

海洋地质大数据信息服务体系建设

戴勤奋,魏合龙,苏国辉,孙记红,刘京鹏,王 诏

(中国地质调查局青岛海洋地质研究所,青岛 266071)

摘要:针对海洋地质信息服务体系建设面临的大数据应用问题,提出了遵循互联网公共协议和 OData 数据开放协议,利用互联网 Web 平台及 Web 开发技术,构建基于 REST 风格 API 的轻量级 Web 服务,以及研究 Web 交互式数据可视化技术的建议。期望借鉴规范、技术、应用相辅相成的互联网成功模式,通过以 API 为核心的应用生态链,逐步推动海洋地质信息资源的开放、共享和互操作。

关键词:海洋地质;大数据;REST API;OData;交互可视化

中图分类号:TE319

文献标识码:A

DOI:10.16028/j.1009-2722.2017.11010

自 1989 年 Tim Berners-Lee 提出万维网设想^[1],20 余年来,互联网以前所未有的速度蓬勃发展,至今,互联网产业深入社会各行各业,并改变了人们的信息访问与共享方式。究其成功原因不外乎 3 点:①跨地域、跨国界互联;②遵循共同的互联网协议;③应用服务驱动发展。海洋地质信息服务体系建设可借鉴规范、技术、应用相辅相成的互联网成功模式,遵循互联网公共协议和数据开放协议、利用互联网 Web 平台及 Web 开发技术,构建安全、开放、平等、互利的数据共享与互操作网络应用环境;同时,通过专业化、个性化信息服务,逐步打开数据割据屏障,创建数据融合模式,开拓数据应用能力,实现海洋地质调查与研究资料从数字化、数据化到信息化的跨越。

1 大数据与轻量级 WEB 应用

信息化核心是数据,大数据核心是融合^[2]。随着不同比例尺海洋区域地质调查、海岸带环境

地质调查、以及海洋矿产资源调查工作不断开展,海量、异构、动态变化数据的存有量快速递增,为多源数据融合创造了条件,也提出了挑战。如何降低大数据的应用门槛,让数据快速成为可供客户消费的信息资源,是海洋地质信息服务体系建设面临的问题。构建轻量级的、基于 API 自由组合的 Web 应用服务体系是其中的一个解决方案。

API(Application Program Interface)是具明确定义、用于提供特定服务的网络应用程序接口。Web 是支撑 API 的技术架构,即 Web 四大基石:统一资源标识符(URI)、超文本传输协议(HTTP)、超文本(HyperText,除 HTML 外,还包括带有超链接的 JSON 或 XML 等)、以及多用途互联网邮件扩展类型(MIME,即媒体类型)。基于 API 的轻量级 Web 应用通过平台独立、低耦合、自包含的应用程序组件实现,与客户端之间采用开放协议通信,客户端只要使用浏览器即可与服务器端交互,消费服务器端提供的 Web 服务。

基于 API 组合的海洋地质大数据信息服务体系,意味着以决策层、管理层、知识层、运作层等不同层次客户的信息需求为导向,拟定不同粒度、不同方式的信息资源服务模式,在问题抽象基础上,明确定义独立的 API 接口,然后通过 API 组合提供信息资源服务,就此逐步构建以

收稿日期:2017-03-17

基金项目:中国地质调查局项目“数字海洋地质工程——海洋地质信息服务体系建设”(DD20160157)。

作者简介:戴勤奋(1963—),女,研究员,现从事海洋地质信息服务体系标准建设工作。E-mail:qddqinfen@163.com

API 为核心的应用生态链,以迭代方式逐渐开放可用的数据资源、提供实用的信息资源服务,并在不断服务与改进过程中寻找新的数据融合模式与服务技术,创造数据新的价值。

2 构建 REST 风格 API

REST(Representational State Transfer)中文译名表述性状态转移是一种基于网络的分布式应用架构风格,2000 年由 Roy Thomas Fielding 在他的博士论文中提出^[3]。REST 风格特征包括面向资源、可寻址性、连通性、无状态性、统一接口以及超文本驱动^[4],适合构建轻量级、可伸缩、松耦合的 Web 服务^[5]。

图 1 给出了 REST 风格 Web 服务的元模型。REST 风格 Web 服务使用 HTTP(S)作为底层协议,以 HTTP 方法(POST、GET、PUT/PATCH、DELETE)作为统一接口实现对资源的 CRUD(创建、读取、更新、删除)基本操作。其中的资源定义为能满足用户需求的可标识对象,REST API 中的资源采用 URI 唯一标识、经由 URL 寻址。API 接口则由资源的 URL(名词)及资源的可执行操作(动词)两部分定义。

HTTP 协议的无状态性以及资源的 URL 可寻址性有利于服务的云平台部署;其次,资源 URL 寻址使得服务与数据的物理存址无关,数据可完全由供方自主管理,不必提交相关管理机构或数据中心集中存储,由此避免数据物理集聚可能带来的机构协调与工作负荷问题,并有助于机

构之间的数据共享和互操作。

客户端通过 API 请求资源,服务器响应客户端请求返回的不是资源本身,而是资源表述。表述是资源对象按一定格式(例如 JSON、XML)序列化形成的文件,便于网络传输。表述中包含的内嵌链接既代表当前资源与其他资源的关联,也代表客户端的可选应用状态,由此驱动客户端选择感兴趣的链接,进入下一个、再下一个应用状态,即所谓的超文本驱动,也体现了资源之间的连通性。资源可按不同媒体类型(MIME 类型)多重表述,以便不同机构通过 HTTP 内容协商机制进行交互。表述的 Schema 模式是人机可读文档,可用于自动化测试、以及客户端输入数据的校验。

3 遵循开放数据协议

REST 是一种架构设计风格而不是标准规范,开放数据协议 OData 4.0 给出了 REST 风格数据服务(OData 服务)的标准化框架,遵循该框架构建的 Web 服务可用于获取并操作依据 OData 实体数据模型(EDM)定义的资源,OData 实体数据模型元素见图 2。

OData 由微软发起,OData 1.0、2.0、3.0 是微软的开放标准,OData 4.0 于 2014 年 3 月 17 日成为 OASIS(结构化信息标准促进组织)的开放工业标准^[6]。目前 OData 4.0 的最新版本为 2016 年 6 月 2 日的更新版,内容包括第 1 部分:协议;第 2 部分:URL 约定;第 3 部分:通用模式定义语言 CSDL 以及 JSON 格式、Atom 格式、数据汇总等扩展标准^[7-12]。2016 年 12 月 OData 4.0 的核心及 JSON 部分成为 ISO/ETC 国际标准^[13-14]。

一个 OData 服务的数据实体通过实体容器组合,实体容器是一个物理或虚拟的数据存储器,其内元素包括实体集、单例实体和导入操作。数据源通过 URL 寻址,来源不限,例如关系数据库、文件系统、传统 Web 站点等。实体之间关系通过导航属性定义,导航属性可在实体容器内按需绑定,绑定的导航属性可用于传递相关实体的链接关系。

OData 服务将服务文档作为服务入口资源,

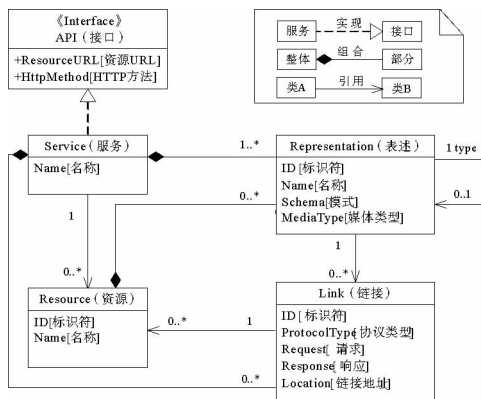


图 1 REST 风格 Web 服务元模型

Fig.1 Metamodel for restful web service

服务暴露的顶级资源(即服务根目录下顶级路径段资源)在实体容器中定义、在 OData 服务文档中罗列,客户可以将顶级资源作为访问起点,之后跟随返回表述中内嵌的超文本链接,实现对目标资源的交互操作,图 3 给出了超文本驱动的 OData 服务示意图。

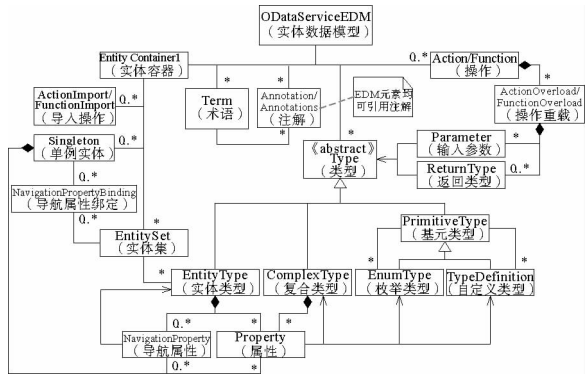


图 2 OData 实体数据模型

Fig. 2 OData EDM

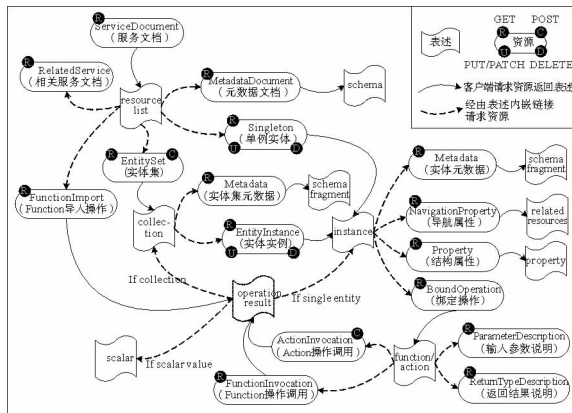


图 3 超文本驱动 OData 服务

Fig. 3 Hypertext-driven OData service

4 交互式数据可视化

数据可视化是将 JSON、XML 等格式的资源表述内容映射到 Web 页面的过程。近 5 年,因大数据兴起以及浏览器性能提高,基于 Web 2.0 的动态交互式可视化技术飞速发展,形形色色的前端框架以及 HTML5、JavaScript 开发库不段涌现,例如 jQuery^[15]、D3^[16]、webGL^[17] 等;同时,用

于数据可视化分析的后端开源库函数也不断涌现,例如 Python 的 matplotlib、Seaborn、ggplot、Bokeh、pygal、Plotly、geoplotlib 等^[18],使得无需插件、直接通过网页的动态响应式数据视图以及动态渲染二维、三维视图并与之交互成为可能。

海洋地质大数据信息体系建设,应将可视化作为大数据应用的一个研究分支,首先研究基于 Web 的解释型交互式数据可视化方法,即通过数据视图表达预先拟定的结果;在此基础上,进一步研究基于 Web 的探索型交互式数据可视化方法,即用多种视图展示同一数据集、揭示数据中可能隐藏的模式和趋势,由此推动多源异构大数据的融合、挖掘和智能化应用。同时切勿轻视交互的力量,因交互模式能激发客户的参与意识和探索数据的兴趣,进一步促进交互可视化技术的发展。

5 结论

大数据时代,信息量猛增,信息无时不在、无处不有,但落实到个体,其关心的信息是十分有限的,基于 REST API 的 Web 服务有利于从大数据中淘取“小数据”,也十分有利于服务的云平台部署。针对海洋地质信息服务体系建设,应在遵循互联网公共协议及数据开放协议基础上,利用互联网 Web 平台及 Web 开发技术,一方面构建目录服务平台,提供公共可访问的入口;另一方面开发服务于决策层、管理层、知识层、运作层等不同层次客户的、专而精的轻量级 Web 应用,释放可用的数据,传达实用的信息,并在服务过程中进一步寻找数据融合的契机,逐步推动海洋地质信息资源的开放、共享和互操作。

参考文献:

[1] Berners-Lee T. Information Management: A Proposal[R]. <http://info.cern.ch/Proposal.html>, March 1989.

[2] 李国杰. 对大数据的再认识[J]. 大数据, 2015(1):1-9.

[3] Roy Thomas Fielding. 李 锟,廖志刚,刘 丹,等译. 架构风格与基于网络的软件架构设计(Architectural Styles and the Design of Network-based Software Architectures)[D]. University of California, Irvine, 2000.

[4] Leonard Richardson, Sam Ruby. RESTful Web Services 中文版[M]. 徐 涵,李红军,胡 伟,译. 北京: 电子工业出版社, 2008.

[5] 李 锟. 深入探索 REST 专栏系列,理解本真的 REST 架

- 构风格[OL]. InfoQ, <http://www.infoq.com/cn/>, 2013-8-26.
- [6] OASIS. OASIS Approves OData 4.0 Standards for an Open, Programmable Web [N/OL]. OASIS, <https://www.oasis-open.org/news/>, 2014-3-17.
- [7] OASIS OData Version 4.0 Part 1: Protocol [S].
- [8] OASIS OData Version 4.0 Part 2: URL Conventions [S].
- [9] OASIS OData Version 4.0 Part 3: Common Schema Definition Language [S].
- [10] OASIS OData JSON Format Version 4.0 [S].
- [11] OASIS OData Atom Format Version 4.0 [S].
- [12] OASIS OData Extension for Data Aggregation Version 4.0 [S].
- [13] ISO/IEC 20802-1:2016—Open data protocol (OData) v4.0 — Part 1: Core [S].
- [14] ISO/IEC 20802-2:2016 — Open data protocol (OData) v4.0 — Part 2: JSON FORMAT [S].
- [15] Jonathan Chaffer, Karl Swedberg. Learning jQuery Fourth Edition Better interaction, design, and web development with simple JavaScript techniques [M]. Birmingham, UK: Packt Publishing Ltd, 2013.
- [16] Scott Murray. 数据可视化实战 使用 D3 设计交互式图表 (Interactive Data Visualization for the Web) [M]. 李松峰, 译. 北京: 人民邮电出版社, 2013.
- [17] Kouichi Matsuda, Rodger Lea. WebGL 编程指南 (WebGL Programming Guide: Interactive 3D Graphics Programming with WebGL) [M] 谢光磊, 译. 北京: 电子工业出版社, 2014.
- [18] Melissa Bierly. 10 Useful Python Data Visualization Libraries for Any Discipline [OL]. <https://blog.modeanalytics.com/python-data-visualization-libraries/>, 2016-8.

CONSTRUCTION OF INFORMATIZATION SERVICE SYSTEM FOR MARINE GEOLOGICAL BIG DATA

DAI Qinfen, WEI Helong, SU Guohui, SUN Jihong, LIU Jingpeng, WANG Zhao
(Qingdao Institute of Marine Geology, CGS, Qingdao 266071, China)

Abstract: In the construction of marine geological information service system, the big data utilization is a challenge. This paper puts forward four suggestions for consideration: using Internet Web platform and Web development technology, following the Internet protocols and OData data open protocol, building restful APIs based lightweight Web services, and researching web-based interactive data visualization methods. Learning from the Internet success model of specification, technology and application complementary strategies, the APIs core application ecosystem will be expected to gradually promote openness, sharing and interoperability for marine geological information resources.

Key words: marine geology; big data; REST API; OData; interactive data visualization