China Data Institute

Center for Geographic Analysis
Harvard University

社会地理计算联合研究中心
Geocomputation Center for Social Science

RMDS

The Guideline
for
Application for Global Research on Novel Coronavirus
“新冠病毒数据资源与全球研究”项目申报指南

http://chinadatalab.net

“The Resources for Novel Coronavirus and Global Research”
“新冠病毒数据资源与全球研究”项目申请

Research Topics (not limited to) 研究议题（建议但不限于）：

- 基于大数据的疫情监测、预警与风险评估
  Outbreak monitoring, early warning and risk assessment based on big data
- 新冠病毒疫情传播扩散时空变化模式研究
  Spatio-temporal Patterns of the Transmission and Spread of Novel Coronavirus
- 新冠病毒疫情对中国与全球贸易与投资影响研究
  Impact of Novel Coronavirus on Trade and Investment in China and the World
- 新冠病毒疫情对地区与全球产业供应链影响研究
  Impact of Novel Coronavirus on Regional and Global Industry Supply Chain
- 新冠病毒疫情中舆情变化与公众行为规律研究
  Public Opinions and Behavior in Response to Novel Coronavirus
- 新冠病毒疫情的时空动态与季节气候环境关系研究
  Spatio-temporal dynamics of the Novel Coronavirus and seasonal climate and environment

Deadline of Application 申请提交截止日期：3/15/2020

Contact： office@chinadatacenter.net

Note 说明：国际合作项目优先考虑 (proposals with international collaboration will be given higher priority)

Participants 支持单位

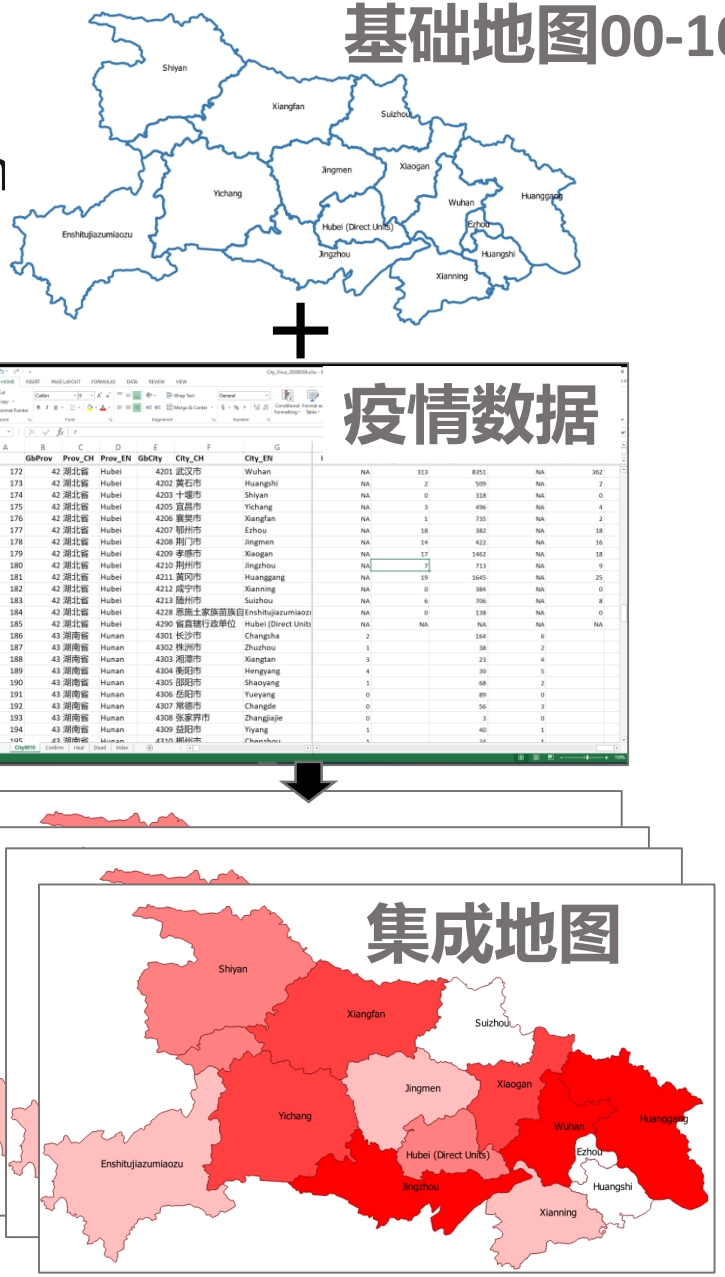
- Geocomputation Center for Social Sciences, Wuhan University
  武汉大学社会地理计算中心
- Center for Geographic Analysis, Harvard University
  哈佛大学地理分析中心
- RMDS Lab 研究方法与数据科学实验室
- China Data Institute 中国数据研究所
- MicroSoft China 微软中国
- Vesystem Inc.和信创天科技有限公司
- Knowledge Sharing Inc.百智享科技有限公司
- All China Marketing Research Inc. 华通人市场研究有限公司

Objectives 项目目标

- 培育先期研究 To develop some pilot studies on COVID-19 for future research
- 专业数据分析 To cultivate professional data analysis
- 新方法与技术探索 To explore new methodology and technology for COVID-19 data analysis
- 鼓励可复制、可重复、可扩展研究 To encourage replicable, reproducible and expandable research
- 鼓励知识与信息共享 To promote knowledge and information sharing
- 促进国际合作 To promote international research collaborations
- 促进教研数据应用 To promote data applications in research and education
- 团队能力建设To build the capacity for future collaborations

Data on COVID-19 疫情数据选择

- **Base data 基础数据**
  - Base maps (province, prefecture and county)
  - Census data (population census and economic cen
  - Statistics (province, prefecture and county)
  - Others
- **Virus data病毒案例**
  - Virus reports from gov't (country, province and city
- **Migration data迁徙人口**
  - Daily floating population data (Baidu)
  - Migration data from census (NBS)
- **Health facility data 健康设施**
  - POI data (AutoNavi)
  - Economic Census (NBS)
- **Social media data社会媒体**
  - Weibo
  - Twitter
- **Papers论文**
- **Reports报告**
- **Policies and regulations政策与法规**



Base Data Sources 基础数据选择

- **Government Statistics**
  - Provincial Statistics (1949 -)
  - City Statistics (1996 -)
  - County Statistics (1997 -)
- **Population Census**
  - Census 1953
  - Census 1964
  - Census 1982
  - Census 1990
  - Census 2000/2010 (province, city, county, township, GRID)
- **Economic Census**
  - Industrial Census 1995 (province, city, county, ZIP)
  - Basic Unit Census 2001 (province, city, county, ZIP)
  - Economic Census 2004/2008 (province, city, county, ZIP)
- **Establishments** (more than 7 millions companies and organizations)
- **Geography and Environment**
  - Land Use data
  - Night-Time data

- **Base Maps**
  - 2000
  - 2010
  - 2000-2010

Tool Selection 工具选择

- ArcGIS
- GeoDa
- GAUSS
- Jupyter
- Knime
- R AnalyticFlow
- Alteryx
- ….

Project Report 项目报告

- 标题 Title
- 关键字 Key words
- 摘要 Abstract
- 研究背景 Background (hypothesis, debates from literature)
- 数据 Data
- 方法 Methodology
- 分析 Results of the Analysis
- 总结和讨论 Conclusions and Discussions
- 致谢 Acknowledgements
- 参考文献 References
- 数据分析流程图 Diagram of flowchart for data analysis
- 数据分析工作流 Workflows for data analysis in the report

Evaluation Criteria for Applications 项目评估准则

☐ Qualification of Applicant 申请人资质:

- Research background 研究背景
- Learning ability 学习能力
- Active participation 积极参与
- Desire to collaborate with others 合作愿望

☐ Research Plan 项目计划:

- Research plans and feasibility 研究计划可行性
- Replicable, reproducible and expandable research 项目的可复制、可重复与可扩展性
- Research results to be shared and delivered 可共享、可发布研究结果

Deployment of Research Data, Reports, Workflows and Presentations at dataverse.harvard.edu 研究成果发布

The Dataverse® Project

Open source research data repository software

Researchers

Enjoy full control over your data. Receive *web visibility, academic credit,* and *increased citation counts.* A personal dataverse is easy to set up, allows you to display your data on your personal website, can be branded uniquely as your research program, makes your data more discoverable to the research community, and satisfies data management plans. Want to set up your personal dataverse?

Journals

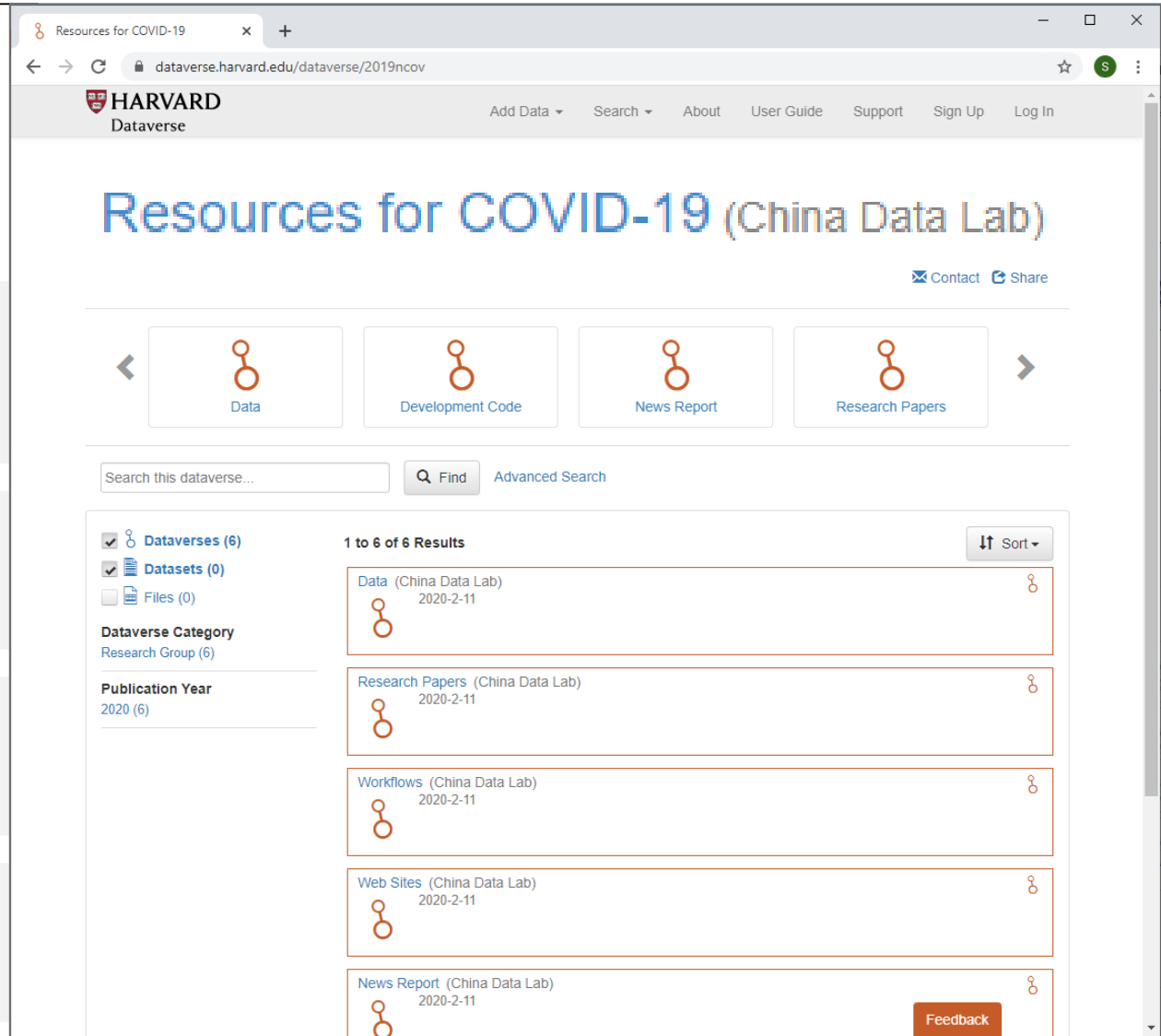
Seamlessly manage the submission, review, and publication of data associated with published articles. Establish an *unbreakable link* between *articles in your journal* and *associated data.* Participate in the open data movement by using Dataverse as part of your journal data policy or list of repository recommendations. Want to find out more about journal dataverses?

Institutions

Establish a research data management solution for your community. Federate with a growing list of Dataverse repositories worldwide for increased discoverability of your community's data. Participate in the drive to set norms for sharing, preserving, citing, exploring, and analyzing research data. Want to install a Dataverse repository?

Developers

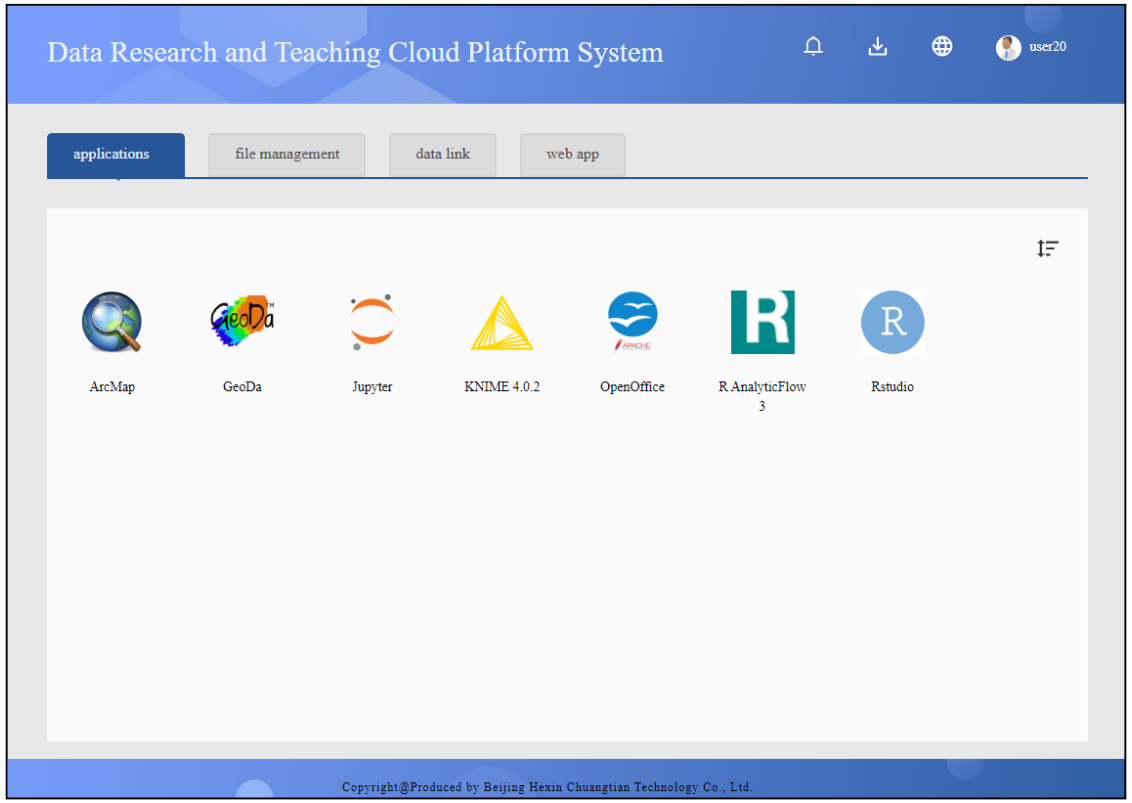
Participate in a vibrant and growing community that is helping to drive the norms for sharing, preserving, citing, exploring, and analyzing research data. Contribute code extensions, documentation, testing, and/or standards. *Integrate research analysis, visualization* and *exploration tools,* or other research and data archival systems with Dataverse. Want to contribute?



Deployment of Executable Workflows on CDL Cloud

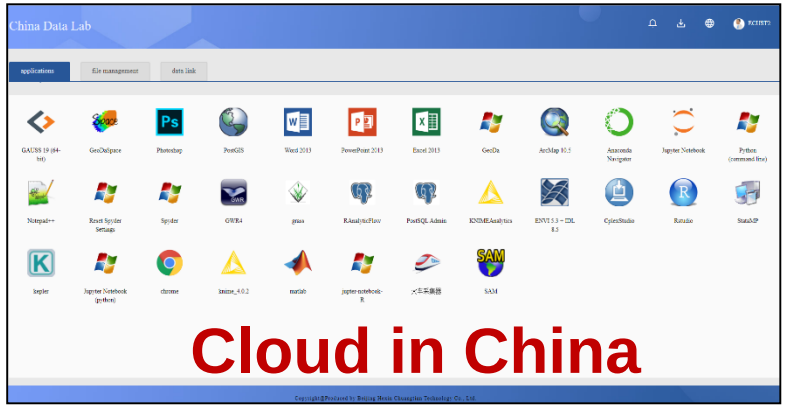
可执行工作流数据案例发布

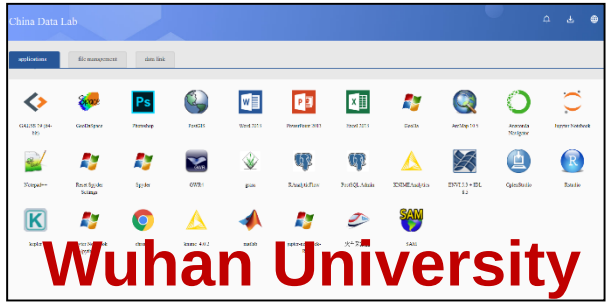
CDL in Harvard

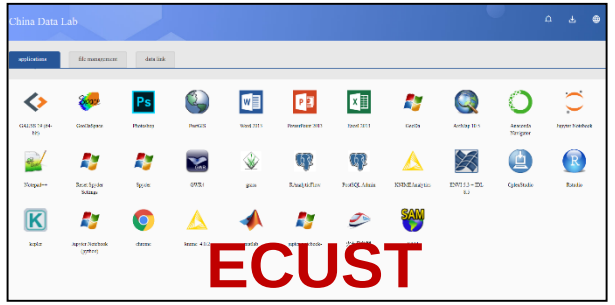


http://harvard.chinadatalab.org

CDL in China







http://chinadatalab.org

Executable Workflow Data Analysis from DataVerse

Tool source: https://www.knime.com/

Open for Innovation KNIME

Workflows for download on dataverse.harvard.edu

HARVARD Dataverse

Add Data ▾ Search ▾ About User Guide Support China Data Lab 2 ▾

Harvard Dataverse > China Data Lab Dataverse > Resources for COVID-19 > Workflows > **COVID19 Stats Anlaysis**

Contact Share Publish Edit ▾

COVID19 Stats Anlaysis

Draft Unpublished

Tao Hu, 2020, "COVID19 Stats Anlaysis", https://doi.org/10.7910/DVN/FWOPW2, Harvard Dataverse, DRAFT VERSION

Cite Dataset ▾ Learn about Data Citation Standards.

Dataset Metrics

0 Downloads

Description	This dataset saves workflows related to the COVID-19 statistics anlaysis. (2020-02-20)
Subject	Medicine, Health and Life Sciences; Computer and Information Science; Social Sciences
Keyword	workflow, statistics, map

Files Metadata Terms Versions

Search this dataset... Find + Upload Files

Filter by File Type: All ▾ Access: All ▾ Sort ▾

1 to 3 of 3 Files Edit Files ▾ Download

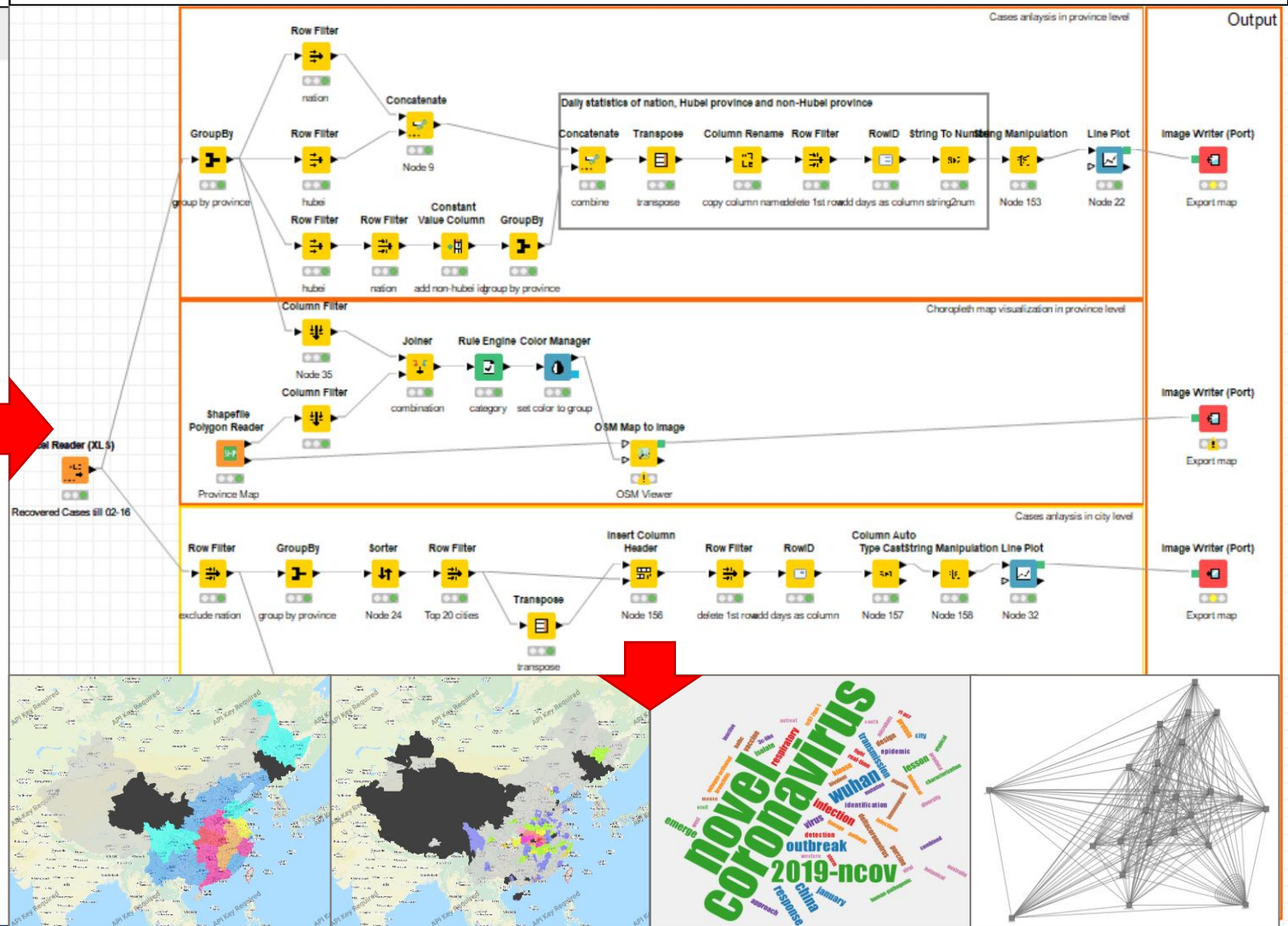
covid19_stats_confirmed.knwf
Unknown - 98.8 KB - 2020-2-20 - 0 Downloads
MD5: 301780709c59bec212ec9de6cb7522cd
Copy and paste the following link to the executable workflow on your
"veappc://88D2AA413AED9040FAE30B4D4FBE0916B56E161735A Download

covid19_stats_death.knwf
Unknown - 99.7 KB - 2020-2-20 - 0 Downloads
MD5: 4b55d67e49361f8952ae1d4e0c2f6885
This workflow demonstrates the analysis of death cases in province
veappc://88D2AA413AED9040FAE30B4D4FBE0916B56E161735A7 Download

covid19_stats_recovered.knwf
Unknown - 98.2 KB - 2020-2-20 - 0 Downloads
MD5: cfb5e9f1ff29c9aa0a01f7f511744098
This workflow demonstrates the anlaysis of recovered cases in provi
veappc://88D2AA413AED9040FAE30B4D4FBE0916B56E161735A7 Download

Feedback

Executable workflows on harvard.chinadatalab.org



Sample Work: Religious Diversity and Regional Development
研究案例：宗教多样性与地区发展

Get Access Export

China Economic Review
Volume 46, December 2017, Pages 1-9

ELSEVIER

Religious diversity and regional development in China

Zheng Ying [a, b], Shibao Liu [c], Shuming Bao [d, e], Jianbo Zhou [a]

Show more

https://doi.org/10.1016/j.chieco.2017.08.003 Get rights and content

Highlights

- Identify possible impacts of religious diversity on regional development at province level under the same political regime
- Apply fragmentation and polarization indices as measures of religious diversity
- Solve the endogeneity issue with the fixed-effect two-stage regression

Abstract

This paper investigates possible impacts of religious diversity on regional development in China. We developed the religious fragmentation and polarization indices as measurements for the religious diversity by using the religious site data from the 2004 economic Census data in China. The results from the panel data regressions within national coverage suggest that the religious diversity has positive and significant impacts on regional development in general. The results from the panel data regressions within regional coverage also suggest that religious fragmentation has a positive and significant association with the economic development in the eastern region of China while religious polarization has a positive and significant association with the economic development in the central and western regions of China.

Aim: Test the impacts of religious diversity on regional development in China

Findings from previous studies: Religious diversity may have positive or negative impacts on economic development

Background of this study: Test the impacts of religious diversity on regional development under the same political regime at sub-national level (province) in a peaceful environment

Data 数据

Name	Format	Description
Data_religion.xlsx	Excel	The number of religion sites of different types at different provinces from 1987 to 2004
Data_statistics.xlsx	Excel	variables included in the regression analysis, including real income per capita, average education attainment, investment in physical capital, policy variable, Growth rate of labor & tech, dep.
province_boundary.shp	Shape	the boundaries of provinces in mainland China

Methodology 方法

Measurements of religious diversity: fragmentation and polarization indices

Fragmentation: $FRAG_i = 1 - \sum_{j=1}^{J} \left(\frac{n_{ij}}{N_i}\right)^2$

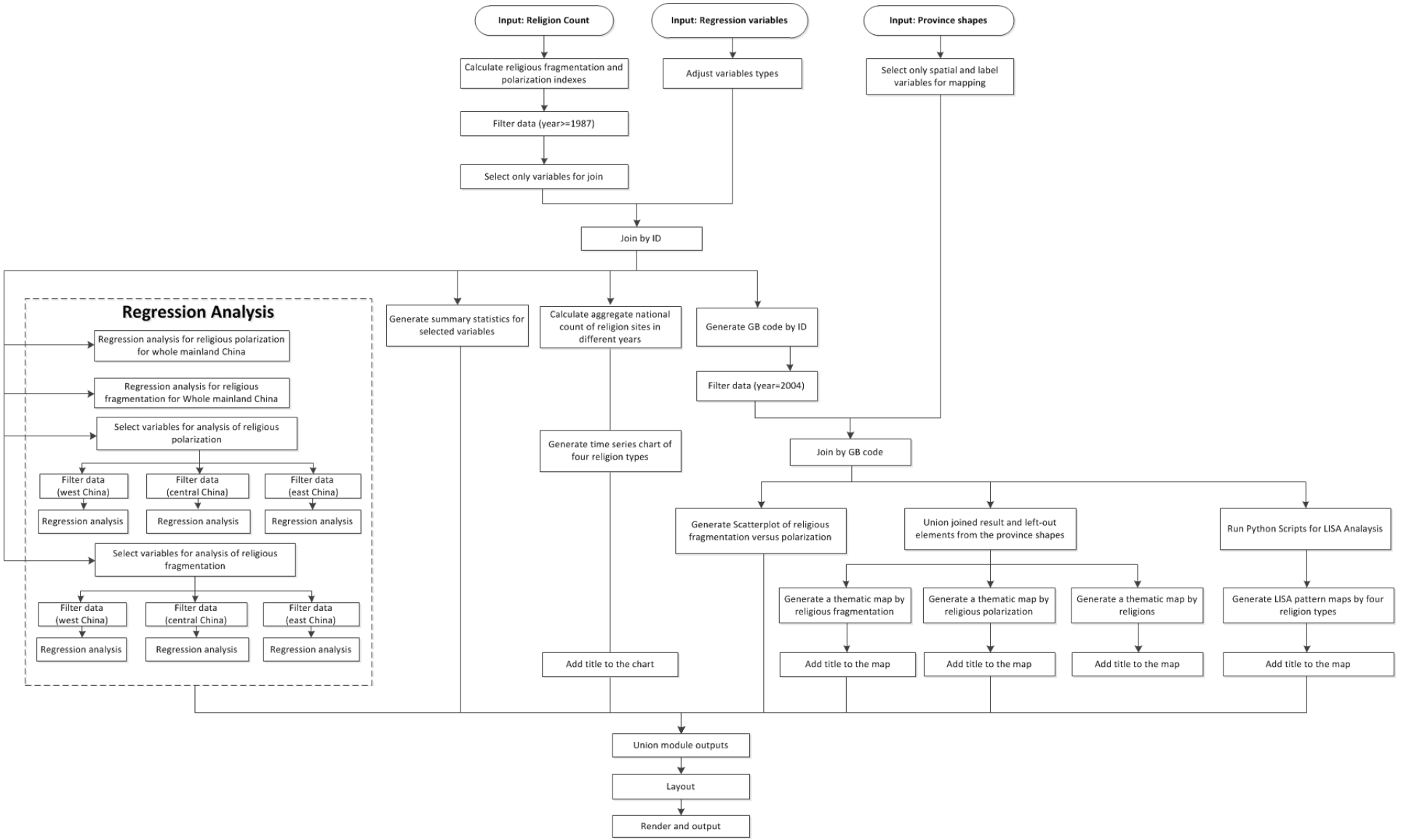
Polarization: $POL_i = 1 - \sum_{j=1}^{J} \left(\frac{0.5 - \pi_{ij}}{0.5}\right)^2 \pi_{ij}$

Regression: $ln\frac{Y(t)}{L(t)} = \beta_0 + \beta_1 lns_k + \beta_2 lns_h + \beta_3(n + g + \delta) + u$

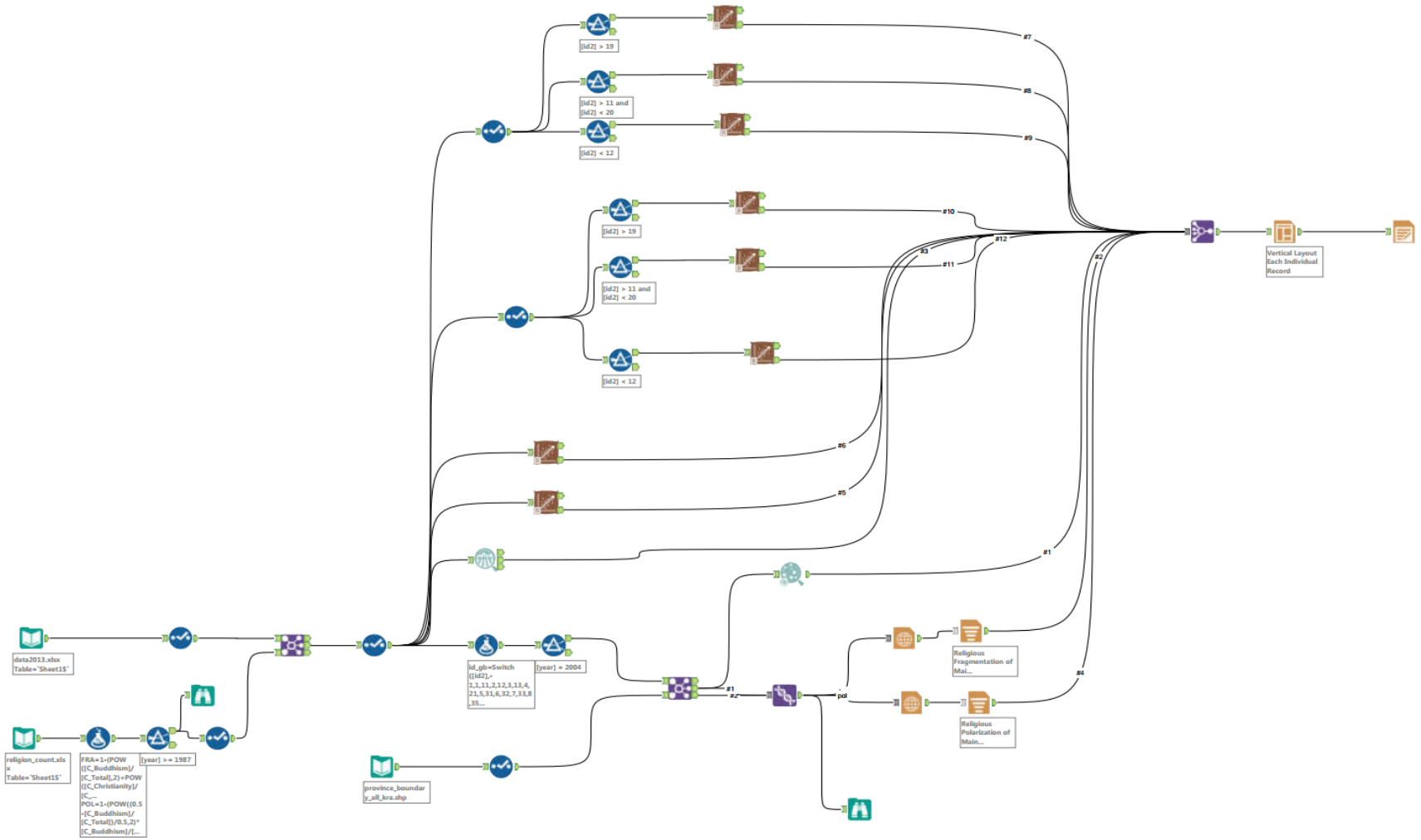
Moran's I test for spatial autocorrelation:

$$I(d) = \frac{\sum_{i}^{n}\sum_{j\neq i}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{S^2 \sum_{i}^{n}\sum_{j\neq i}^{n} w_{ij}}$$

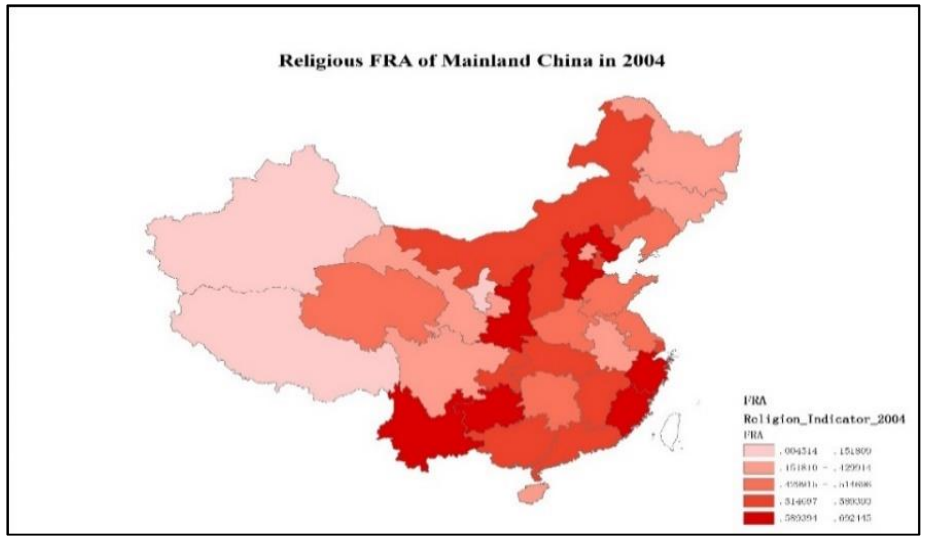
The Flowchart for Data Analysis 概念性数据分析流程图

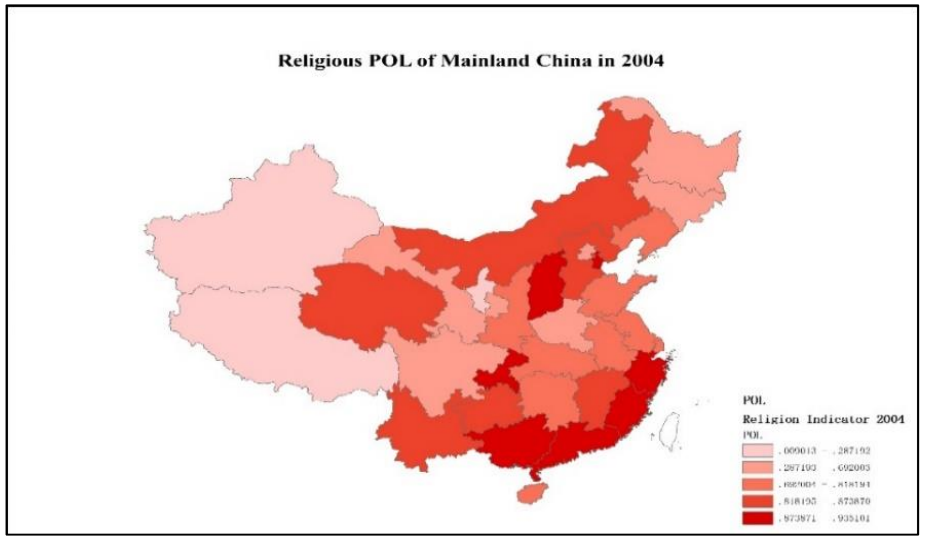


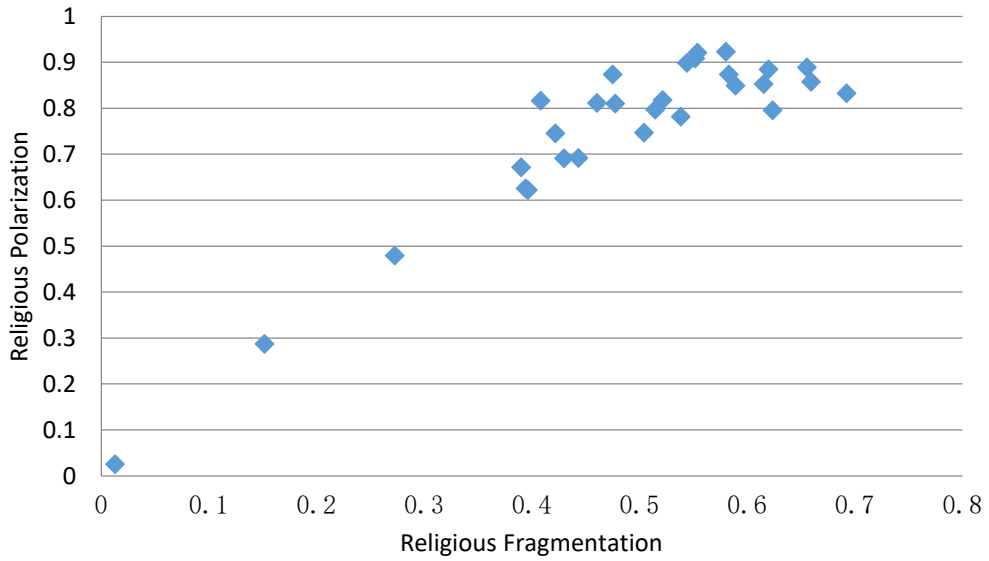
The Workflow for Data Analysis 技术性数据分析工作流



Results from Analysis 数据分析结果







	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Whole	East	Central	West	Whole	East	Central	West
frag	0.482***	0.550***	0.0433	-0.243				
	(4.31)	(3.89)	(0.44)	(-0.88)				
(n+g+δ)	-0.00281	-0.00185	-0.0169*	-0.0484***	-0.00286	-0.00271	-0.0178*	-0.0458***
	(-0.54)	(-0.36)	(-1.71)	(-3.58)	(-0.54)	(-0.51)	(-1.81)	(-3.37)
lnedu	-0.415***	0.221	0.213	-0.565***	-0.402***	0.202	0.203	-0.580***
	(-2.91)	(0.89)	(1.08)	(-3.63)	(-2.77)	(0.77)	(1.03)	(-3.77)
lncap	0.0198	0.227***	0.0638**	0.0990**	0.0275	0.235***	0.0734**	0.0919**
	(0.67)	(5.30)	(1.99)	(2.22)	(0.91)	(5.21)	(2.27)	(2.08)
lnpub	-0.119***	-0.281***	0.0840	-0.128**	-0.131***	-0.308***	0.0880	-0.108*
	(-2.84)	(-5.49)	(1.53)	(-2.00)	(-3.08)	(-5.81)	(1.60)	(-1.69)
pol					0.141	0.165	-0.0408	0.417*
					(1.52)	(1.29)	(-0.50)	(1.69)
_cons	6.967***	6.089***	6.469***	7.690***	7.067***	6.233***	6.570***	7.333***
	(24.91)	(13.18)	(15.20)	(23.02)	(24.19)	(12.04)	(15.32)	(20.27)
N	504	180	144	180	504	180	144	180
R^2	0.972	0.987	0.994	0.982	0.971	0.986	0.994	0.982
A-R^2	0.969	0.984	0.992	0.978	0.968	0.983	0.992	0.978
F Sta.	727.2	503.6	794.4	363.0	701.5	461.6	794.8	368.3

Conclusions and Discussions 结论和讨论

- The results from the panel data regressions within national coverage suggest that the religious diversity has positive and significant impacts on regional development in general.
- The results from the panel data regressions within regional coverage suggest that religious fragmentation has a positive and significant association with the economic development in the eastern region of China while religious polarization has a positive and significant association with the economic development in the central and western regions of China.

References 参考文献

- Ying, Zheng, Shibao Liu, Shuming Bao and Jianbo Zhou, 2017. Religious Diversity and Regional Development. China. China Economic Review, Vol. 46, P1-9. https://doi.org/10.1016/j.chieco.2017.08.003.
- Alesina, A., & Zhuravskaya, E. (2011). Segregation and the quality of government in a cross section of countries. American Economic Review, 101 (5), 1872–1911.
- Alesina, A., Harnoss, J., & Rapoport, H. (2016). Birthplace diversity and economic prosperity. Journal of Economic Growth, 21 (2), 1-38.
- Knack, S., & Barro, R. (1998). The Determinants of Economic Growth. Determinants of economic growth: MIT.
- Kodila-Tedika, O. (2012). Determinants of Peace : A Cross-Country Analysis. (Vol.2, pp.180-200). University Library of Munich, Germany.
- Lu H, Page S E. (1998).Diversity and Optimality. Research in Economics.
- Mankiw, N. G., Weil, D. N., & Romer, D. (1992). A contribution to economic growth. , 107(2), 407-437.
- Montalvo, J.G. and M. Reynal-Querol, 2003. Religious polarization and economic development. Economics Letters, 80 (2), 201-210.
- Montalvo, J.G., Reynal-Querol, M. (2000). “The effect of ethnic and religious conflict on growth”, IVIE WP-EC 2000-04. An updated version can be found in http: / /www.wcfio.harvard.edu/programs/ prpes.
- Okten, C., & Osili, U. O. (2004). Contributions in heterogeneous communities: evidence from Indonesia. Journal of Population Economics,17(4), 603-626.
- Ottaviano, Gianmarco, and Giovanni Peri. (2003）.“The Economic Value of Cultural Diversity: Evidence from US Cities.” University of California, Davis. Unpublished.

The List of Documents for Report 结果文件列表

Documents	Description	File
Workflows	Workflow files	*.yxmd
Flowcharts	The flowchart of workflows	*.doc
Case study report	Word or PDF file	Religion_study.docx
Case study presentation	Presentation file	Religion_case.pptx
Data	Religious data table in Excel	Data_religion.xlsx
	Statistical table in Excel	Data_statistics.xslx
	Province map in Shape file	Province_boundary.shp
Output files	Output files	Output_religion.pdf

Output Files 结果输出文件

Description	Format
Map of Religious Fragmentation by Province in 2004	Map
Map of Religious Polarization by Province in 2004	Map
Scatterplot of Religious fragmentation versus Religious polarization in 2004	Chart
Summary statistics of regression variables	Text
Regression analysis (Residual table, coefficient table, ANOVA Analysis, Basic Diagnostic Plots)	Text

The Cloud Platform for Global Research on Novel Coronavirus

Qs & As 问题与回答

Q: Where we can find the data for the project?

A: Most aggregated data can be found from dataVerse at Harvard for public access. Some microdata and basic data will be located on the cloud with restricted access. All participants are also encouraged to collect the data from other sources and share with participants of other research groups. For those data on the wish list from other sources, we can help reach out to other data sources for possible support.

Q: Where will be the data depository for the final research data and reports?

A: We'll use dataVerse at Harvard primarily for data depository, which is free and linked to the online map viewer (WorldMap) and the cloud platform for COVID-19 study.

Q: What are the criteria for evaluating our applications?

All applications will be reviewed by the review committee with some leading scholars and professionals from Harvard, Wuhan Univ., Peking Univ., RMDS and China Data Institute. For those pilot projects, we hope that the project will help those participants to shape and build their research ideas for other grant applications, and also learn new methodology and technology for future development. Thus, we'll pay more attention to the applicants' learning ability, active participation, and desire to collaborate with others.

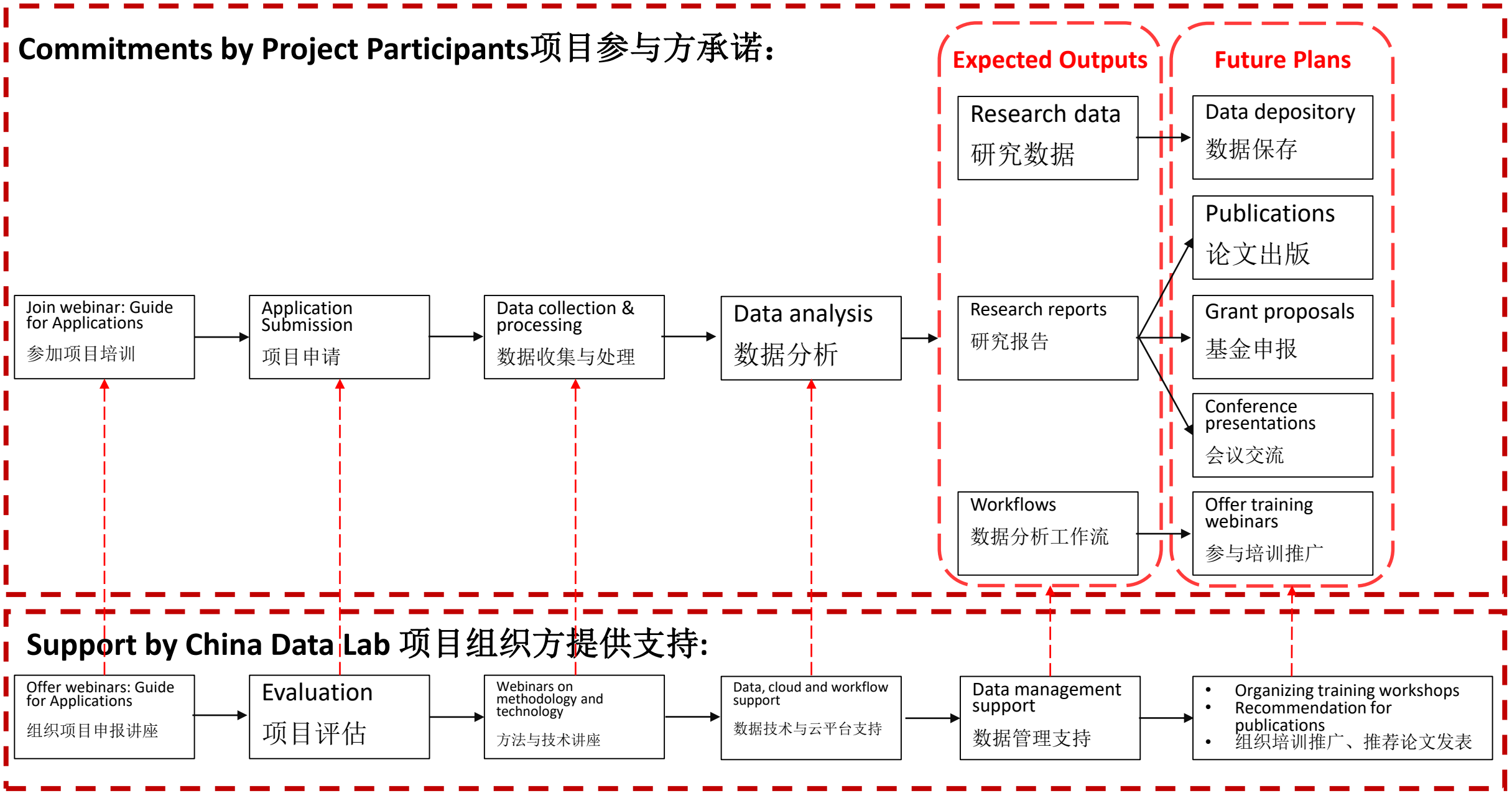
Q: What support can those participants receive?

A: Each participants will receive a package of various support during the whole process, including webinars for project application, training webinars for data, methodology and technology, research data we can offer, cloud platform account, depository support for research data and reports, opportunity for offering training workshops, and recommendations for publications.

Q: Who will own the copyright of those research results or publications from the project?

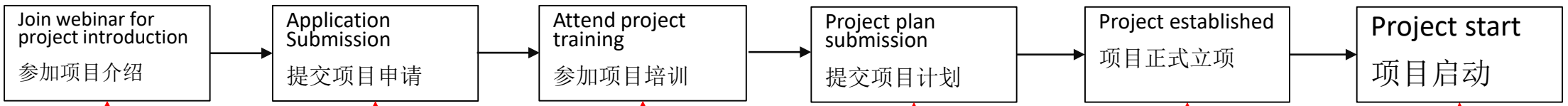
A: The author(s) will own the copyright. The organizer would like to have distribution right for those research results from the project, including derived research data, research reports, workflows and presentations. The author(s) may acknowledge the support of the China Data Lab in their publications.

Project Flowchart: Commitments and Support项目流程

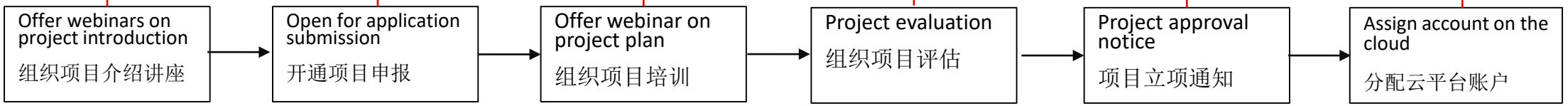


Project Flowchart 项目申报流程

Project Participants 项目申请方：



China Data Lab 项目组织方:



Training Workshops on Workflows with Knime工作流培训讲座

1. **Training on building workflows using Knime** -2019/11/21 (1 hr 28 mins)
https://chinadatainstitute.my.webex.com/recordingservice/sites/chinadatainstitute.my/recording/playback/e9cedc5354194819891e3ee6f9fd1aa3
Password: 9xJUTmSe

2. **Spatial Statistical Analysis with Knime**, Lingbo Liu-2019/11/27 (1 hr 17 mins)
https://chinadatainstitute.my.webex.com/recordingservice/sites/chinadatainstitute.my/recording/playback/6691ea0040f8456d81acffb1930e3186
Password: Pi8qgAMn

3. **Industrial co-agglomeration analysis with Knime**, Lu Chen-2019/11/29 (1 hr 35 mins)
https://chinadatainstitute.my.webex.com/recordingservice/sites/chinadatainstitute.my/recording/playback/5b6be24540694592861b73788f48ebe4
Password: Tb3sarph

4. **Online training on workflow tool KNIME**-2020/02/07 (1 hr 5 mins)
https://chinadatainstitute.my.webex.com/recordingservice/sites/chinadatainstitute.my/recording/playback/f45fec687cb644198b0c7686ee85153c
This recording does not require a password.

Project Introduction and Application 项目介绍与项目申请

Project Introductions in PPT:

- **An Introduction to the Data Resources for Covid-19 Studies**
https://doi.org/10.7910/DVN/OTYQUY

- **Global Research on COVID-19**
https://doi.org/10.7910/DVN/9ZIDYR

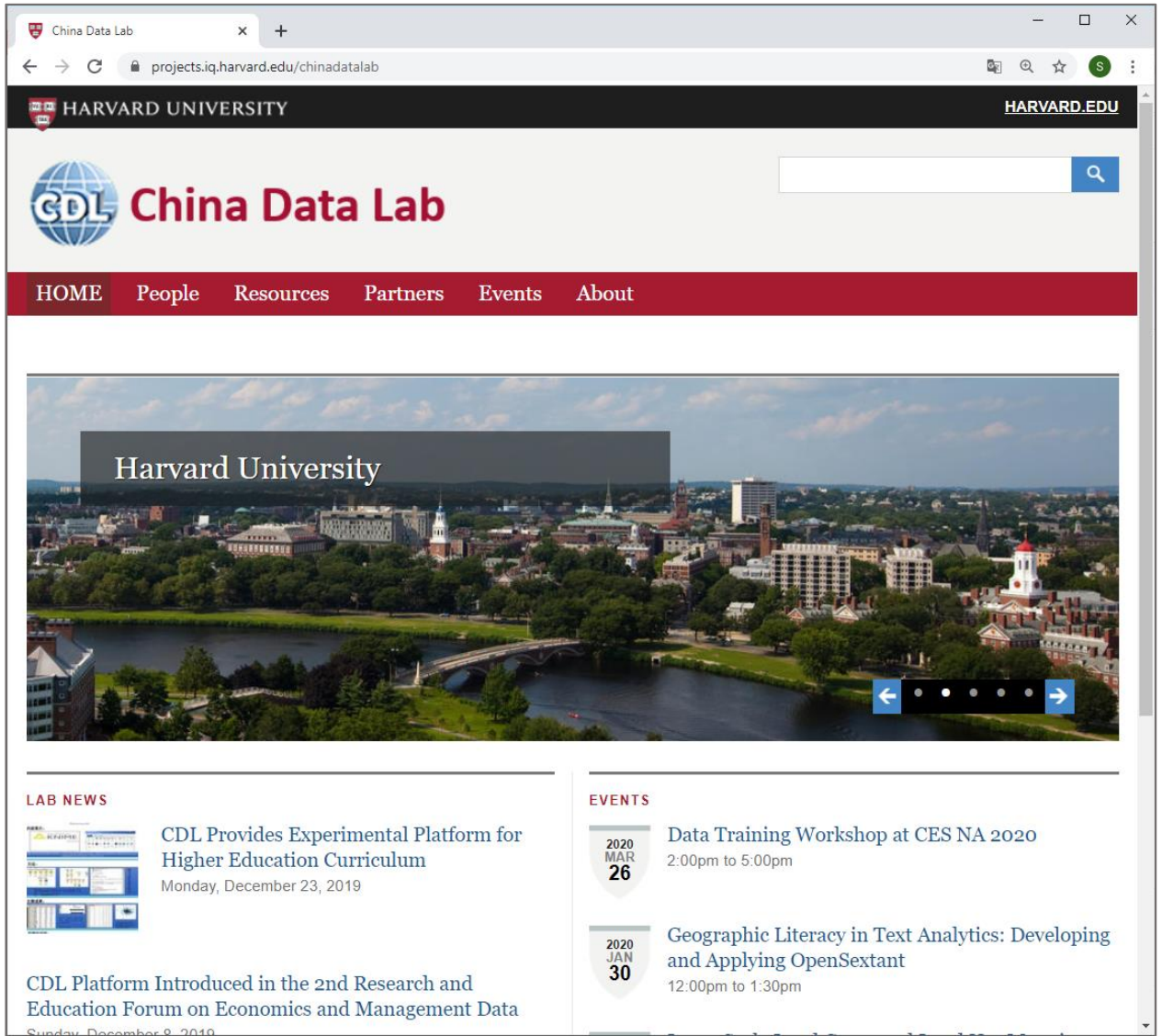
The Online Applications:

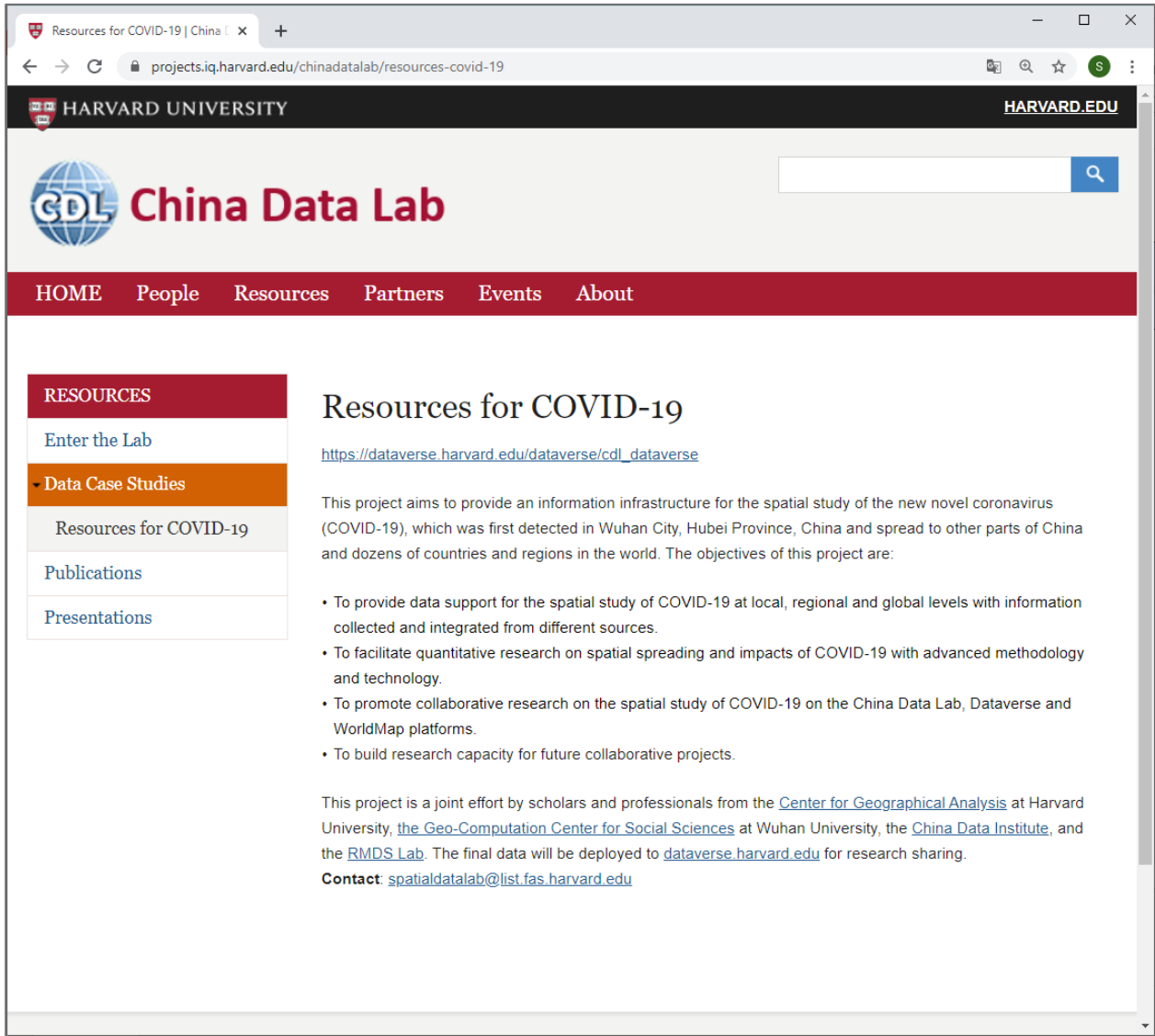
- **The Online Application in English**
https://harvard.az1.qualtrics.com/jfe/form/SV_9Qy3p1jOE1SbIKF

- **The Online Application in Chinese**
https://harvard.az1.qualtrics.com/jfe/form/SV_djwCFDTkJtOGiBT

Project Home: Resources for COVID-19 项目网站

http://chinadatalab.net





Websites and Contacts主要网站与联系方式

中国数据实验室项目网站 China Data Lab

https://projects.iq.harvard.edu/chinadatalab

中国数据在线China Data Online

http://china-data-online.com

联系方式Contacts

SpatialDataLab@list.fas.harvard.edu

office@chinadatacenter.net