**艺术人文赛道博物馆数据说明**

1. **简介**

本次比赛提供了来自英国维多利亚与艾尔伯特博物馆(V&A)、法国卢浮宫博物馆（Le Louvre）、美国大都会艺术博物馆（The MET）、日本京都国立博物馆（Kyohaku）、台北故宫博物院、澳大利亚国家博物馆（National Museum of Australia）、日本正仓院（Shosoin）7家博物馆的部分开放数据，共计4312条数据（4321张图片，一条藏品数据可能对应多张图片）。

其中2到2884行的藏品元数据中包含“中国（China）”（相关字段为“来源地”或“描述”），2885到4313行藏品元数据中不包含“中国”。藏品数据表单以XLSX格式提供，藏品图片以压缩包格式提供。

此次提供的数据通过博物馆开放数据API下载，人为对数据进行了部分重构、清洗和翻译，建议选手从博物馆官网通过开放数据接口下载更多原始数据进行分析和处理。

1. **字段说明**

|  |  |  |
| --- | --- | --- |
| **字段名称** | **字段含义** | **相关说明** |
| **博物馆** | 藏品所在的博物馆 |  |
| **ID** | 藏品的命名ID | 与藏品图片一一对应，藏品图片命名格式为“ID\_藏品名称“，比如”34489\_Print”。注意，此ID在数据表单中可能不是唯一值。 |
| **Main\_ID** | 藏品的唯一ID | 该ID在数据表单中是唯一值。 |
| **版权声明** | 藏品图片的版权信息 | 来自博物馆元数据。 |
| **Credit Line** | 藏品的版权信息 | 来自博物馆元数据。包含了对藏品的创作者、捐赠者、购买者或藏品来源的致谢信息 |
| **博物馆索引号** | 藏品所在博物馆的自有索引号。 | 来自博物馆元数据。 |
| **作者** | 藏品的作者 | 来自博物馆元数据。 |
| **来源地** | 藏品的来源地 | 来自博物馆元数据。注意区分几种不同的来源地（made/manufactured/designed/photographed） |
| **时代** | 藏品被制作或被生产的时代 | 来自博物馆元数据。通常为历史纪年，中英文的表示方法不同，注意区分。 |
| **年代** | 藏品被制作或被生产的年代 | 来自博物馆元数据。通常为公元纪年，用阿拉伯数字表示。 |
| **尺寸** | 藏品的尺寸 | 来自博物馆元数据。 |
| **分类\_EN**  | 藏品的分类（英文） | 来自博物馆元数据。如果原始数据非英文，此处为DeepL机器翻译结果。 |
| **分类\_ZN** | 藏品的分类（中文） | 来自博物馆元数据。如果原始数据非中文，此处为DeepL机器翻译结果。 |
| **材质\_EN** | 藏品的材质（英文） | 来自博物馆元数据。如果原始数据非英文，此处为DeepL机器翻译结果。 |
| **材质\_ZN** | 藏品的材质（中文） | 来自博物馆元数据。如果原始数据非中文，此处为DeepL机器翻译结果。 |
| **技法\_EN** | 藏品相关的技法（英文） | 来自博物馆元数据。如果原始数据非英文，此处为DeepL机器翻译结果。 |
| **技法\_ZN** | 藏品相关的技法（中文） | 来自博物馆元数据。如果原始数据非中文，此处为DeepL机器翻译结果。 |
| **Metadata\_EN** | 使用NLP算法从藏品元数据中自动提取的关键词（英文） | 后期添加字段。如果原始数据非英文，此处为DeepL机器翻译结果。可能存在中英混杂、简繁混杂的情况，使用时请谨慎处理。 |
| **Metadata\_ZN** | 使用NLP算法从藏品元数据中自动提取的关键词（中文） | 后期添加字段。如果原始数据非中文，此处为DeepL机器翻译结果可能存在中英混杂、简繁混杂的情况，使用时请谨慎处理。 |
| **AAT\_Keyword\_EN** | 根据Metadata\_EN中的关键词与AAT词表进行了匹配，此处为匹配后的结果（英文） | 后期添加字段。AAT词表使用的是AAT-Taiwan「藝術與建築索引典」。链接如下： https://aat.teldap.tw/index.php?\_lang=en |
| **AAT\_Keyword\_ZN** | 根据Metadata\_EN中的关键词与AAT词表进行了匹配，此处为匹配后的结果（中文） | 后期添加字段。AAT词表使用的是AAT-Taiwan「藝術與建築索引典」。链接如下： https://aat.teldap.tw/index.php?\_lang=en |
| **藏品名称\_EN** | 藏品名称（英文） | 来自博物馆元数据。如果原始数据非英文，此处为DeepL机器翻译结果。 |
| **藏品名称\_ZN** | 藏品名称（中文） | 来自博物馆元数据。如果原始数据非中文，此处为DeepL机器翻译结果。 |
| **描述\_EN** | 藏品描述（英文） | 来自博物馆元数据。如果原始数据非英文，此处为DeepL机器翻译结果。 |
| **描述\_ZN** | 藏品描述（中文） | 来自博物馆元数据。如果原始数据非中文，此处为DeepL机器翻译结果。 |
| **时期** | 藏品被制作或被生产的历史时期（中文） | 后期添加字段，选择了藏品较多的几个特殊的历史时期进行了归一化处理和中文翻译。 |
| **起始年** | 藏品被制作或被生产的时期的起始年份 | 后期添加字段。从博物馆元数据中提取或根据“时代”信息进行公元纪年的映射处理。 |
| **终止年** | 藏品被制作或被生产的时期的终止年份 | 后期添加字段。从博物馆元数据中提取或根据“时代”信息进行公元纪年的映射处理。 |

1. **提示**
2. 关于字段：由于各个博物馆的字段格式不同，所有藏品的元数据字段名称都进行了人为的规整和合并，其中分类、材质、技法、Metadata、藏品名称、描述6个字段内的信息经过了中英双语的翻译（DeepL机翻），EN代表英文，ZN代表中文，分析时请注意检查翻译的准确性。
3. 关于数据与图片的对应关系：图片命名中的ID与比数据表单中的ID一一对应，Main\_ID为藏品在数据库中的唯一ID。如果想要只保留一个Main\_ID，可以把藏品图片按照Main\_ID重命名一下。
4. 所有样例数据均为从博物馆或其他官方API渠道获取，仅供此次大赛公益使用。请参赛选手使用时遵守数据提供博物馆的相关约定。

博物馆开放数据示例：

英国维多利亚与艾尔伯特博物馆(V&A)：<https://developers.vam.ac.uk/guide/v2/welcome.html>；

法国卢浮宫博物馆（Le Louvre）：<https://collections.louvre.fr/en/page/documentationJSO>；美国大都会艺术博物馆（The MET）：<https://www.metmuseum.org/articles/open-access-at-the-met>；

台北故宫博物院：<https://openapiweb.npm.gov.tw/APP_Prog/cht/overview_cht.aspx>；

澳大利亚国家博物馆（National Museum of Australia）：<https://www.nma.gov.au/about/our-collection/museum-api>。