FIELD GUIDE · 101 → 301

UK Data Service

数据简介与使用说明书

一本按美国课堂思路写的“入门到上手”手册：先告诉你为什么值得学，再画出全景地图，然后一步步带你从注册、找数据、申请权限，走到下载、合并、分析和规范引用。每章都有学习目标、类比、易错点和自测题。

1967
起步年份，近 60 年持续建设

3 级
Open · Safeguarded · Controlled 访问体系

£37.5m
UKRI/ESRC 资助，保障运行至 2030 年

Stata
SPSS / TAB 三种标准下载格式

00 怎么用这本手册
01 Big Picture：UKDS 是什么
02 藏了什么：数据全景地图
03 访问等级与 Five Safes
04 Step by Step：从注册到下载
05 找对数据：检索与读记录
06 深潜：Understanding Society 文件体系
07 Worked Example · 选题灵感库
08 易错点与高手习惯
09 资源 · FAQ · 术语 · 速查表

整理依据：ukdataservice.ac.uk 官网（About、Access conditions、Types of data access、Download & order、Formats、Using data、Citing data、Learning Hub）及数据目录 SN 6614、SN 5050 记录，读取于 2026-11-18。政策会更新，以官网为准。

HOW TO USE THIS GUIDE

00 怎么用这本手册

THE BIG IDEA

UK Data Service（UKDS）不是“一个数据库”，而是**一座图书馆**：里面有几千个独立的数据集（study），每个都有自己的编号（SN）、借阅规则（access level）和说明书（documentation）。学会它，本质上是学会三件事：**找书、办借书证、读说明书**。

美国学生学新东西，通常走一条固定路线：**Why → What → How → Practice → Reflect**。先知道为什么值得花时间，再建立整体地图，然后按步骤操作，用真实例子练手，最后用自测题检查是否真的掌握。本手册每一章都按这个节奏写。

手册里的五种“提示框”

框	作用
Learning Objectives	每章开头：读完你应该能做到什么。读之前先看，读完再对照。
Analogy 类比	把陌生概念映射到你熟悉的东西，比如 NHANES、MIMIC、PhysioNet。
Pro Tip	熟手才知道的捷径。
Watch Out	新手最常踩的坑，尤其是会导致结果错误或违反协议的。
Check Your Understanding	章末小问题，答案在第 9 章。

你现在在哪一步？

已完成

已注册 UKDS 账号；已建立非商业项目“Household winter warmth adequacy and subsequent pain interference”（有效期至 2028-11-14）；SN 6614 Understanding Society 已获批；End User Licence 已于 2026-11-13 签署。

下一步

按第 6–7 章完成本地下载（Stata 格式）、按 pidp 合并多波数据、处理负值缺失码与权重；第 8 章的“输出披露”和“数据销毁”规则从第一天就要遵守。

账号状态取自 2026-11-17 读取的 UKDS 账户页面。账户联系邮箱为单位邮箱，UKDS 的通知可能发到那里，建议留意。

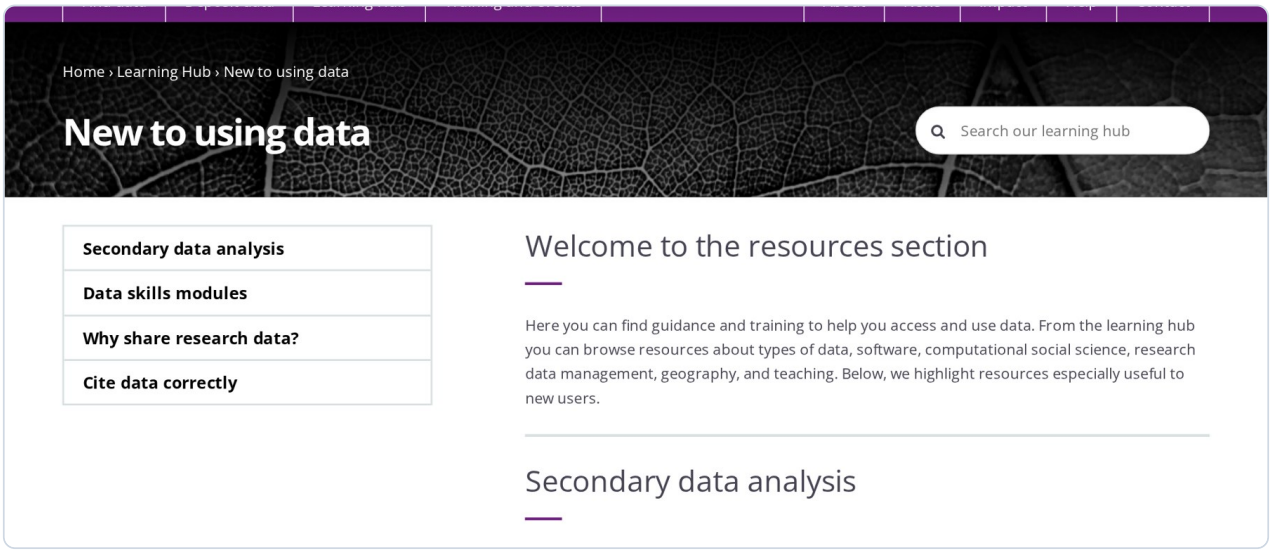


图 0-1　UKDS Learning Hub「New to using data」页面：本手册的主要参照之一（2026-11-18 截图）

THE BIG PICTURE

01 UKDS 是什么，为什么值得学

LEARNING OBJECTIVES

读完本章，你能：① 用一句话说清 UKDS 的定位；② 说出它由谁资助、由谁运营；③ 解释它为什么对麻醉与围术期研究者有用。

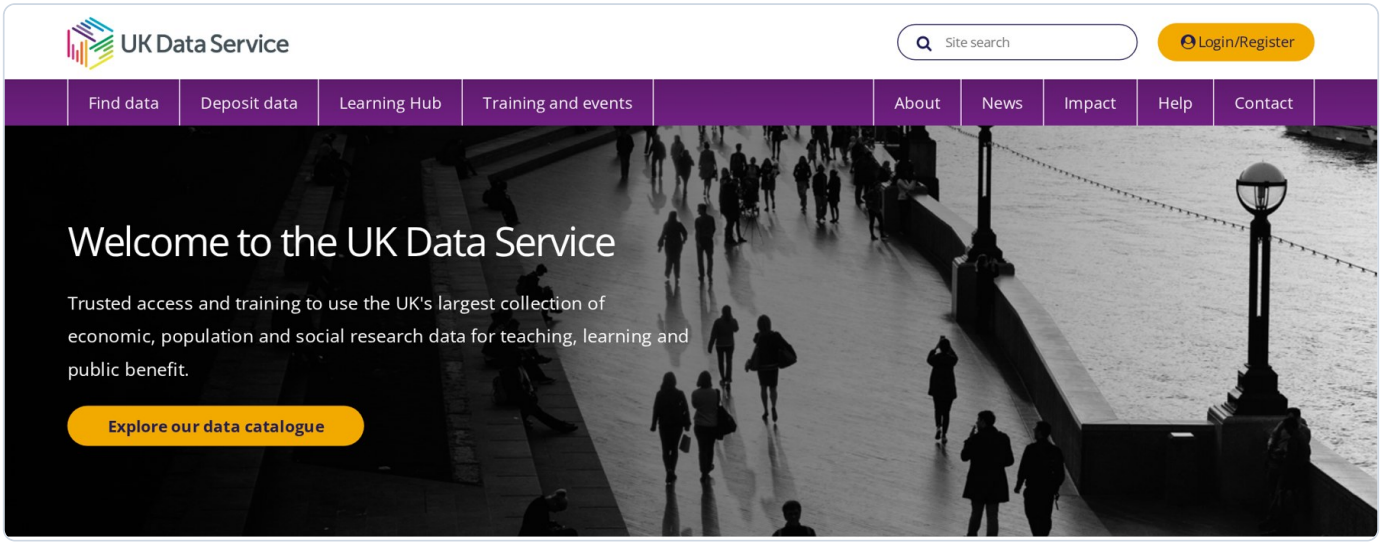


图 1-1　UK Data Service 官网首页（ukdataservice.ac.uk，2026-11-18 截图）

一句话定义

UK Data Service 是英国**经济、人口与社会研究数据的主要存储库**，由 UKRI 通过经济与社会研究理事会（ESRC）资助，自 1967 年起积累，是同类中规模最大的可信数字档案。它**不拥有**所提供的数据，而是按数据所有者（政府部门、ONS、研究团队）商定的条件提供访问。

1967	2011	2020	3 种	2030
开始积累数据	SecureLab 启用	按《数字经济法》获认证	标准下载格式	当前资助覆盖至

Analogy　类比

如果你熟悉美国的 **ICPSR**（密歇根大学的社会科学数据联盟），UKDS 就是英国版 ICPSR；如果更熟悉 **PhysioNet**，可以这样理解：PhysioNet 管 MIMIC 这类临床数据，UKDS 管的是“人们的生活”——收入、住房、就业、家庭、健康自评、衰老轨迹。它们互补，而不是替代。

谁在运营

资助方

UKRI / ESRC。新一轮 3,750 万英镑资助保障服务运行至 2030 年。

合作机构

University of Essex（牵头，UK Data Archive 所在地）、University of Manchester、Jisc、UCL、University of Edinburgh。

安全资质

UK Data Archive 是首个获 ISO 27001 认证的大学机构；2020 年获英国统计局按《数字经济法 2017》认证；参与制定 Five Safes 框架。

为什么一个麻醉医生要学它

你关心的问题	UKDS 能提供什么
衰弱、认知、老年患者	ELSA（50 岁以上英格兰人群），含体检、步速、认知、加速度计（Wave 10）和临终访谈。
社会决定因素与慢性疼痛	Understanding Society：每年随访全户成员，含 SF-12、GHQ-12、住房、能源、收入等。
生命早期暴露与成年结局	CLS 出生队列（1958、1970、千禧队列），可做几十年跨度的生命历程分析。
与 NHANES/CHARLS 做跨国对照	ELSA 提供 Gateway Harmonized 版本，便于与 HRS、CHARLS 等姊妹研究对比。

Pro Tip
审稿人常问“为什么用英国数据”。最有力的回答通常是：**纵向重复测量 + 全户设计 + 社会经济变量的深度**，这三点恰好是医院电子病历数据库（MIMIC、VitalDB）缺的。

它能做什么，不能做什么

能做 ✓
- 提供去标识化的调查、队列、普查与宏观数据
- 提供文档、引用工具、培训与 Helpdesk
- 代为向 ICPSR 等国外档案馆订购数据
- 为敏感数据提供远程安全分析环境

不能做 ✗
- 替你决定数据所有者设定的访问规则
- 为你定制 UKDS 已有数据的子集
- 提供统计软件支持或格式选择建议
- 出售数据（版权始终属于数据所有者）

Check Your Understanding 1
UKDS 拥有它提供的数据吗？如果某个数据集的使用规则让你困惑，最终由谁决定？

02 WHAT'S INSIDE 藏了什么：数据全景地图

LEARNING OBJECTIVES
读完本章，你能：① 区分横断面调查、纵向面板、出生队列、普查、国际宏观数据和质性数据；② 说出至少 4 个与健康研究直接相关的核心数据系列。

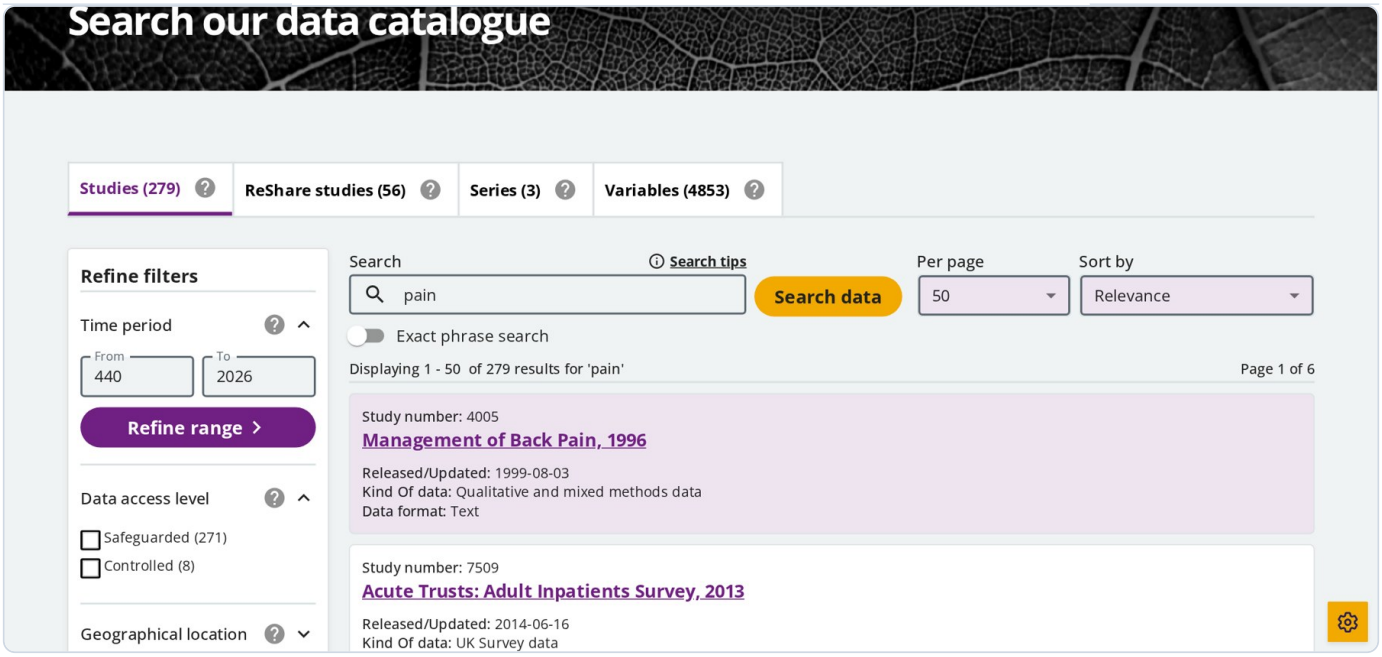


图 2-1　在数据目录搜索 “pain”：279 项研究、4,853 个变量，左侧可按时间与访问等级筛选（2026-11-18 截图）

主要数据提供方

官网列出的“策展型存放者”包括：Centre for Longitudinal Studies、ONS、NHS Digital、Department for Work and Pensions、Home Office、Department for Transport、DEFRA、ISER（Understanding Society 所属）、NatCen Social Research、Verian、NISRA，以及 Eurostat、IMF、OECD、UNIDO、World Bank 等国际机构。

两个最值得你先认识的数据集

项目	Understanding Society（SN 6614）	ELSA（SN 5050）
当前版本	Waves 1–15（2009–2024）+ 协调后 BHPS Waves 1–18（1991–2009）	Waves 0–11（1998–2024）
人群	英国全户成员，含 16 岁以上成人、10–15 岁青少年自填问卷	英格兰 50 岁以上人群；Wave 11 新样本来自 Family Resources Survey
规模	Wave 1 样本超过 40,000 户（实际访谈约 30,000 户）	因数据文件而异，见数据字典
访问等级	Safeguarded EUL	Safeguarded EUL，另有附加保密条款
健康亮点	SF-12、GHQ-12；生物标志物与遗传评分另见 SN 7251	护士访视、认知、Wave 10 加速度计、临终访谈
DOI（当前）	10.5255/UKDA-SN-6614-22	10.5255/UKDA-SN-5050-37

Understanding Society 里的健康数据清单

对临床研究者来说，UKHLS 最被低估的是它的**生物学层**。除了每年问卷里的量表，还有护士访视、血液标志物、表观遗传时钟、蛋白组学和多基因评分，分散在不同的 SN 里。

层次	数据集	内容
问卷量表	SN 6614，每波	GHQ-12、SF-12、WEMWBS、CAGE、IPAQ、SDQ；自报慢病与健康行为；孕期与分娩（PEACH 汇总文件 SN 9075）
护士体检	Wave 2–3	约 13,500 名成人：血压、身高体重、腰围、体脂、握力、肺功能；Wave 16（2024–2026 现场）再次采集，预计 2026 年底发布
血液标志物	SN 7251	血脂、HbA1c、hs-CRP、纤维蛋白原、CMV、血红蛋白、铁蛋白、肝肾功能、睾酮、DHEAS、IGF-1；5 种表观遗传年龄算法；心血管代谢与神经两组蛋白组学面板
遗传	SN 7587（SL）	多基因评分；详细基因组与表观基因组数据需向研究团队申请

来源：Understanding Society 网站 Health and biomarkers 页面，2026-11-18 读取。

Pro Tip　给麻醉科研者

hs-CRP、纤维蛋白原这类炎症指标，配合后续多波 SF-12 疼痛条目，可以做“基线炎症负荷 → 慢性疼痛轨迹”的纵向研究；表观遗传时钟则可作为“生物学年龄”替代衰弱测量。两者都在 SN 7251，下载前请在目录 Access 标签确认其访问等级与附加条件。

Watch Out

ELSA 不允许“个人/无机构隶属”使用，并要求你承诺：不尝试把 Wave 0 数据与 Health Survey for England 链接、不用于识别参与者、不尝试识别抽样地理单元。下载前请逐条读 Access 页的附加条件。

Check Your Understanding 2

你想研究“1970 年出生者童年家庭收入与中年慢性疼痛的关系”，应该先去哪个板块找数据?

ACCESS LEVELS

03 访问等级与 Five Safes

LEARNING OBJECTIVES

读完本章，你能：① 判断一个数据集属于哪一级访问；② 说出每一级需要你做什么；③ 解释为什么官方建议“先用 EUL 版本”。

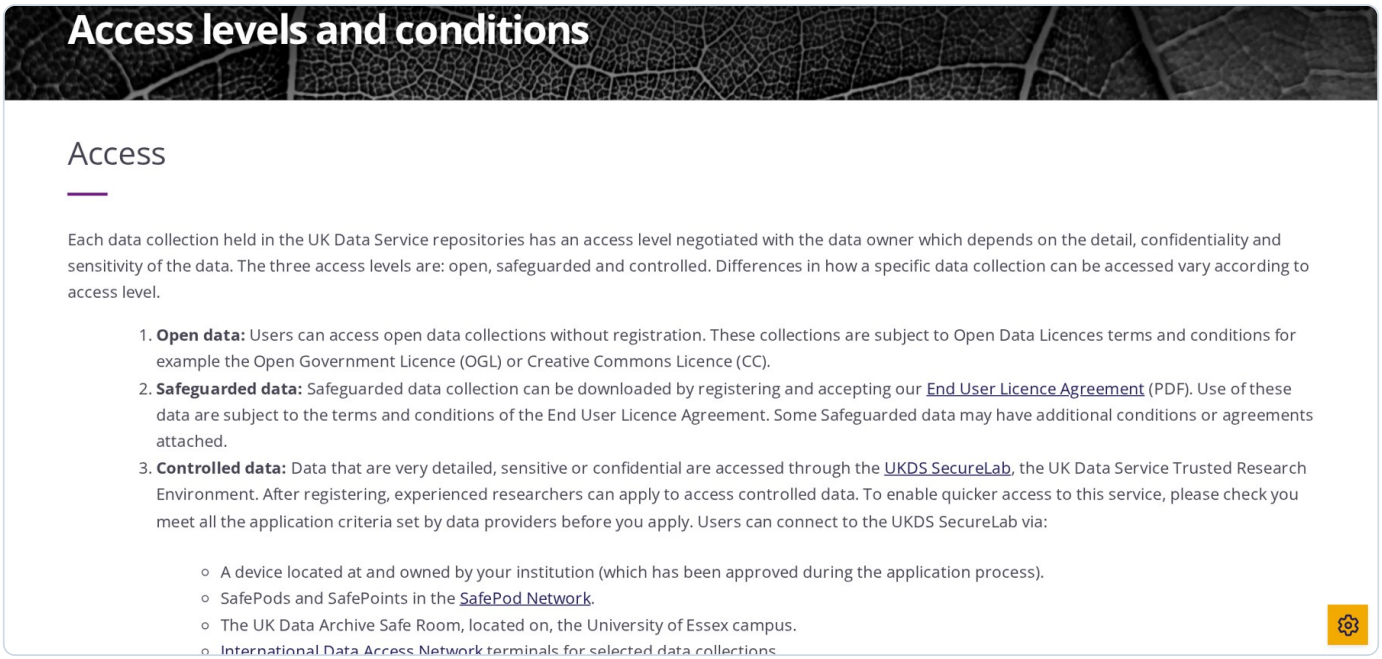


图 3-1　UKDS 官网 Access levels and conditions 页面：Open、Safeguarded、Controlled 三级说明

Safeguarded 内部还有四档

档位	你需要做什么	能否商用/教学
EUL 标准	注册并接受 End User Licence	可申请
EUL + Special Conditions	额外接受特殊条件，如发表前审查	视条件
EUL + 数据所有者许可	填写用途说明，等待存放者批准	视条件
Special Licence（SL）	填写 SL 表格交 Helpdesk，转数据所有者审批，签 SL 用户协议	不可商用、不可教学

法律上的区别

Open 与 Safeguarded 数据在法律上**都不是“个人数据”**；两者的区别在于 Safeguarded 仍有与其他数据链接后被识别的残余风险，所以要“知道谁在用、用来干什么”。Controlled 数据可被界定为个人数据，已去标识但披露风险高。

Analogy　类比
这和 MIMIC 很像：MIMIC-IV 需要 CITI 培训 + PhysioNet credentialed access + 数据使用协议，相当于 UKDS 的 Safeguarded/Special Licence 之间；而 UKDS 的 SecureLab 更接近“数据不出门、只能在安全机房里分析”的模式，类似美国 Census 的 FSRDC。

同一项研究，三个版本：以 Understanding Society 为例

版本	SN	多出来的内容
EUL	6614	标准版本，出生只到年份，部分收入变量顶部截断（top-coded）
SL	6931	出生月份、更细的国家与职业编码、未截断的收入变量等
Secure	6676	SL 全部变量 + 出生日 + 1 米精度的邮编网格坐标

Pro Tip
官方原话：**先拿 EUL 版本，看够不够用**。你的保暖–疼痛课题按年份分析即可，EUL 足够；只有当你需要精确到月的季节暴露或小区域地理链接时，才值得走 SL 申请。

Watch Out
商业用途只能用 EUL 数据，且收取管理费（英国纳税人每项目 £450 + 每个数据集 £50 + VAT；非英国纳税人 £540 + £60）。学术论文属于 Non-commercial，不收费。

Controlled 数据背后的逻辑：Five Safes



图 3-2　UKDS 官方 Five Safes 图标



图 3-3　杜伦大学的 SafePod：可在其中远程访问 Controlled 数据（图源 SafePod Network）

Reality Check　你在中国，能不能用 Controlled 数据?

官网写明：除通过 IDAN 协议提供的部分非 ONS 数据外，Controlled 数据**只对“人在英国、且隶属于愿承担法律责任的英国机构”的研究者开放**；博士生须与导师联合申请，导师也要完成表格和培训；项目必须证明公共利益。数据不能下载，只能在机构终端、Safe Room 或 SafePod 远程分析。对你而言，**EUL 和 Special Licence 才是主战场**；若将来确需 Secure 数据，现实路径是与英国合作者联合申请。

Check Your Understanding 3

你的项目需要受访者“出生月份”，应该申请 SN 6614、6931 还是 6676？为什么不直接申请 6676?

STEP BY STEP

04 从注册到下载：完整流程

LEARNING OBJECTIVES

读完本章，你能独立完成：找到数据集 → 加入账户 → 建项目 → 解锁权限 → 选格式下载。

1. **在 Data Catalogue 找到研究。**可按主题浏览（Ageing、Health、Food and nutrition 等），或直接搜 SN。点 **Access** 标签确认访问等级和附加条件。
2. **Add to account。**未注册会被引导先注册。注册时选的是 **user type**（你是谁），不是项目用途。
3. **进入 My account，把数据集放进项目。**可放进已有项目，也可新建。
4. **选择用途类型：**Non-commercial（研究、学位论文）、Commercial（以营利为主要目的）、Teaching（课堂）。学术论文选 Non-commercial。
5. **看状态。**显示 **Active** 可直接下载；显示 **Request access** 就点开，逐条完成 Complete actions（接受特殊条件、下载并填写表格等）。Special Licence 的勾选由工作人员手动更新，暂时没打勾属正常。
6. **选格式下载。**主馆藏通常提供 Stata、SPSS、TAB 三种格式。

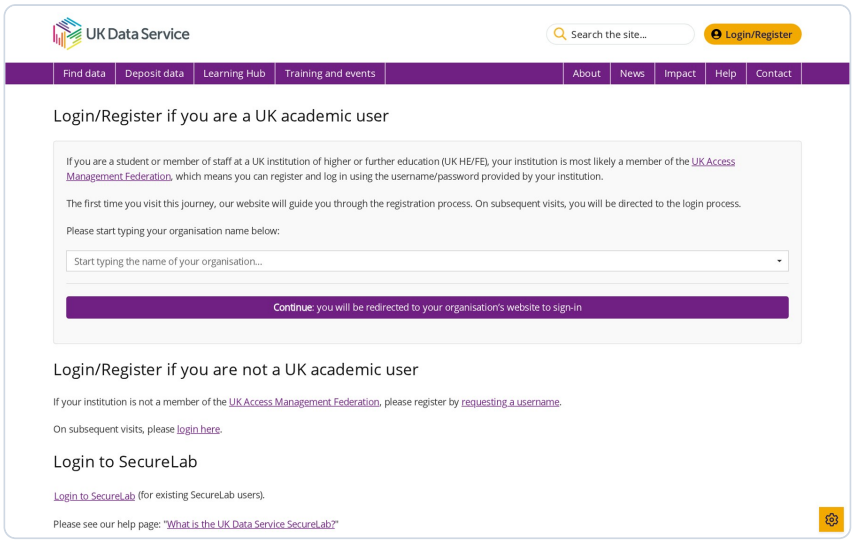


图 4-1　登录/注册页：英国高校走 UK Access Management Federation，非英国高校走“申请用户名”

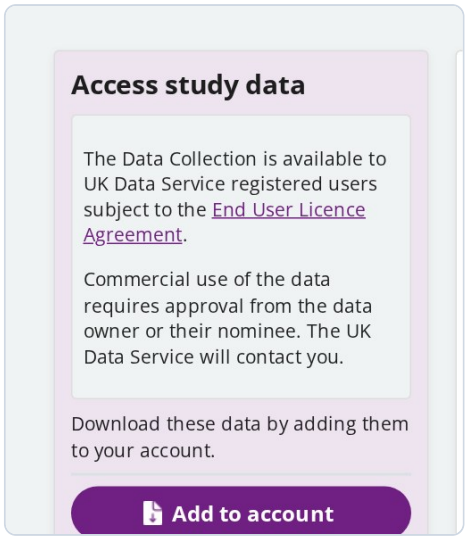


图 4-2　目录记录左栏的“Add to account”按钮

该选哪种格式

格式	适合谁	注意
Stata	熟悉微观数据分析的研究者；R、SAS 也能读	Understanding Society 由存放者以 Stata 格式提供，官方**强烈建议用 Stata 分析**，转换格式可能出问题
SPSS	统计软件新手	老数据可能是 .por，需在 SPSS 内导入
TAB	没有统计软件的人	制表符分隔，需在 Excel 里“打开 → 所有文件 → 分隔符向导”

Pro Tip
你用 R 也完全可以：`haven::read_dta()` 能保留 Stata 的变量标签和值标签。但大文件超过 2,047 个变量时，Stata 需要 SE 或 MP 版本；ELSA 的 Harmonized 文件有 11,779 个变量，必须 Stata MP。

Watch Out
“Special request” 只用于向国外档案馆（如 ICPSR）代订数据，**不能**用来请求 UKDS 已有数据的定制子集。

Check Your Understanding 4
你是医院职工（user type），用 UKDS 数据写 SCI 论文，项目用途应选哪一类？会产生费用吗？

05 SEARCH & READ 找对数据：检索与读懂记录

LEARNING OBJECTIVES
读完本章，你能：① 用“三层检索法”从研究问题定位到具体变量；② 在 2 分钟内从一条目录记录里读出编号、版本、访问条件、引用、抽样、权重和已知问题。

三层检索法：问题 → 数据集 → 变量

第 1 层 选数据集：在 Data Catalogue 搜关键词，用左栏按时间和访问等级筛选（见图 2-1），产出 1–3 个候选 SN。**第 2 层 确认变量：**目录的 Variables 标签，或研究自带的变量检索（UKHLS 的 Main Survey variables + index terms；跨研究用 CLOSER Discovery），产出变量名与出现波次。**第 3 层 读文档：**Documentation 标签里的 User Guide、问卷和数据字典，产出可执行的分析计划。

选哪个数据集？一张决策表

你的研究问题长这样	优先考虑	理由
全年龄段、家庭环境、每年随访	Understanding Society（SN 6614）	全户设计，年度追踪
50 岁以上、衰老、认知、退休	ELSA（SN 5050）	老年专题，含体检
童年暴露 → 中年结局	CLS 出生队列	跨度几十年
需要血液标志物	UKHLS 护士评估（SN 7251）	Wave 2–3 血样
需要客观活动/睡眠	ELSA Wave 10 加速度计	Axivity AX3 8 天佩戴
与中国 CHARLS 做对比	Gateway Harmonized ELSA	变量已统一口径

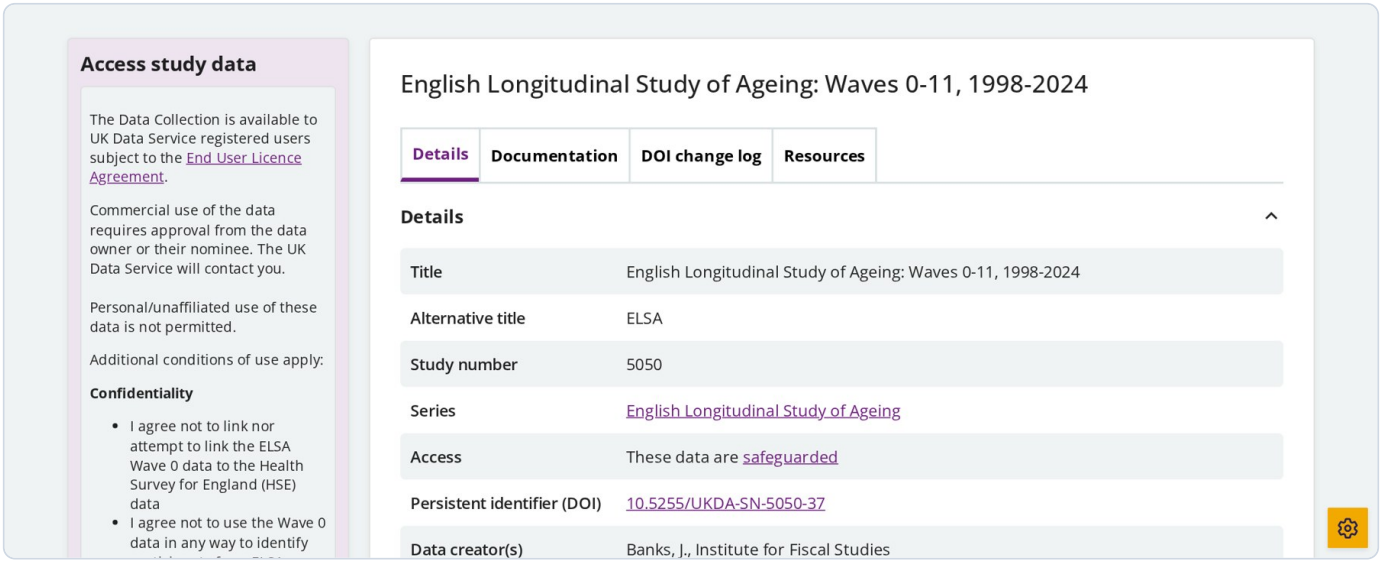


图 5-1　ELSA（SN 5050）目录记录 Details 页：左栏列出附加保密条款，右侧为编号、系列、访问等级与 DOI

读懂记录：四个标签页分别看什么

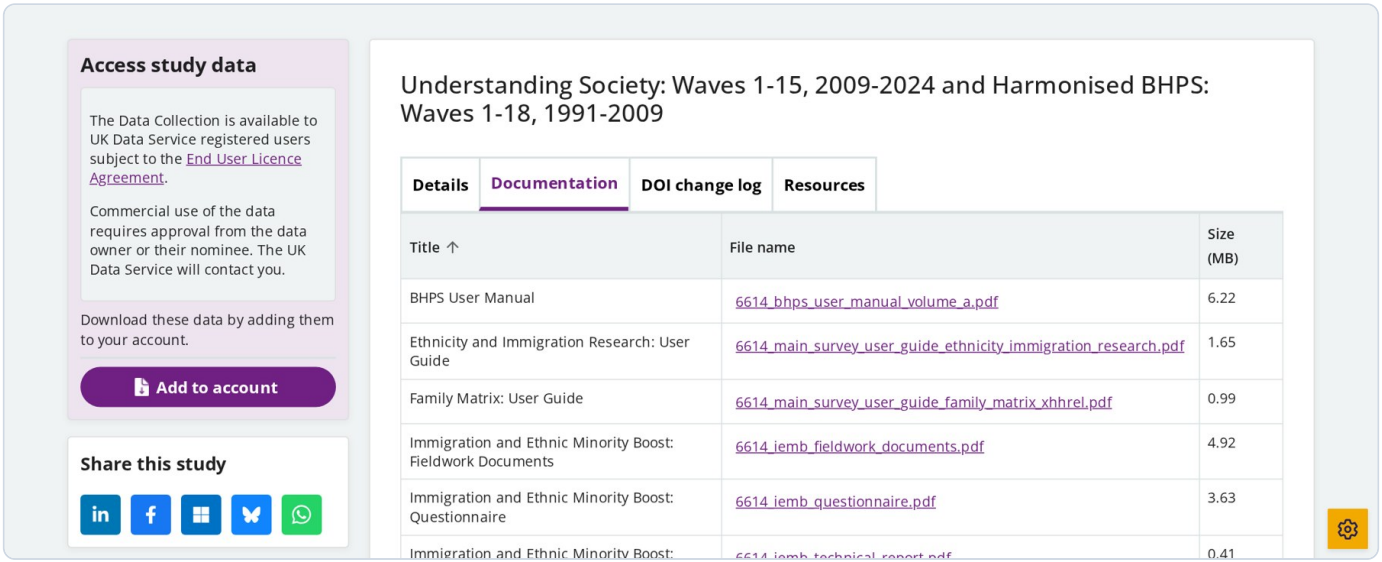


图 5-2　SN 6614 的 Documentation 标签：User Guide、问卷、数据字典等 PDF 均可直接下载

标签	看什么
Details	标题、SN、访问等级、DOI、引用、摘要、抽样、权重、数据集文件清单
Documentation	User Guide、问卷、数据字典、技术报告。**真正的“说明书”在这里。**
DOI change log	每个版本改了什么。数据更新后结果对不上，先查这里。
Resources	相关出版物、教程链接

Analogy　类比
就像读一篇 RCT 先看 CONSORT 流程图和方法学部分，读数据集先看 Documentation 里的 User Guide。跳过它直接写代码，就像不看纳入排除标准就开始解读结果。

Watch Out　ELSA 健康状况变量
目录摘要明确写着：ELSA 团队发现历史数据中部分健康状况测量有问题，涉及眼病、心血管疾病（高血压、心绞痛、心梗、心衰、糖尿病、卒中等）和慢性病（慢阻肺、哮喘、关节炎、骨质疏松、癌症、帕金森、痴呆等）。做这类分析前，**必须先读 ELSA User Guide**，有疑问可写信给 elsadata@natcen.ac.uk。

Pro Tip
除了 UKDS 目录，还有两个“只看变量、不用下载”的工具：**Nesstar 目录**（可在线看变量描述和频数）和 **CLOSER Discovery**（可跨多个英国纵向研究搜问卷和变量）。写申请摘要前先用它们确认变量存在。

Check Your Understanding 5
半年后你重新下载 SN 6614，发现样本量和之前差了几百人。第一步应该看哪里？

DEEP DIVE
06 Understanding Society 文件体系

LEARNING OBJECTIVES
读完本章，你能：① 看懂文件名前缀和后缀；② 知道个人、家庭、跨波文件分别装什么；③ 用 pidp 和 w_hidp 正确合并。

THE BIG IDEA
Understanding Society 是**“宽表按波次拆开、按层级拆开”**的数据：每一波一组文件，每组再分个人、家庭、儿童、收入等。分析的第一步永远是**按 ID 把需要的片段拼起来**。

文件名规则

前缀	含义
`a_ … o_`	Understanding Society 第 1–15 波
`ba_ … br_`	协调后 BHPS 第 1–18 波
`x…`	跨波文件，如 xwaveid、xwavedat、xhhrel

后缀	内容
`indresp`	16 岁以上成人完整访谈（含 SF-12、GHQ-12）
`hhresp`	家庭问卷：住房、取暖、租金、物质剥夺
`indall`	家庭内所有成员的基本信息
`youth / child`	10–15 岁自填问卷 / 家长代答的儿童信息
`income · egoalt`	收入来源明细 • 家庭成员两两关系

Filename	Description
w_indall bw_indall	Household grid data for all persons in household, including children and non-respondents. *The variable pidp or the combination of variables "w_hidp w_pno" uniquely identifies each row of w_indall. The variable pidp or the combination of variables "bw_hidp bw_pno" uniquely identifies each row of bw_indall.*
w_hhresp bw_hhresp	Substantive data collected from responding households. *The variable w_hidp uniquely identifies each row in b_hhresp. The variable bw_hidp uniquely identifies each row in bw_hhresp.*
w_indresp bw_indresp	Substantive data collected from responding adults (16+) including proxies. Some information collected in these questionnaires are better presented in multi-level files (see Table 2). *The variable pidp or the combination of variables "w_hidp w_pno" uniquely identifies each row of w_indresp. The variable pidp or the combination of variables "bw_hidp bw_pno" uniquely identifies each row of bw_indresp.*
w_youth bw_youth	Substantive data from youth questionnaire. *The variable pidp uniquely identifies each row in these files. The variable pidp or the combination of variables "w_hidp w_pno" uniquely identifies each row of w_youthl. The variable pidp or the combination of variables "bw_hidp bw_pno" uniquely identifies each row of bw_youth.*
w_child	Childcare, consents and school information of all children (0-15 years) in the household. This is a derived data file collecting information pertaining to children as reported by their parents and guardians in the adult questionnaire. *The variable pidp or the combination of variables "w_hidp w_pno" uniquely identifies each row of w_child.*

图 6-1　UKHLS User Guide 原文 Table 1：主要数据文件与唯一标识变量

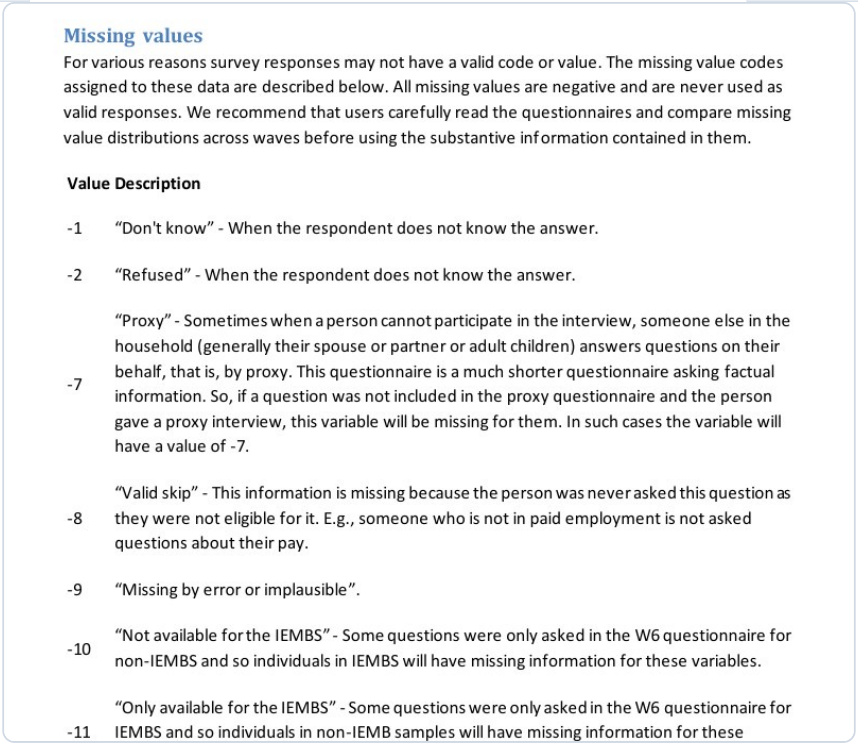
两个 ID，记住就不会合并错

ID	含义	用法
`pidp`	跨波不变的个人 ID	把同一个人不同波次的记录连起来
`w_hidp`	该波次的家庭 ID（每波前缀不同，如 a_hidp）	只在同一波内把个人与家庭文件连起来；**不能跨波使用**，因为人会搬家、分户

负值不是数据，是“为什么缺失”

编码	含义（User Guide 原文释义）	编码	含义
`-1`	Don't know，不知道	`-9`	错误或不合理导致缺失
`-2`	Refused，拒答	`-10`	非 IEMBS 样本未问此题
`-7`	Proxy，代答问卷未含此题	`-11`	仅 IEMBS 样本被问到
`-8`	Valid skip，不适用（跳题）	`-20/-21`	xwavedat 中无 BHPS / 无 UKHLS 数据

Watch Out
直接对含负值的变量求均值或做回归，会把“拒答”当成真实的低分。**第一行清洗代码就应该把负值统一转为缺失**，并在方法部分说明缺失处理。



Missing values

For various reasons survey responses may not have a valid code or value. The missing value codes assigned to these data are described below. All missing values are negative and are never used as valid responses. We recommend that users carefully read the questionnaires and compare missing value distributions across waves before using the substantive information contained in them.

Value	Description
-1	"Don't know" - When the respondent does not know the answer.
-2	"Refused" - When the respondent does not know the answer.
-7	"Proxy" - Sometimes when a person cannot participate in the interview, someone else in the household (generally their spouse or partner or adult children) answers questions on their behalf, that is, by proxy. This questionnaire is a much shorter questionnaire asking factual information. So, if a question was not included in the proxy questionnaire and the person gave a proxy interview, this variable will be missing for them. In such cases the variable will have a value of -7.
-8	"Valid skip" - This information is missing because the person was never asked this question as they were not eligible for it. E.g., someone who is not in paid employment is not asked questions about their pay.
-9	"Missing by error or implausible".
-10	"Not available for the IEMBS" - Some questions were only asked in the W6 questionnaire for non-IEMBS and so individuals in IEMBS will have missing information for these variables.
-11	"Only available for the IEMBS" - Some questions were only asked in the W6 questionnaire for IEMBS and so individuals in non-IEMB samples will have missing information for these

图 6-2　UKHLS 主调查 User Guide（2026-02-23 修订版）第 25 页缺失值说明原文

权重：纵向分析的必修课

Understanding Society 和 ELSA 都是多阶段分层随机抽样，目录明确写着“Weighting used”。UKDS 学习中心提供《What is weighting?》《What is complex sample design?》两份指南，Understanding Society 另有权重专题网络研讨会。基本原则：横断面描述用横断面权重；多波纵向分析用对应波次组合的纵向权重；同时考虑 PSU 与 strata 以获得正确标准误。具体变量名以 User Guide 的权重章节为准。

Check Your Understanding 6
为什么不能用 a_hidp 去匹配第 5 波的家庭文件 e_hhresp?

Midpoint Review　前 6 章复盘

美国教材习惯在动手练习前停一下。一次合规的 UKDS 分析要依次过六关：选对 SN → 选对访问等级 → 读 User Guide → 用 pidp/w_hidp 正确合并 → 清洗与加权 → 引用 DOI 并避免泄露小格子。

如果只记住三句话

1. 先 EUL，后 SL，Secure 是例外。
2. pidp 连人，w_hidp 连家，负值先清。
3. 写论文必写 SN、版本和 DOI。

和你熟悉的数据库对照

NHANES 是横断面 + 复杂抽样，权重逻辑相通；MIMIC 是住院病例，没有家庭层级；UKHLS 则是“人 × 家庭 × 年份”三维结构，合并步骤比前两者多一层。

WORKED EXAMPLE

07 演练：你的保暖–疼痛课题

LEARNING OBJECTIVES

用你已获批的项目，把前 6 章串起来走一遍：问题 → 变量 → 合并 → 清洗 → 分析 → 引用。

研究问题

家庭冬季取暖是否充足，与随后的疼痛干扰是否相关?

暴露

hheat（家庭文件 hhresp）：家里冬天能否保持足够温暖

结局

scsf5（个人文件 indresp）：SF-12 中“疼痛对工作的干扰”条目

Stata 骨架代码

```
* 1) 循环各波：个人文件 + 家庭文件，同波内按 w_hidp 合并
local waves a b c d e f g h i j k l m n o
local i = 1
foreach w of local waves {
    use pidp `w'_hidp `w'_scsf5 `w'_age_dv `w'_sex_dv using "`w'_indresp.dta", clear
    merge m:1 `w'_hidp using "`w'_hhresp.dta", keepusing(`w'_hheat) keep(master match) nogen
    rename `w'_* *            // 去掉波次前缀
    gen wave = `i'
    tempfile t`i'
    save `t`i''
    local ++i
}
* 2) 纵向追加成长格式
use `t1', clear
forvalues k = 2/15 {
    append using `t`k''
}
* 3) 负值统一设为缺失（仅数值变量）
ds, has(type numeric)
mvdecode `r(varlist)', mv(-21/-1)
* 4) 设定面板，暴露滞后一波
xtset pidp wave
gen hheat_lag = L.hheat
```

骨架用于说明逻辑。变量是否在每一波都有、取值方向、权重变量名，请以 SN 6614 的 Documentation 和变量检索为准；hheat 在部分波次可能未询问，循环中需做存在性检查。

分析路线建议

1. **描述：**按波次画出“取暖不足”比例和疼痛干扰分布，确认数据质量。
2. **主模型：**滞后一波暴露 → 当期结局，混合效应有序 logistic 或线性模型，调整年龄、性别、收入、慢病。
3. **因果加强：**个体固定效应模型，只用同一人内部的变化，自动控制不随时间变化的混杂。
4. **敏感性：**加权 vs 不加权；排除代答；排除仅出现一波者。

项目路线图（编者建议）

阶段	建议时间	产出
下载与合并	2026 Q4	长格式分析数据集与变量字典
清洗与描述	2026 Q4 – 2027 Q1	Table 1、缺失模式报告
主模型	2027 Q1 – Q2	主效应估计与图表
敏感性分析	2027 Q2	加权、多重插补、E 值
写作与投稿	2027 Q2 – Q4	稿件与数据可用性声明
项目到期与数据销毁	2028-11-14 前	按 EUL 删除本地数据

麻醉与围术期选题灵感库

以下是基于本手册所列数据可行性提出的方向，供你筛选、查新后再立项。

01 基线炎症负荷与慢性疼痛轨迹
UKHLS 护士评估 hs-CRP、纤维蛋白原（SN 7251）→ 后续多波 SF-12 疼痛干扰。
设计：纵向轨迹 + 混合模型；亮点：生物标志物与社会因素同框。

02 表观遗传年龄加速与疼痛、功能下降
SN 7251 的 5 种表观遗传时钟 → SF-12 躯体分量与疼痛条目。
亮点：把“生物学年龄”引入围术期风险分层的上游证据。

03 能源价格冲击的准实验研究
UKHLS 多波取暖与能源变量 + 2021–2023 价格上涨 → 疼痛与心理健康。
设计：双重差分或事件研究；注意与 2026 年 Cha 等研究区分。

04 客观睡眠、活动与认知储备
ELSA Wave 10 加速度计（步数、睡眠分期）→ 认知测试与后续认知变化。
临床延伸：术前睡眠与活动作为术后认知障碍风险标志的人群证据。

05 英中对比：慢性疼痛与抑郁共病
Gateway Harmonized ELSA × CHARLS，统一口径比较疼痛–抑郁关联。
亮点：跨国比较，契合你已熟悉的 CHARLS。

写进论文的数据声明与引用

```
University of Essex, Institute for Social and Economic Research. (2026).
Understanding Society: Waves 1-15, 2009-2024 and Harmonised BHPS: Waves 1-18,
1991-2009. [data collection]. 21st Edition. UK Data Service. SN: 6614,
DOI: http://doi.org/10.5255/UKDA-SN-6614-22
```

以上为 2026-11-18 目录页的 APA 引用。注意：同一页面在 Edition history 处写的是“22nd Edition（2026-04-21）”，与引用中的“21st Edition”不一致，**下载时以你实际下载文件附带的版本号为准**，并在论文中写明。

PITFALLS & PRO HABITS

08 易错点与高手习惯

LEARNING OBJECTIVES

读完本章，你能避开最常见的 8 个错误，并养成从第一天起就合规的工作习惯。

EUL 规定你必须做到的事

#	规定（来自 Types of data access 页面）	对你意味着什么
1	不把数据分享给任何未在 UKDS 注册的人	合作者要自己注册、加入项目；不能用网盘或微信直接传数据文件
2	未经许可不用于商业目的	学术论文没问题
3	妥善保管密码	账号不共享
4	保护个人、家庭、机构的保密性	不发布小格子频数；不尝试识别个人
5	发表时使用正确的引用与致谢	每篇论文写 SN、版本和 DOI
6	把发表成果的书目信息发给 UKDS	论文接收后发送

八个新手常见错误

① 负值当真值
-1 到 -9 是缺失原因码，必须先转缺失。

② 跨波用家庭 ID
w_hidp 只在同一波内有效，跨波用 pidp。

③ 忽略权重和设计
描述性比例不加权会偏离总体；标准误不考虑分层和聚类会偏小。

④ 不记版本号
数据会更新，不记版本就无法复现。

⑤ 跳过 User Guide
已知数据问题、跳题逻辑、量表计分都写在里面。

⑥ 一上来就申请 Secure
先确认 EUL 不够用，再升级。

⑦ 发表小格子
某亚组只有极少数人时，交叉表可能导致可识别。

⑧ 把数据传给他人
违反 EUL，也会影响你所在机构。

高手习惯：一个项目文件夹的样子

```
UKHLS_warmth_pain/
├── 00_raw/          # 原始下载，只读，不改动；记录 SN、版本、下载日期
├── 01_code/         # 01_merge.do  02_clean.do  03_models.do
├── 02_derived/      # 合并后的分析文件，可由代码完全重建
├── 03_output/       # 只放汇总表和图，不放个体级数据
└── README.txt       # 数据来源、版本、EUL 签署日期、项目编号
```

Pro Tip

研究结束或协议到期时，按 UKDS《Research Data Handling and Security》指南删除数据副本。你的项目有效期到 2028-11-14，可以提前在日历上标记。

RESOURCES · QUIZ · CHEAT SHEET

09 资源、自测答案、FAQ 与速查表

继续学习：官方免费资源

资源	适合什么时候用
Learning Hub · New to using data	第一周：二次数据分析入门、数据引用
Data skills modules（互动课程）	系统学“Introduction to survey data”，含权重与抽样设计
What is weighting? / complex sample design（PDF）	写方法学部分之前
What are hierarchical files? / Analysing change over time（PDF）	做个人–家庭合并和纵向模型时
A Guide to Weighting in Understanding Society（视频）	纵向权重选择
Health Studies User Conference 2026	看同行如何用这些数据做健康研究
CLOSER Discovery · UKDS Data Explorer	跨研究搜变量与问卷

自测答案

题	答案
1	不拥有。访问等级与条件由数据所有者商定，UKDS 按约提供。
2	出生队列板块，由 CLS 提供的 1970 British Cohort Study（BCS70）。
3	SN 6931（Special Licence）。6676 是 Secure Access，手续最重且一般不对学生开放，官方建议用够用的最低一级。
4	Non-commercial，不收费。user type 与项目用途是两回事。
5	目录记录的 DOI change log 与版本说明（如 SN 6614 第 21/22 版的 revisions 文档）。
6	w_hidp 只在当波有效，人会搬家或分户，跨波必须用 pidp 先连人，再在各自波次内连家庭。

常见问题 FAQ

Q 我人在中国，能注册和下载吗?

可以。你的账号与 SN 6614 已获批。但 Controlled 数据要求人在英国并隶属英国机构；个别数据集仅限英国高校研究者，以 Access 标签为准。

Q 合作者能直接用我下载的文件吗?

不能。EUL 禁止把数据给未注册者。合作者需自行注册并接受协议。

Q EUL 够用吗，什么时候需要 Special Licence?

按年份分析、国家级地理通常够用。需要出生月份、未截断收入或低层级地理标识时，再申请 SL（UKHLS 为 SN 6931，地理文件另有多个 SN）。

Q 用 R 而不是 Stata 可以吗?

可以用 haven 读取 .dta。但 UKHLS 官方建议用 Stata，转换格式可能出现意外问题，转换后务必核对样本量和标签。

Q 数据更新了，我要重新下载吗?

写论文时固定一个版本即可，在方法部分写明版本号和 DOI；投稿前可查 DOI change log，看是否有影响你变量的修订。

Q 遇到变量含义不清怎么办?

先查 User Guide 与问卷；UKHLS 有用户支持论坛；ELSA 数据问题可写信 elsadata@natcen.ac.uk；访问与下载问题找 UKDS Helpdesk。

术语表 Glossary

SN（Study Number） 数据集唯一编号。

EUL（End User Licence） 标准用户许可协议，Safeguarded 数据的基础门槛。

Special Licence（SL） 更细变量的许可，需单独申请并经数据所有者批准。

SecureLab UKDS 的可信研究环境，数据不可下载。

Five Safes 数据、项目、人员、环境、输出五项安全原则。

DOI 数据集的永久标识，末尾数字随版本变化。

Edition 数据集版本号，复现研究的关键。

pidp UKHLS 跨波个人 ID。

w_hidp UKHLS 当波家庭 ID。

indresp / hhresp 个人问卷文件 / 家庭问卷文件。

Proxy 由他人代答的简短问卷。

Top-coding 为防识别把极端值截断到上限。

PSU / Strata 初级抽样单元 / 分层，计算标准误需用到。

Cross-sectional weight 单波描述用权重。

Longitudinal weight 多波追踪分析用权重。

Harmonised data 不同调查间统一口径后的数据。

QualiBank UKDS 的质性数据在线检索工具。

CLOSER Discovery 跨英国纵向研究的问卷与变量检索平台。

一页速查表

找数据	datacatalogue.ukdataservice.ac.uk，按主题或 SN 搜，先看 Access 标签
访问三级	Open 免注册 • Safeguarded 注册 + EUL • Controlled SecureLab
项目用途	Non-commercial（论文） • Commercial（收费） • Teaching
下载格式	Stata（UKHLS 首选） • SPSS • TAB
UKHLS 编号	EUL 6614 • SL 6931 • Secure 6676 • 生物标志物 7251 • COVID 8644 • Innovation Panel 6849
UKHLS 前缀	a_–o_ = 第 1–15 波；ba_–br_ = BHPS 第 1–18 波；x = 跨波
ID	pidp 跨波连人；w_hidp 同波连家庭
缺失码	负值一律先转缺失（-1 不知道，-2 拒答，-7 代答，-8 不适用，-9 缺失）
ELSA	SN 5050，Waves 0–11，Harmonized 文件需 Stata MP
引用	作者（年份）. 标题. [data collection]. 版本. UK Data Service. SN. DOI
求助	UKDS Helpdesk 在线表单；help@ukdataservice.ac.uk；Understanding Society 用户支持论坛

本手册的事实依据为 UKDS 官网与数据目录公开页面（2026-11-18 读取）。缺失码依 2026-02-23 版 UKHLS User Guide，权重变量名等细节以对应 User Guide 为准；“类比”与“分析路线建议”为编者的教学性补充，不代表 UKDS 官方意见。