 三至四环

容积率 2.93 [关于容积率]

开盘时间 2004-07-20 [开盘详情]

物业费 2.49元/平米·月

开发商 北京裕泽房地产开发有限责任公司 [查看预售许可证]

预售许可证 京房售证字(2003)10号 [更多]

售楼地址 阜成路44号(肿瘤医院向东500米路南,金悦海鲜大酒店东侧)

物业地址 海淀五路居(玲珑塔北1公里),蓝靛厂南路 [交通图]

交通状况 公交:33路 地铁:6号线、10号线 [更多]

二手房销售电话:暂无数据 搜房评估参考价:元/平方米 [房价走势]

项目特色 水景地产, 中式地产

装修状况 精装修 [装修相册][建材卖场]

所属商圈 八里庄

绿化率 30% [关于绿化率]

入住时间 2004

物业公司 北京



# 采掘示例： 黄页、楼盘信息

## 项目简介

绿波漫板小区位于西三环外,昆玉河畔,北京唯一的活水河。小区整体由8栋商品楼组成,1号楼--5号楼是板楼组成,6、7、8号楼是板塔结合。小区环境幽雅,绿化环境达标,独立健身会所,超市,双向地铁交叉,贯穿北京。

## 楼层状况

共700多户。其中7号楼为14层塔楼,其中1、2层为底商,265户

## 车位信息

停车位:400多个,地上车位无,地下车位售价12万,租200元/月

## 相关信息

占地面积:37500平方米

建筑面积:110000平方米

开工时间:2003-05-20

竣工时间:2007-07-20

物业管理附加信息:2.49元/平米·月

开发商:北京裕泽房地产开发有限责任公司

投资商:北京裕泽房地产开发有限责任公司

物业管理公司:北京集兴物业管理公司

建筑单位:中国新兴保信建设总公司

代理商:北京精测行房地产经纪公司

建筑设计单位:台湾筑镇工程设计股份有限公司

户数:总户数700户 当期户数700户

安居客信息





# 采掘示例：新浪微博用户信息

verified_reason	followers	location	statuses	screen_name	description	friends_c
	12485	北京	2976	规划师石楠	资深规划师，注册规划师，以交友为事业。	572
	11225	上海	3764	孙旖文	关注城市以及城市中的人和事。 同济大学建筑与城市规划学院城市规划系	297
	275	江苏 苏州	1248	莫梦辰_殇_规划	规划学徒Tao+缘木求鱼的笨蛋	189
	72	北京	79	txtian	注册城市规划师	150
	1393	上海	7	董鉴泓	同济大学建筑城规学院教授 《城市规划学刊》主编	43
上海同异城市设计有限公司官	80	上海	177	上海同异城市设计	上海同异城市设计有限公司官方微博（2001-2012） 城市规划、城市设计	230
	219	广东 深圳	1060	MU董的一些事一些情	爱生活，爱睡觉，未来城市规划师	212
	3	广东 广州	1	丁逸聪-规划	有志者，共规划！道亦在。	11
	172	浙江 宁波	449	宁波市城市交通研究所	宁波市规划设计研究院 城市交通研究所	204
	107	湖北 武汉	22	武大牛强	武汉大学城市规划系教师，GIS迷，主攻规划信息技术。	94
	9	湖北 黄石	42	磁湖一梦	规划黄石	72
	38	河南 郑州	19	睦岚	规划工作者	297
	15	上海 杨浦区	26	丁国胜Planner	同济大学城市规划与设计博士研究生	70
	366	上海	622	同济CAUP叶海	同济大学建筑与城市规划学院副教授，建筑技术科学专业教师，国家注册	370
	22	新疆 乌鲁木齐	346	Lily刘昱	城市规划师；2002年乌鲁木齐《晨报》钻石新娘得主！	121
	62	甘肃 兰州	141	魏孔军-甘肃	甘肃省规划院	85
	148	重庆 沙坪坝	1064	hapyhq01	利川；重庆大学；规划人	263
	16	上海	51	目心木木	一个从事规划行业的菜鸟选手.....	115
	39	江西 九江	188	小马奔腾勇康	规划小兵	163
	23	广东 深圳	135	lukiar晶	城市规划师不是生活的创造者，而是形式的赋予者	256
	166	湖北 宜昌	691	里拉很安静	三峡大学城市规划学生，简简单单，安安静静，开心就好。	411
	68	广东 深圳	77	_范嵘_	蕃奥让规划行动起来	215
	17	山东 青岛	55	chen_c旭	学习阶段.....规划行业菜鸟一枚.....巨恶心英语~~~	176
	15	云南 昆明	73	唯珍卿	云南大学大四在读，城市规划专业，以后出去可能会是穷画图匠...	90
	20	吉林 长春	60	OnlyforDream_dc	专情于地理，放眼于规划！	167
	25	陕西 西安	24	zimeizi	西北大学城市规划学生	131
	24	北京 东城区	67	人地咨询	一批毕业于中国科学院、中国社会科学院等研究所的博士，以专业知识	81
《城市规划》杂志 官方微博	1488	北京 海淀区	152	城市规划杂志	《城市规划》杂志官方微博	63
中国城市规划设计研究院深圳	324	广东	16	中国城市规划设计研究院	中国城市规划设计研究院深圳分院认证微博	423
	609	北京 海淀区	337	北京万合创景集团	北京万合创景国际规划设计研究院，欢迎喜欢设计规划的朋友加盟，我们	2000
	175	北京 丰台区	15	金地停车集团招聘网	提供停车全面解决方案，商业模式贯穿整个停车产业。公司全国招聘规划	1046
*						



## 加工：

- 基于网络服务的社会化数据空间化
- 火星、百度坐标的纠偏
- 北京地方坐标的转换
- 语义消歧，数据同一化
- 多源数据复合加工
- .....



# 加工：技术难点1

## • 非结构化数据到结构化数据

### 小区简介

**基本信息**

物业类别 住宅、公寓、公寓  
建筑类别 板楼 塔楼 小高层  
环线位置 三至四环  
容积率 2.93 [【关于容积率】](#)  
开盘时间 2004-07-20 [【开盘详情】](#)  
物业费 2.49元/平米·月  
物业公司 北京集兴物业管理有限公司

项目特色 水景地产,中式地产  
装修状况 精装修 [【装修相册】](#) [【建材卖场】](#)  
所属商圈 八里庄  
绿化率 30% [【关于绿化率】](#)  
入住时间 2004-09-30 [【入住详情】](#)

开发商 北京裕泽房地产开发有限公司 [【查看预售许可证】](#)  
预售许可证 京房售证字(2003)110号 [【更多】](#)  
售楼地址 阜成路44号(肿瘤医院向东500米路南,金悦海鲜大酒店东侧)  
物业地址 海淀五路居 09号楼北1公里),蓝靛厂南路 [【交通图】](#)  
交通状况 公交:33路 地铁:6号线、10号线 [【更多】](#)  
二手房销售电话:暂无数据 搜房评估参考价:元/平方米 [【房价走势】](#)

**项目简介**

绿波漫板小区位于西三环外,昆玉河畔,是北京唯一的活水河。小区整体由8栋商品楼组成,1号楼-5号楼是板楼组成,6、7、8号楼是塔楼塔楼。小区环境优美,绿化环境达标,独立健身会所,超市,双向地铁交叉,贯穿北京。

**楼层状况**

共700多户,其中7号楼为14层塔楼,其中1、2层为底商,265户

**车位信息**

停车位:400多个,地上车位无,地下车位售价12万,租200元/月

**相关信息**

占地面积:37500平方米  
建筑面积:110000平方米  
开工时间:2003-05-20  
竣工时间:2007-07-20  
物业管理附加信息:2.49元/平米·月  
开发商:北京裕泽房地产开发有限公司  
投资方:北京裕泽房地产开发有限公司  
物业管理公司:北京集兴物业管理有限公司  
建筑单位:中国新兴保信建设总公司  
代理商:北京精润行房地产经纪公司  
建筑设计单位:台湾筑诚工程设计股份有限公司  
户数:总户数700户 当期户数700户

### 置业专家

在线咨询



李宇坤  
13661258553  
我爱我家



韦青  
13718142776  
链家地产



王莹  
18801167533  
麦田房产

[绿波漫板地图](#) [查看大图](#)

起点  终点



**相关资讯** [更多>>](#)

- 海淀尾房绿波漫板 精装小一居现房 2008-10-13
- 海淀尾房在售 绿波漫板一居均价1 2008-08-15
- 海淀精装小户型绿波漫板 现房44平 2008-07-17
- 海淀区尾房最新动态 一二层选择余 2008-08-21
- 业主“忽略”了!跳水降50万抛售三 2012-02-09
- 北京楼市最具潜力五大“价值洼地” 2009-11-08

**便民信息** [更多>>](#)

今天星期一 限行尾号3和8

绿波漫板社区邮编 邮编:100142  
北京利康货运搬家公司 4008175588-5124  
北京玉中建设运输队 010-87013344  
北京利达万通搬家公司 010-57267778 010-57267779  
北京市非紧急救助热线 12345

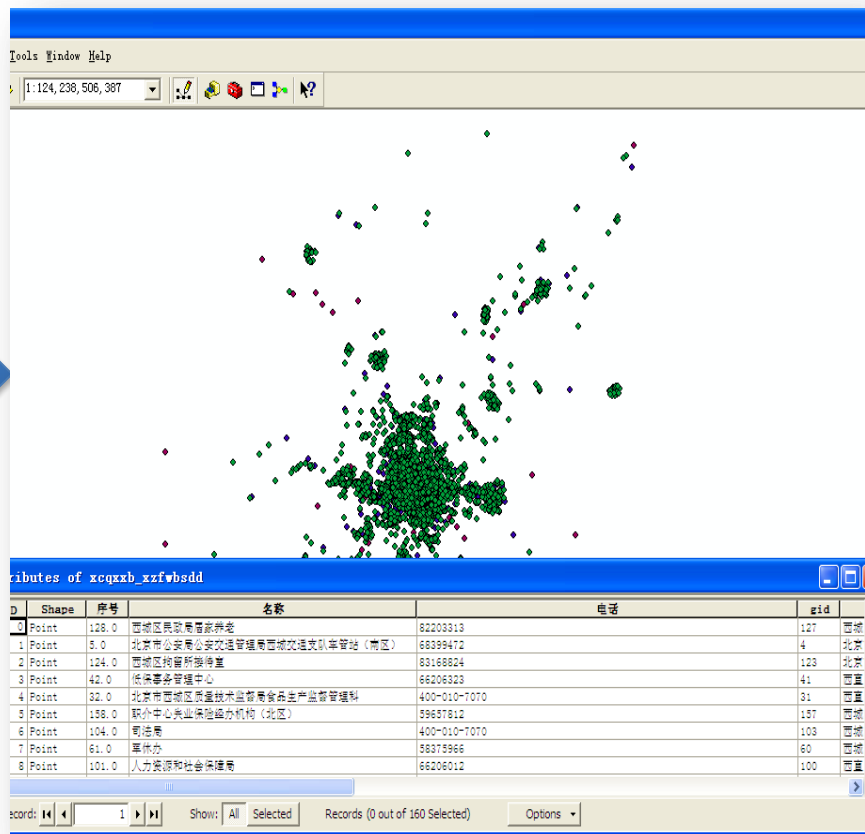


序号	区域	小区名称	物业公司	总户数	停车位	绿化率	容积率	总建筑面积	小区地址
1	昌平区北	东昌亚平上区	房修一物业公司	3967户	1549	30.10%	2.5	238237.3平方米	回龙观西大街9号院
2	昌平区	矩昌阵平小区	唐人物业管理有限公司	2078户	11030	35%	1.58	230000平方米	京昌高速公路回龙观出口正东3公里处
3	昌平区	天通苑东一区	方佳物业管理有限公司	7745户	89035	35%	1.1	490000平方米	立水桥北太平庄东路
4	昌平区	昌新平龙	新龙集团物业管理公司	6000户	35040	40%	2.46	600000平方米	龙泽苑城地铁站南侧200米

# 加工：技术难点2

- 从文本数据到空间数据

序号	区域	小区名称	物业公司	总户数	停车位	绿化率	容积率	总建筑面积	小区地址
1	昌平区北	东亚平上	房修一物业公司	3967	1549	30.10%	2.5	238237.3平方米	回龙观西大街9号院
2	昌平区	矩昌阵平	唐人物业管理	2078	1100	35%	1.58	230000平方米	京昌高速公路回龙观出口正东3公里处
3	昌平区	天通苑东	方佳物业管理	7745	8903	35%	1.1	490000平方米	立水桥北太平庄东路
4	昌平区	昌平新平	新龙集团物业管理	6000	3500	40%	2.46	600000平方米	龙泽苑城地铁站南侧200米





# 加工：基于网络服务的社会化数据空间化

Baidu 地图 | API



MapABC地图服务平台新版发布  
荣获奖项 优惠促销

MapABC API 2

搜狗地图API升级为 2.5

2.5新特性:

- 代码动态加载
- 初始化加载代码更小
- 速度更快

API 阿里云地图 上线啦!

图吧地图 API

极速加载-精准定位

SOSO 地图 API

新鲜发布!

轻快小巧 简单易用

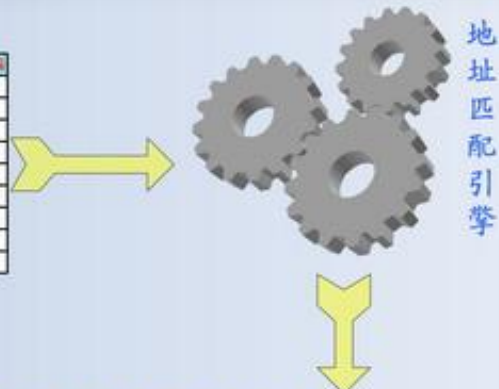
天地图

国家测绘地理信息局 | 关于天地图 | 更新日志 | 意见反馈 | 帮助

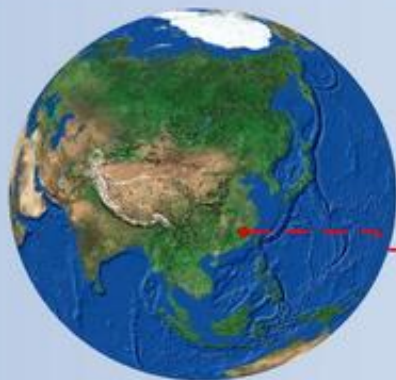
在线地图 手机地图 地图API 省市直通 专题应用 服务资源

地址列表

序号	地址	联系电话	负责人	邮编
1	福建省福州市台江区瀛洲街道瀛东社区红星新村92幢			
2	福建省福州市台江区上海街道上海新村117号			
3	福建省福州市鼓楼区华林路11号西湖宾馆2号楼			
4	福建省福州市晋安区新店镇健康村518号			
5	福建省福州市晋安区六一北路50-52号			
6	福建省福州市晋安区六一北路126号洋下新村126号			
7	福建省福州市鼓楼区五四北路273号			
8	福建省福州市鼓楼区五四路313号融侨花园			
9	福建省福州市鼓楼区五一北路1号天马广场			



地址匹配引擎



带经纬度坐标的地址列表

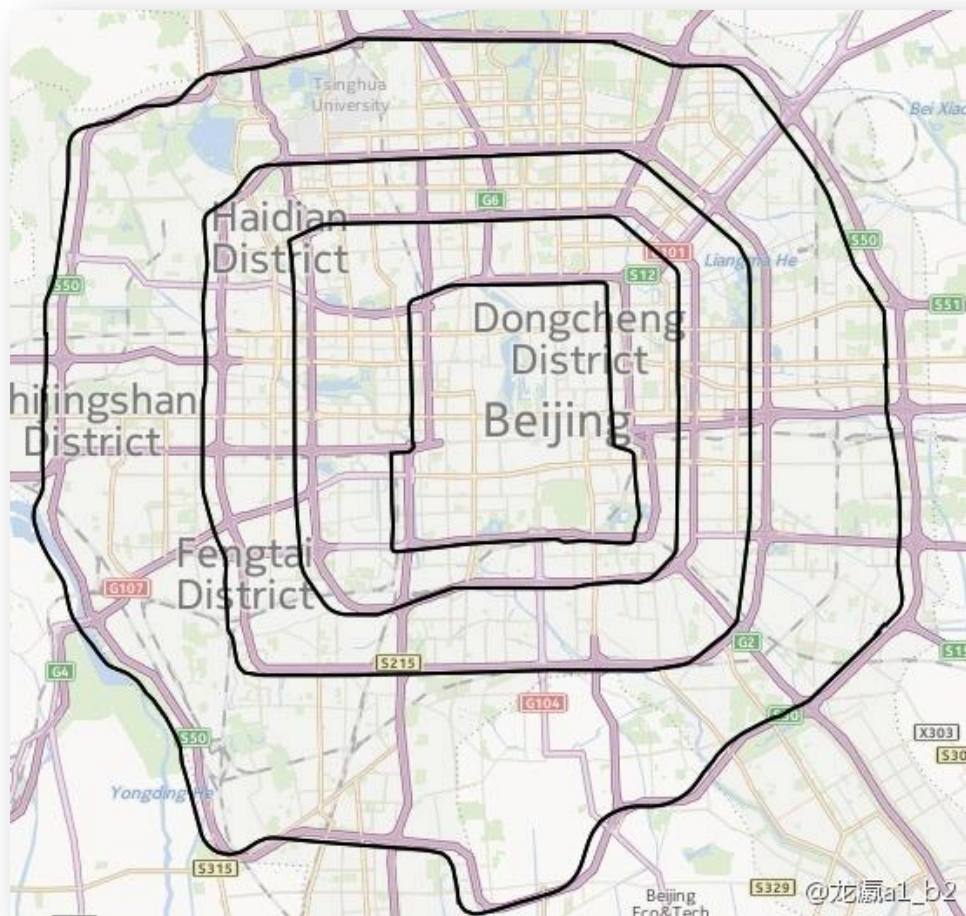
序号	地址	经纬度坐标
1	福建省福州市台江区瀛洲街道瀛东社区红星新村92幢	119.455475, 26.879239
2	福建省福州市台江区上海街道上海新村117号	119.305408, 26.079469
3	福建省福州市鼓楼区华林路11号西湖宾馆2号楼	119.306112, 26.079421
4	福建省福州市晋安区新店镇健康村518号	119.964407, 26.357841
5	福建省福州市晋安区六一北路50-52号	119.225986, 26.745893
6	福建省福州市晋安区六一北路126号洋下新村126号	119.745624, 26.079548
7	福建省福州市鼓楼区五四北路273号	119.338941, 26.587695
8	福建省福州市鼓楼区五四路313号融侨花园	119.586475, 26.755862
9	福建省福州市鼓楼区五一北路1号天马广场	119.147864, 26.685741

地址匹配流程



# 加工：技术难点3

- 火星、百度坐标纠偏：利用函数拟合纠正坐标偏差

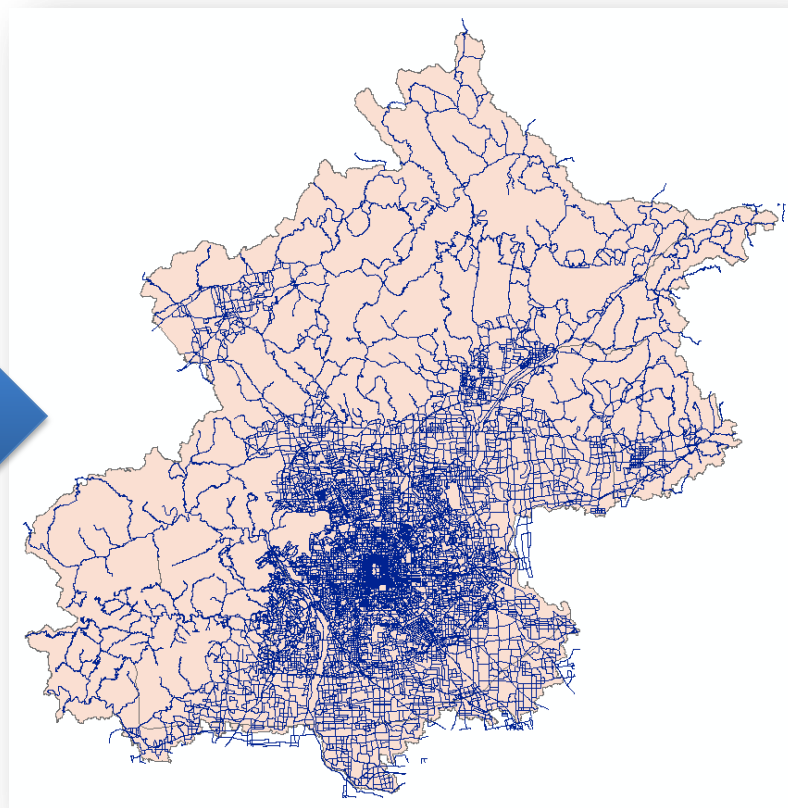
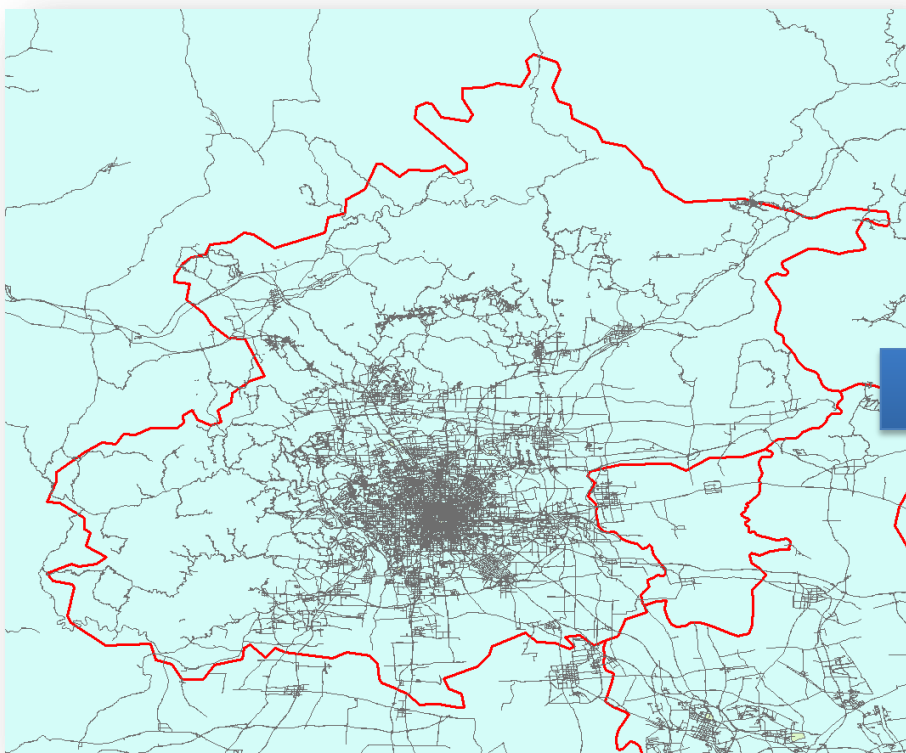


- **火星坐标**：出于安全考虑，国家测绘局要求所有互联网及导航电子地图的数据都使用其制定的加密算法将真实的坐标非线性加偏成虚假的坐标，在北京其偏移量大约有700米，其坐标系被称为火星坐标；
- **百度坐标、北京经信委坐标**：百度、经信委等则在火星坐标的基础上又进一步加偏，分别形成了自己坐标系；
- 从bjdata下载的数据及利用OpenAPI空间化的数据都是经信委坐标系、火星或者百度坐标系，必须经纠偏处理变成标准的经纬度坐标数据才可使用。



## 加工：技术难点4

- 从WGS84经纬度坐标到北京地方坐标



# 加工：技术难点5

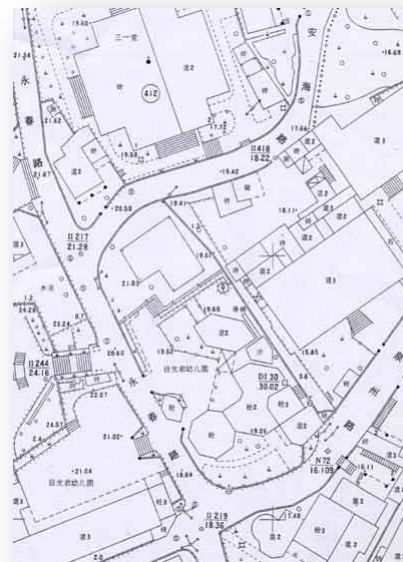
## • 数据同一化与多源复合加工

例：将房产信息，诸如门牌号码、建筑年代、户数此类，落到1：2000地形图的建筑物数据上，丰富地物属性。

**方嘉国际基本信息**

**基本信息**

<b>物业类别</b> 综合体	<b>项目特色</b> 小户型, 地铁沿线, 打折优惠楼盘
<b>建筑类别</b> 板塔结合 小高层	<b>装修状况</b> 毛坯, 公共部分精装修 [装修相册] [建材卖场]
<b>环线位置</b> 三至四环	<b>所属商圈</b> 方庄
<b>容积率</b> 2.90 [关于容积率]	<b>绿化率</b> 35.10% [关于绿化率]
<b>开盘时间</b> 预计2012年5月26日开盘 [开盘详情]	<b>入住时间</b> 2013-10 [入住详情]
<b>物业费</b> 4.90元/平方米·月	<b>物业公司</b> 北京东光物业管理有限公司
<b>开发商</b> 北京华业通置业有限公司 [查看预售许可证]	
<b>预售许可证</b> 京房售证字(2012)29号 [更多]	
<b>售楼地址</b> 丰台区方庄桥南300米路东(天天渔港对面)	
<b>物业地址</b> 丰台方庄桥东南角 [交通图]	
<b>交通状况</b> 紧邻地铁5号线、10号线、亦庄线、十号线二期的枢纽宋家庄站、成寿寺站, 2014年通车的1 [更多]	
<b>方嘉国际房价</b> 均价33000元/平方米 [房价走势] [我要纠错] [房贷计算器]	





中国网络中心北京代表处

华夏银行24小时自助银行(北京长安支行) 华夏银行(北京长安支行)



河北驴肉火烧

三里河二区4号楼

济南金惠建设服务有限公司北京分公司北京东方邮务商务调查

三里河二区(南一门)

三里河二区三里河二区1-5号楼(南门)

三里河六巷水

三里河河南横街交叉口

金凤成楼(三里河店)

蓝色城

三里河东路3号

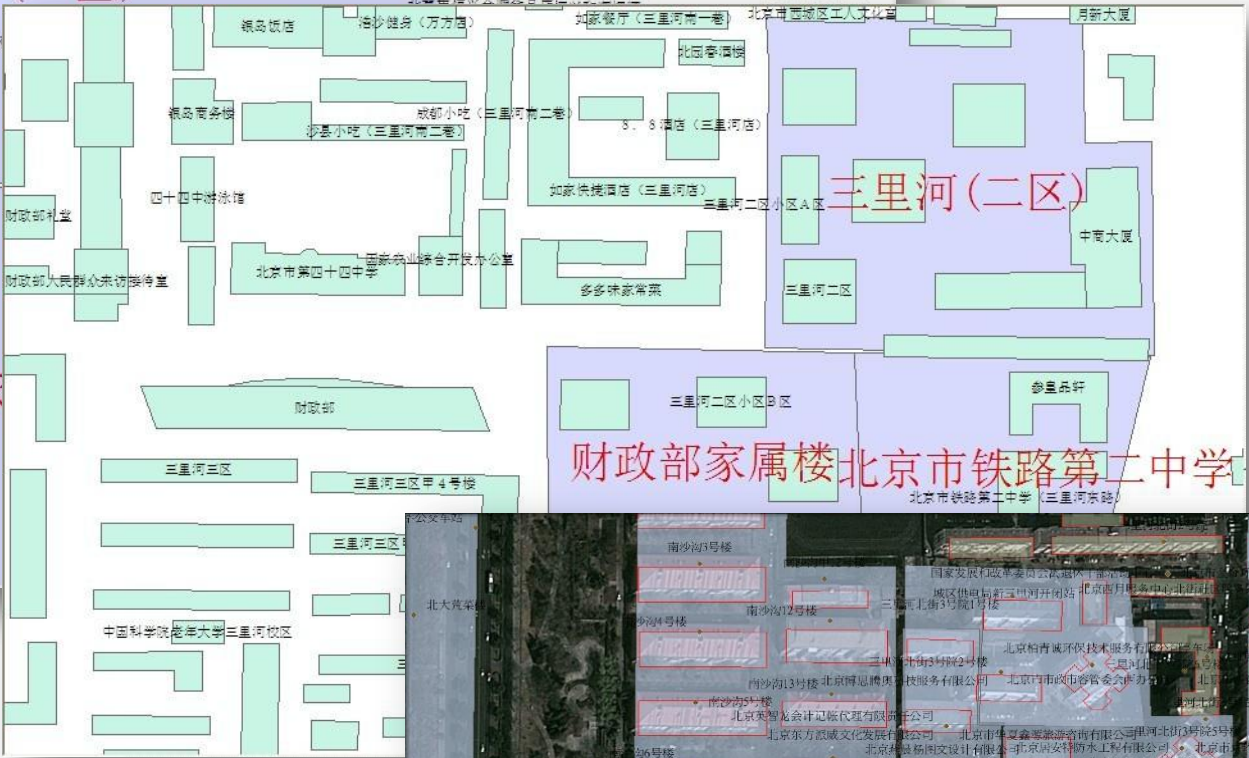
三里河二区6号楼

麻辣烤串

烤鱼吧

## 三里河(二区)

## 北京市铁



## 财政部家属楼北京市铁路第二中学

# 加工:

## 从点到建筑到用地的空间数据产品

2014/1/2

北京市

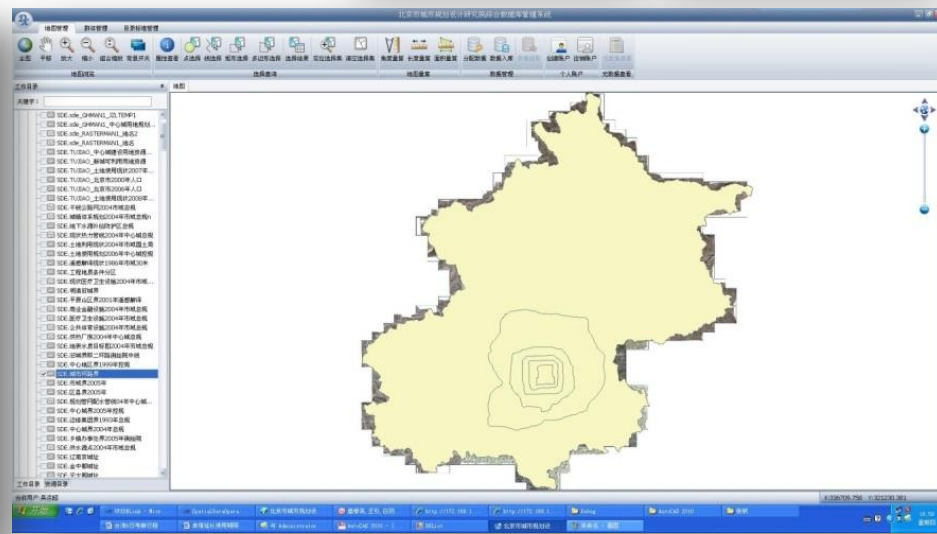
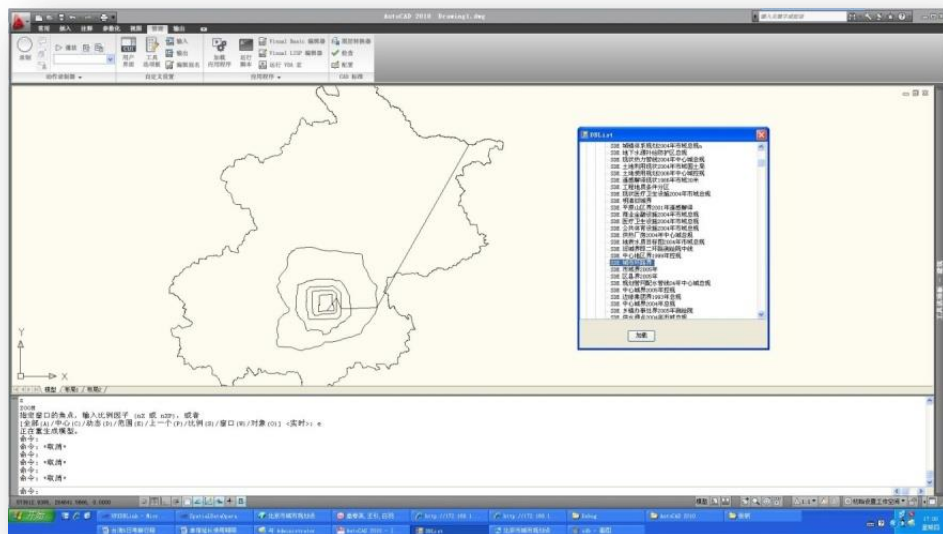
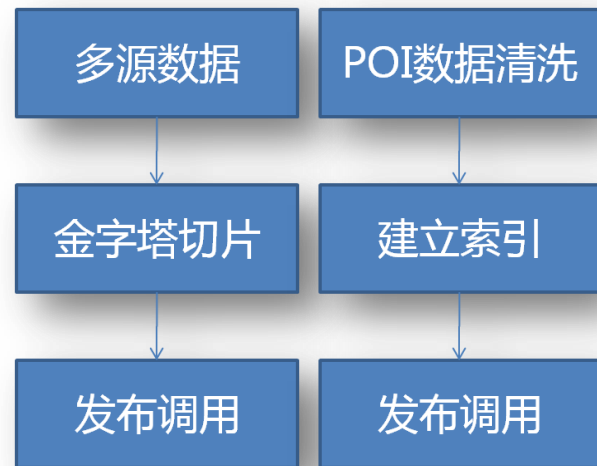


Google Earth

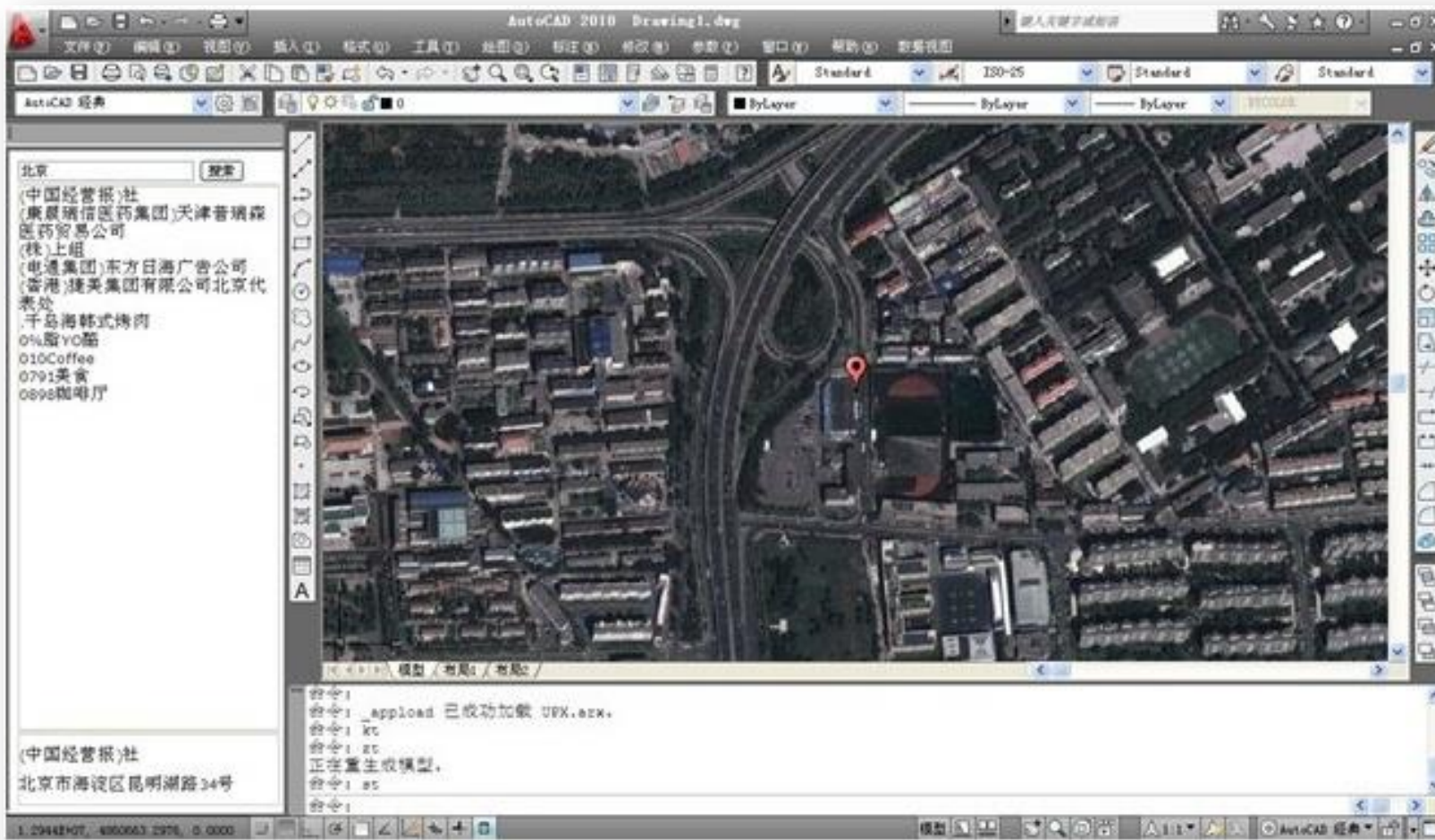


# 应用：

- 基于AutoCAD的专题图浏览
- 基于AutoCAD的定位检索
- 基于AutoCAD的空间库调用



# 应用：基于AutoCAD的定位检索





# 课题成果1：OPEN数据源

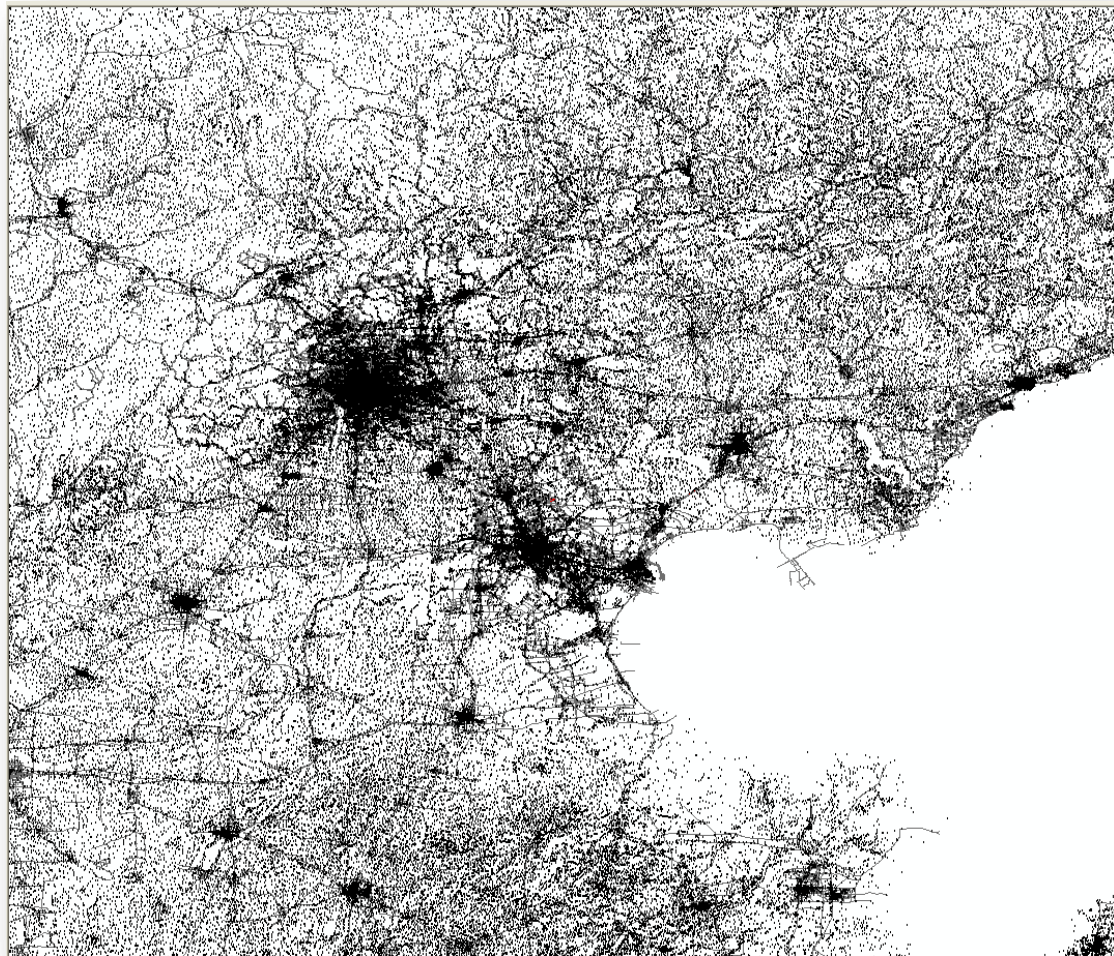
类别	名称	网址
OS: Country	开源荷兰	Oscity.nl
OS: Ctiy	开源城市计划	<a href="https://github.com/">github.com/</a>
OS: City	开源城市列表	<a href="http://futureeverything.org/2012/11/open-data-cities/">futureeverything.org/2012/11/open-data-cities/</a>
OS: 商业网站	数据堂	<a href="http://www.datatang.com">www.datatang.com</a>
OS: 政府	北京市政务数据资源网	<a href="http://www.bjdata.gov.cn">www.bjdata.gov.cn</a>
OS: 学术机构	中国人口地理信息系统	<a href="http://cpgis.fudan.edu.cn/cpgis">cpgis.fudan.edu.cn/cpgis</a>
CS: VGI	Open Street Map	<a href="http://www.opentreetmap.org">www.opentreetmap.org</a>
CS: VGI	Global cities Data	<a href="http://Download.bbbike.org/osm/bbbike">Download.bbbike.org/osm/bbbike</a>
CS: 拍照地址	sightsmap	<a href="http://www.sightsmap.com">www.sightsmap.com</a>
.....	.....	.....

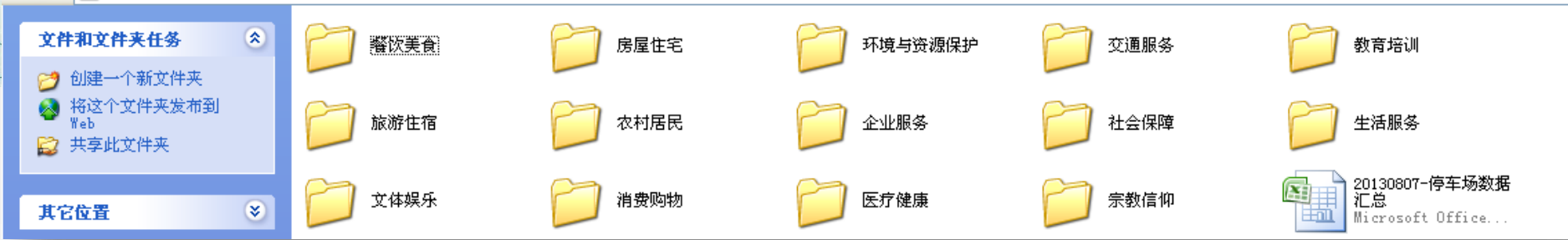
注：OS: Open Source, 开源数据  
CS: Crowdsourcing, 众包数据



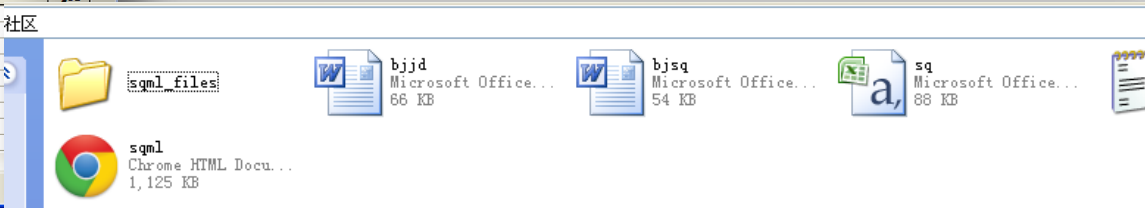
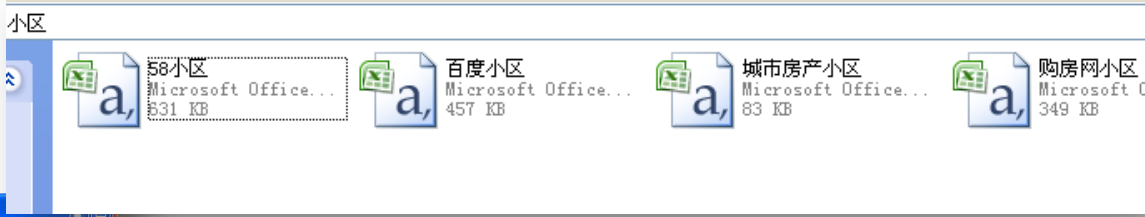
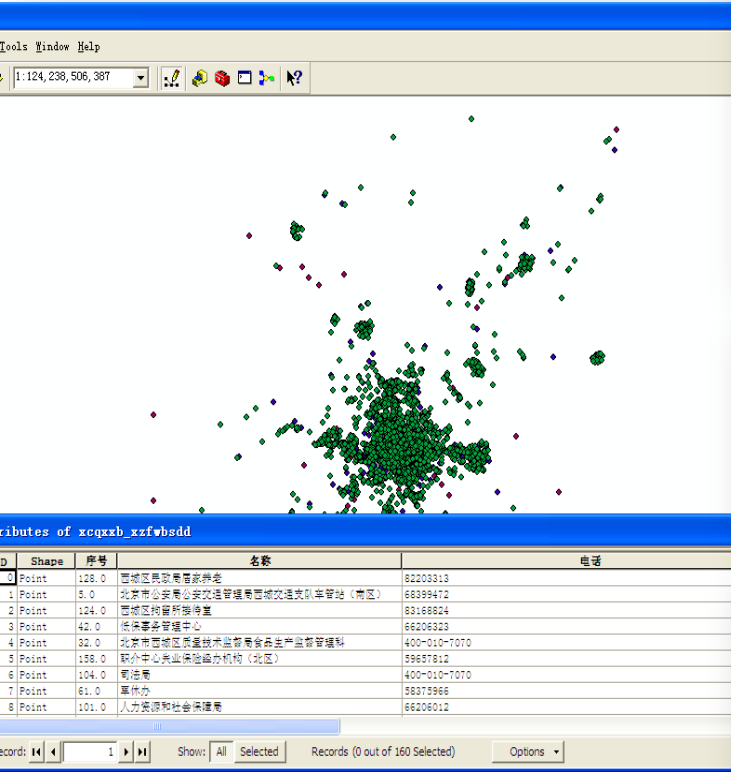
## 课题成果2：数据集

- 各类专题
- 路网
- 边界
- POI
- 历史地图
- 签到
- 人



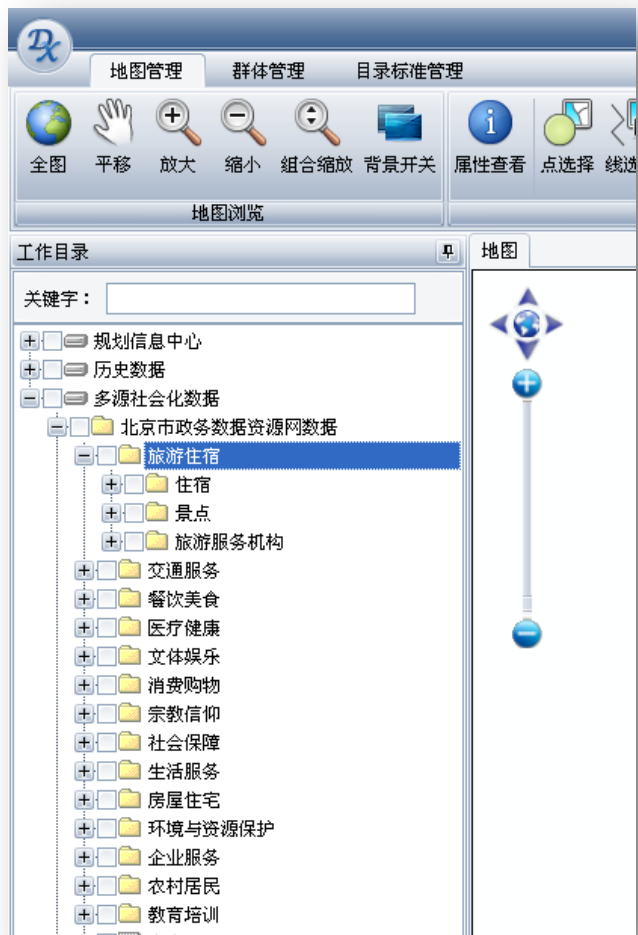


# 专题数据：原始





# 专题数据：入库

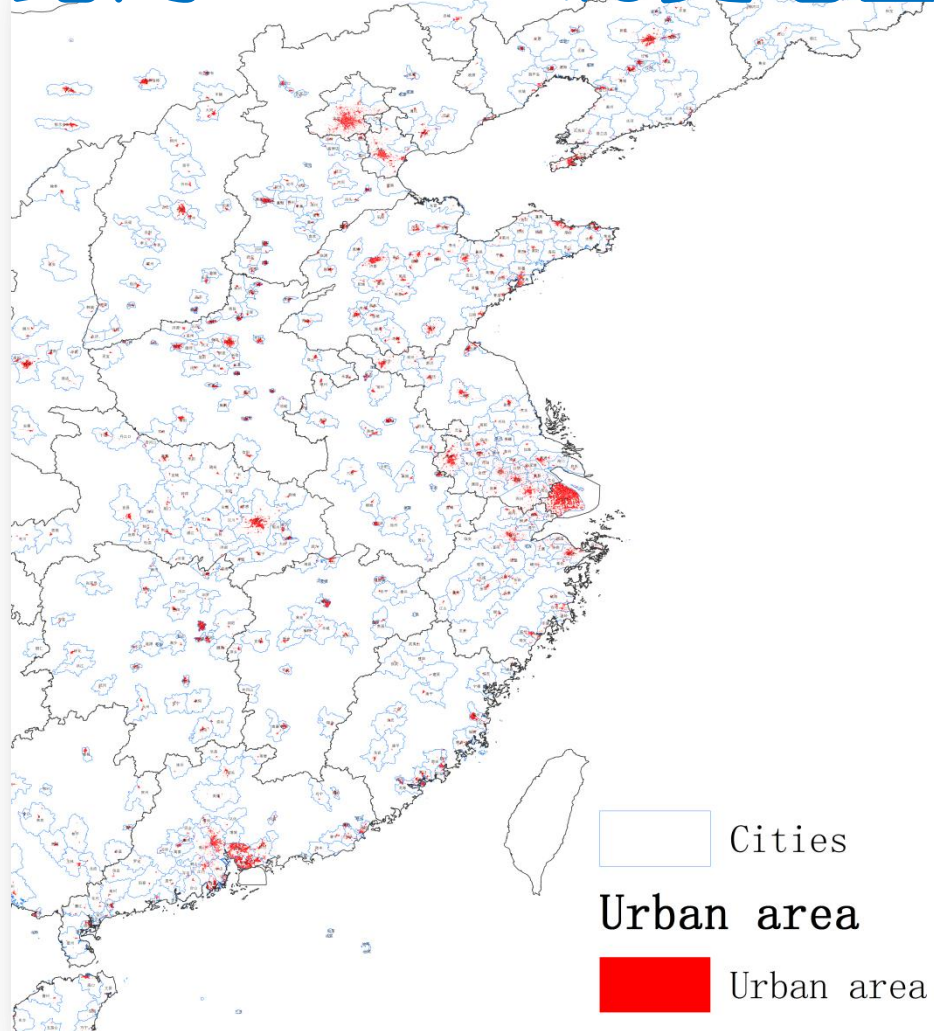


2014/1/2

旅游住宿		
住宿		
	旅游饭店	北京人家.xlsx (已更新) 宾馆旅店.csv (原始数据, 网站未找到相关数据)
	星级酒店	星级饭店.csv (原始数据)
	绿色饭店	金叶级绿色饭店.csv (原始数据) 银叶级绿色饭店.csv (原始数据)
旅游服务机构		
	旅游向导	外资旅行社.csv (原始数据) 旅游咨询服务站.xlsx (已更新) 无出境权旅行社.csv (原始数据) 有出境权旅行社.csv (原始数据)
景点		
	公园广场	森林公园.xlsx (已更新) 注册公园.csv (原始数据) 郊野公园.csv (原始数据)
	博物馆	北京地区博物馆.csv (原始数据)
	工农业旅游	A级景点.csv (原始数据) 农委观光果园.csv (原始数据, 网站未找到相关数据) 民俗村.xlsx 民俗村旅游.csv (原始数据, 网站未找到相关数据) 观光农园.csv (原始数据, 网站未找到相关数据) 观光垂钓.csv (原始数据, 网站未找到相关数据)
	特色街区与建筑	民族工作重点街道、民族乡村.csv (原始数据)



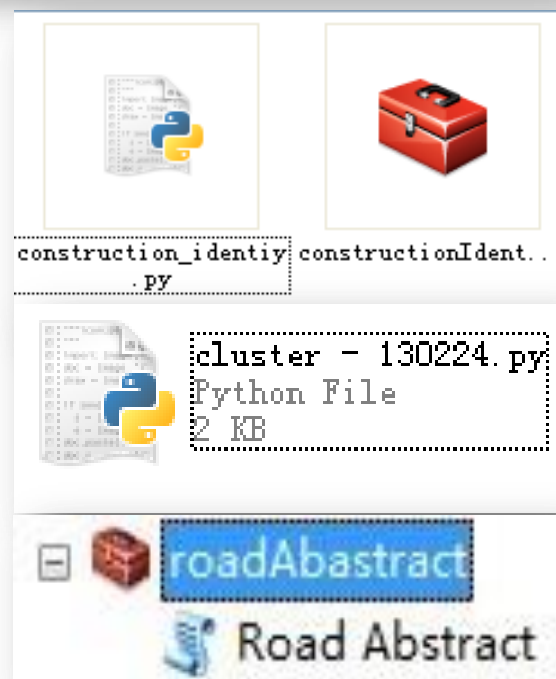
# 路网、POI、历史地图、签到和人的数据



名称	来源
北京市土地出让	<a href="http://www.bjgtj.gov.cn">www.bjgtj.gov.cn</a>
北京路网	<a href="http://www.datatang.com">www.datatang.com</a>
全国POI, 2008年	Beijing City Lab
全国POI, 导航数据, 2009年	Beijing City Lab
297个城市的地块图	Beijing City Lab
全球Open flights datasets	<a href="http://www.openflights.org">www.openflights.org</a>
中国历史地图GIS数据	<a href="http://www.fas.harvard.edu/~chgis/data/chgis/">www.fas.harvard.edu/~chgis/data/chgis/</a>
新浪微博用户数据	Weibo.com
北京POI, 新浪微博签到, 2012	Weibo.com
某酒店数据	Baidu 云
某QQ群数据	Baidu 云

# 课题成果3：应用程序

- 数据抓取程序
  - 微博数据抓取程序
  - 黄页数据爬取程序
  - 百度路况识别提取工具
  - .....
- 数据处理程序
  - 基于OpenAPI的地址匹配程序
  - 坐标纠偏及转换程序
- 数据应用系统
  - 基于AutoCAD的专题图浏览
  - 基于AutoCAD的定位检索
  - 基于AutoCAD的空间库调用





## 课题成果4：论文发表

- 《Automated identification and characterization of parcels (AICP)》-龙瀛等
- 《How Mix Beijing Is ? 》-龙瀛
- 《A Simulation and Application of Population and Employment of Beijing》-何莲娜等
- 《大数据挖掘助力微观尺度下经规对话》-何莲娜等
- 《中规院、清华院、北规院的微博人脉特征分析》-茅明睿
- 《基于微博数据的规划圈识别初探》-茅明睿等



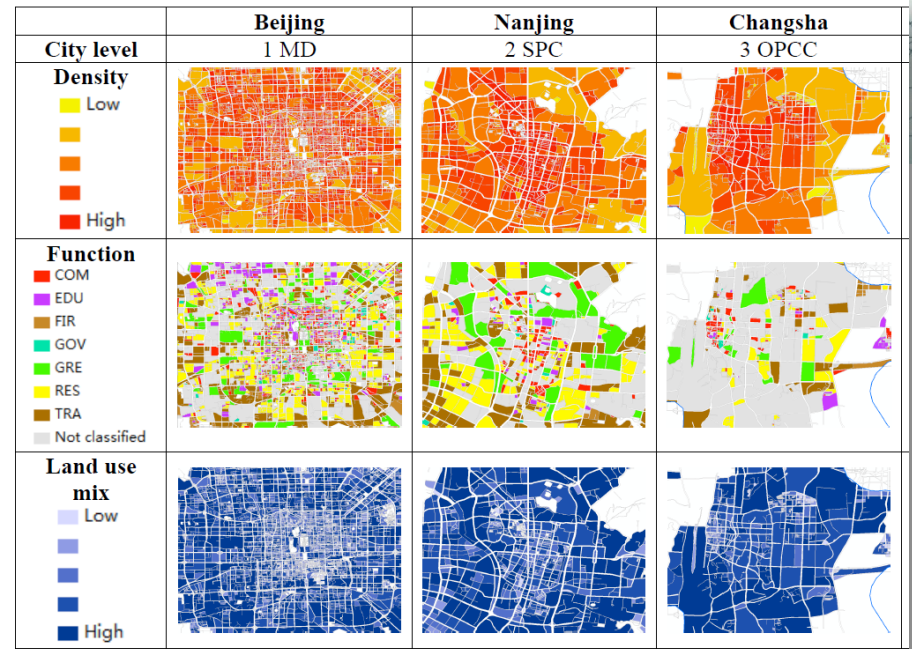
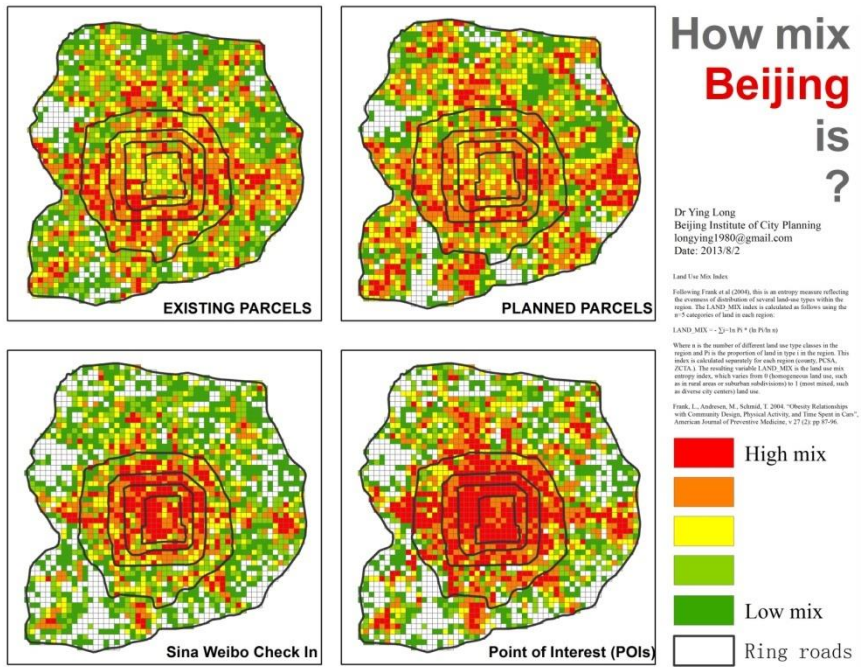
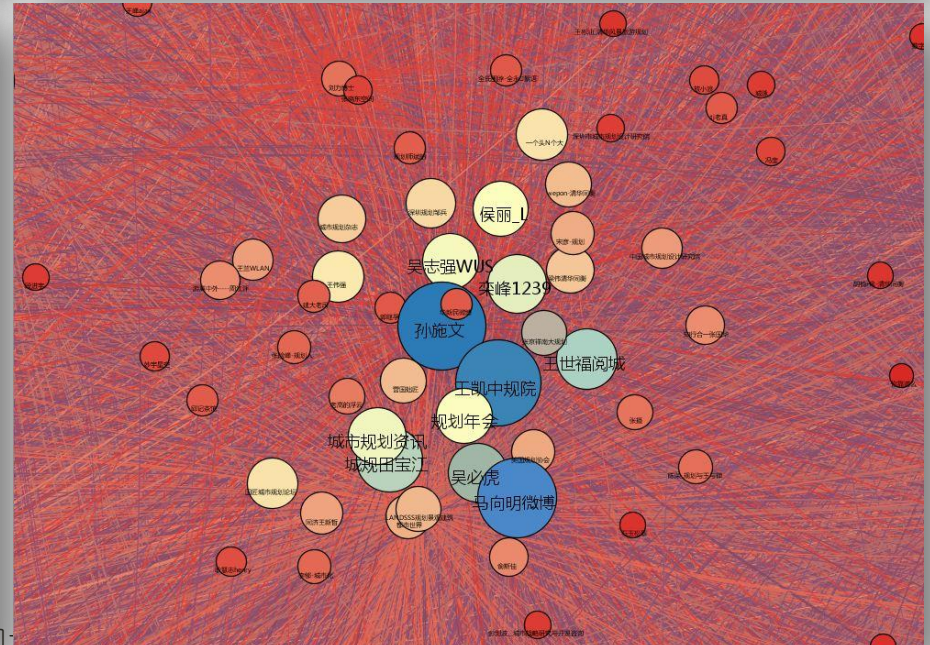
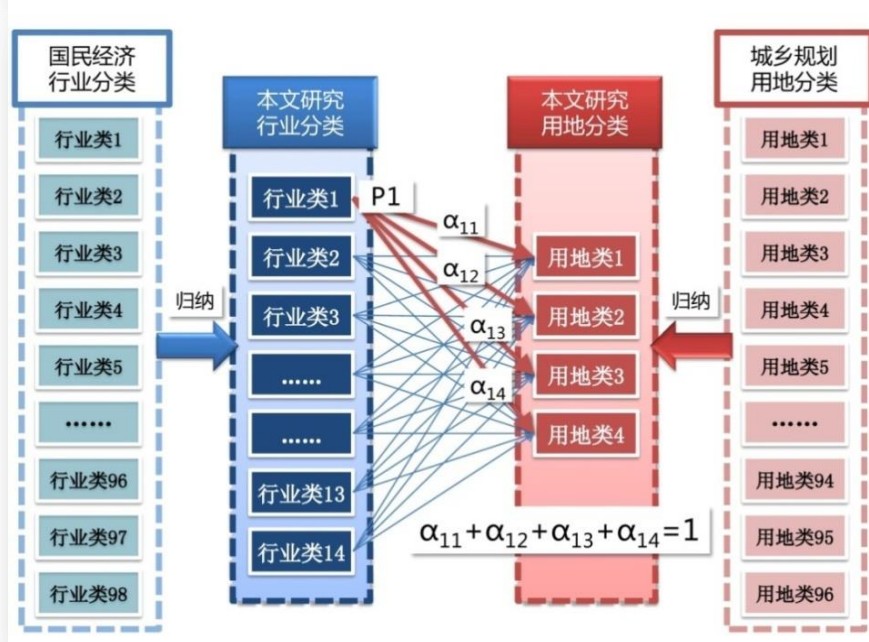


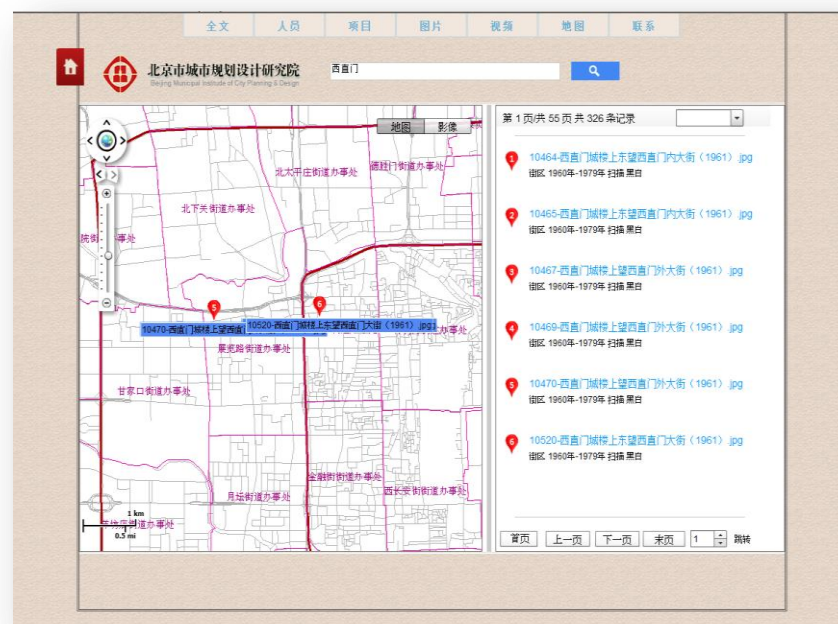
Figure 5 The generated parcels and their attributes in typical cities of China





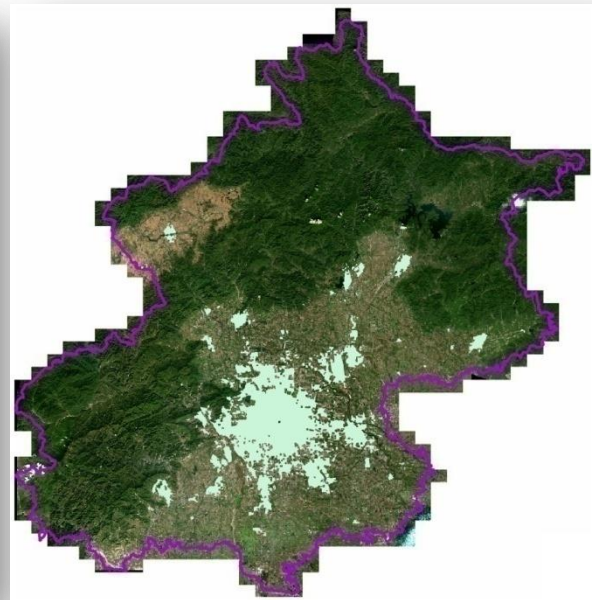
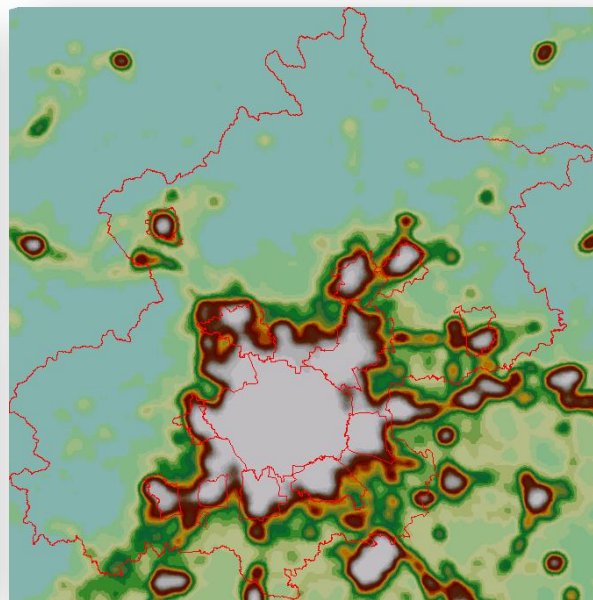
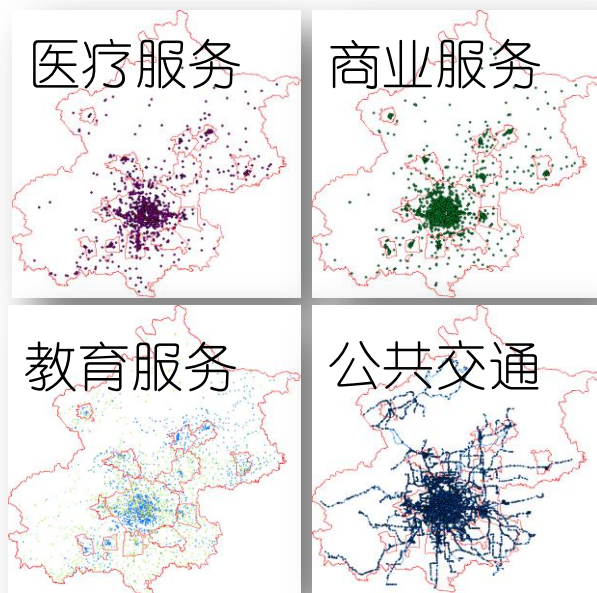
# 成果应用：

- 北京城市建成区界定研究
- 规划设计知识管理与协同工作研究
- 北京市人口、空间、功能及规划实施综合分析



# 成果应用1：北京城市建成区界定研究

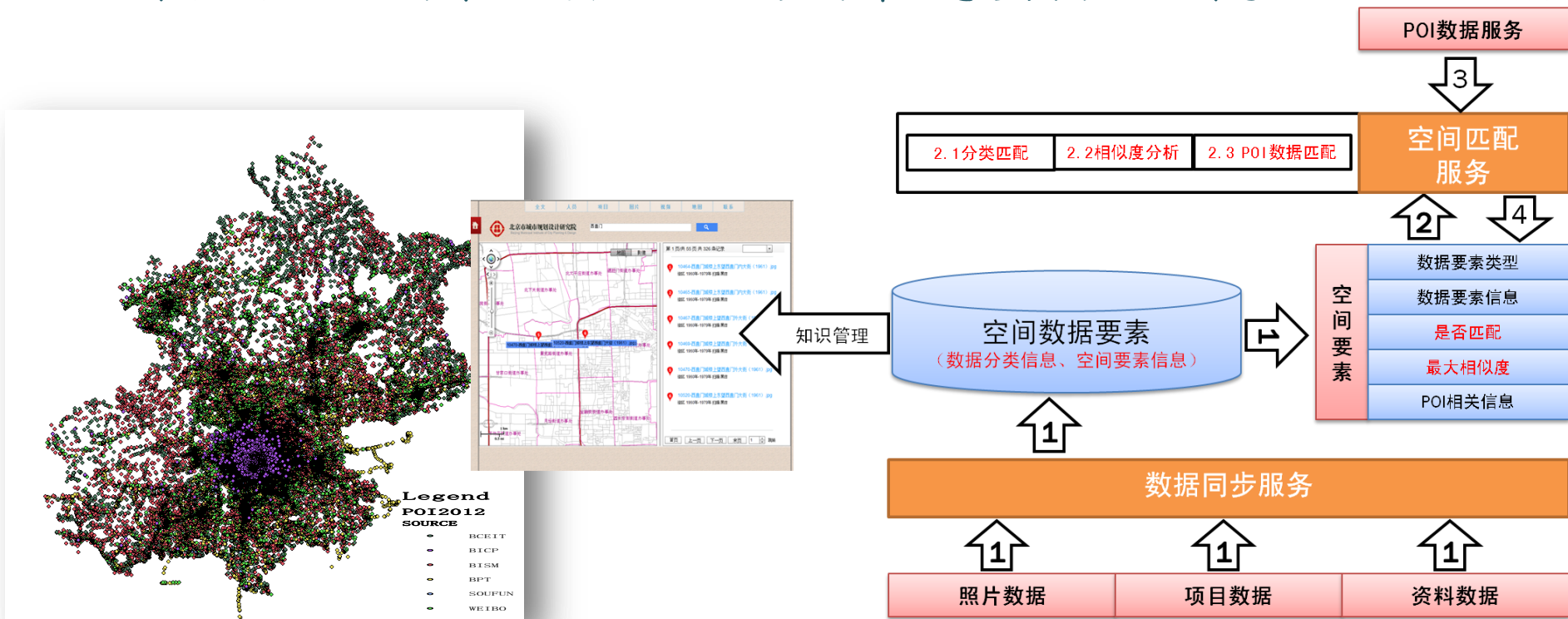
- 利用导航数据作为公共设施的分布依据
- 利用美国国家气象局的DMSP/OLS数据
- 形成了北京城市建成区的界定成果





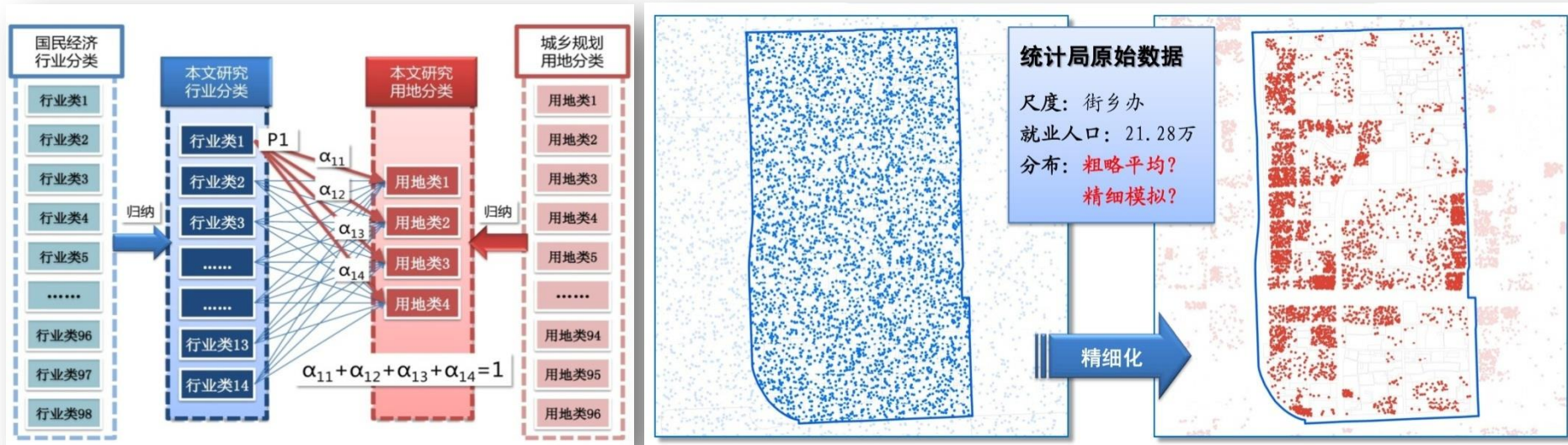
# 成果应用2：规划知识管理与协同工作研究

- 利用了新浪微博的POI数据建立了北京地址库
- 利用地址库对照片、文本等数据进行地址匹配



# 成果应用3：人口、空间、功能及规划实施综合分析

- 基于社会化数据智能获取技术抓取的2010年北京市企业黄页数据，含52万多个就业单位的空间位置、行业分类（依据国民经济分类标准）、就业容量等信息。
- 利用该数据实现了就业人口在微观空间尺度上的分布模拟。





## 结论与展望：

- 社会化数据是城市规划工作的**可靠**数据来源
- 社会化数据能给规划带来新的**思路**，能拓展我们对城市的**认知**和**研究方法**
- **大数据**的应用和相关研究方法将是未来的城市**研究热点**
- **开源**和**众包**已经并且会进一步影响城市的公共治理，不仅会成为我们**获取数据**的**渠道**更会成为**研究城市、治理城市**的重要途径



# 谢谢

