

流域水环境管理大数据存储与交换规范

(征求意见稿)

编制说明

标准编制组

2021年2月

项 目 名 称：国家流域水环境管理大数据平台关键技术研究

项目统一编 号：2017ZX07302004

承 担 单 位：北京师范大学

北控水务（中国）投资有限公司

山东省生态环境规划研究院

中科宇图科技股份有限公司

主 编 人 员：王国强、王运涛、冯艳霞、王宝刚、黄鹏飞、朱宜、谢刚、彭岩波。

目 录

一、任务来源	1
二、标准制定的必要性.....	3
三、主要工作过程	8
四、国内外相关标准研究	10
五、标准制定的原则和依据.....	14
六、标准主要内容和情况说明	16
七、标准实施的环境效益与经济技术分析.....	19
八、标准实施建议	21

一、任务来源

党中央、国务院高度重视水环境保护工作，“十一五”以来，积极探索流域水环境保护新路，取得积极成效。根据 2014 年《中国环境状况公报》，2001-2014 年，长江、黄河、珠江、松花江、淮河、海河、辽河七大流域和浙闽片河流、西北诸河、西南诸河总体水质明显好转，I-III类水质断面比例上升 32.7 个百分点，劣V类水质断面比例下降 21.2 个百分点。尽管我国水污染防治工作取得了积极进展，但水资源保障能力弱、水环境质量差、水生态受损重、环境隐患多等问题依然十分突出。严峻的水环境问题需要利用大数据等技术手段进行科学评价、预测预警和优化决策方案，辅助流域水环境管理。

(1) 推进流域环境治理体系和治理能力现代化的重要手段

大数据技术不仅能解决传统数据管理方面存在的问题，而且使数据管理趋于智能化，加强政府公共服务和市场监管，推动简政放权和政府职能转变。国务院《促进大数据发展行动纲要》等文件要求推动政府信息系统和公共数据互联互通，促进大数据在各行业创新应用。

《国务院关于积极推进“互联网+”行动的指导意见》明确提出构建“互联网+”绿色生态，实现生态环境数据互联互通和开放共享。

2016 年 3 月，生态环境部正式发布了《生态环境大数据建设总体方案》，陈吉宁部长对生态环境大数据平台建设提出了明确要求，指出“大数据、互联网+等信息技术已成为推进环境治理体系和治理能力现代化的重要手段，要加强生态环境大数据综合应用和集成分析，为生态环境保护科学决策提供有力支撑”。

(2) 提升流域水环境保护科学决策支撑水平的有效途径

流域水环境管理与单纯的水环境管理具有更为不同的内涵与外延。流域水环境管理不仅包括水污染防治，还包括水土保持、河道治理、生态保护等内容，相关管理涉及环境保护部、发展改革委、工业和信息化部、国土资源部、住房城乡建设部、交通运输部、水利部、农业部、卫生计生委等多部门。当前流域水环境管理相关业务应用的数据类型高达几十种，来源于不同的数据生产部门，其组织管理的方式、标准、参考体系也各不相同，给环境大数据的快速形成与综合应用提出了挑战。

流域水环境管理是一种典型的跨行业、多类型的大数据综合业务应用，对多源、复杂的流域水环境大数据信息进行综合应用和集成分析。因此，针对流域水环境管理的大数据平台

建设是提升流域水环境保护科学决策支撑水平的重要保障。

(3) 实现流域水环境精细化管理的必然要求

2015年4月，国务院印发《水污染防治行动计划》（以下简称“水十条”），提出“完善流域协作机制，流域上下游各级政府、各部门之间要加强协调配合、定期会商，实施联合监测、联合执法、应急联动、信息共享”、“深化重点流域污染防治，研究建立流域水生态环境功能分区管理体系”、“完成全国排污许可证管理信息平台建设”等要求。2016年生态环境部成立水环境管理司，体现了国家以水环境问题为导向，统筹开展水环境精细化管理的决心，理顺了内部职责和业务关系，提高了工作效率。

通过“国家流域水环境管理大数据平台关键技术研究”的实施，重点攻克水环境多模集合模拟和流域水环境管理大数据平台构建等关键技术，构建流域水环境管理大数据平台，为流域水环境管理提供科学决策的准确依据，实现流域水环境管理向精细化转变的技术支撑，对推动流域水环境管理创新，提升管理水平，做好新常态下的流域水环境保护工作具有重大意义。

(4) 流域水环境管理大数据标准化需求

为贯彻落实《国务院关于印发促进大数据发展行动纲要的通知》精神，积极开展生态环境大数据建设与应用工作，生态环保部组织编制了《生态环境大数据建设总体方案》。“国家流域水环境管理大数据平台关键技术研究”课题组通过旨在解决大数据平台架构、精细化大数据分析方法及智能化业务分析手段等方面的难点，构建流域水环境管理业务化平台，实现流域水环境智能化、精细化管理、评估与决策。

构建流域水环境管理大数据平台对数据资源汇集整合的广度、信息综合开发利用的深度，以及业务应用和管理决策的智能化水平方面均提出了更高要求，面临视频、图片、语音等非结构化数据的大量涌入，通过标准化的途径整合资源，固化成果，促进各方达成共识，形成统一的数据格式、接口、安全、开放等规范，是流域水环境管理大数据平台构建成功的关键。

综上所述，“国家流域水环境管理大数据平台关键技术研究”课题在流域水环境管理大数据领域历时四年的技术攻关，攻克了大数据标准化、数据汇集、多模集合、数据挖掘、智慧管理和决策支持等一系列难题，在实践中获得了一系列专业技术成果。《流域水环境管理大数据存储与交换规范》是流域水环境管理大数据平台的核心技术规范，它是平台大数据挖掘、多模集合模拟、智慧管理和业务应用间联动交互往来的核心枢纽。课题组结合国内外相

关技术标准，构建了大数据存储和交换的技术体系，经过近两年的应用示范，平台数据存储稳定、实时数据读取快速、功能模块数据传输畅通，实践证明该大数据存储与交换技术科学、全面、可行、适用，由此编制了《流域水环境管理大数据存储与交换规范》以供参考。

标准编制工作组主要由国家流域水环境管理大数据平台关键技术研究课题组成员参加，包括北京师范大学、北控水务（中国）投资有限公司、山东省生态环境规划研究院、中科宇图科技股份有限公司。标准编制工作主要是由工作组内的成员参与完成，北京师范大学、北控水务（中国）投资有限公司、山东省生态环境规划研究院主要负责规范审核、立项申请及发布；中科宇图科技股份有限公司主要负责编制。

二、标准制定的必要性

大数据是以容量大、类型多、存取速度快、应用价值高为主要特征的数据集合，正快速发展为对数量巨大、来源分散、格式多样的数据进行采集、存储和关联分析，从中发现新知识、创造新价值、提升新能力的新一代信息技术和服务业态。全面推进大数据发展和应用，加快建设数据强国，已经成为我国国家战略。

（一）政策要求落实的必要性

党中央、国务院高度重视大数据在推进生态文明建设中的地位和作用。习近平总书记明确指出，要推进全国生态环境监测数据联网共享，开展生态环境大数据分析。李克强总理强调，要在环保等重点领域引入大数据监管，主动查究违法违规行为。国务院《促进大数据发展行动纲要》等文件要求推动政府信息系统和公共数据互联共享，促进大数据在各行业创新应用；运用现代信息技术加强政府公共服务和市场监管，推动简政放权和政府职能转变；构建“互联网+”绿色生态，实现生态环境数据互联互通和开放共享。陈吉宁部长要求，大数据、“互联网+”等信息技术已成为推进环境治理体系和治理能力现代化的重要手段，要加强生态环境大数据综合应用和集成分析，为生态环境保护科学决策提供有力支撑。

（二）流域管理提效的必要性

流域水环境管理是一种典型的跨行业、多类型的大数据综合业务应用，对多源、复杂的流域水环境大数据信息进行综合应用和集成分析。针对流域水环境管理的大数据平台建设是提升流域水环境保护科学决策支撑水平的重要保障。

流域水环境管理与单纯的水环境管理具有更为不同的内涵与外延。当前流域水环境管理相关业务应用的数据类型高达几十种，来源于不同的数据生产部门，其组织管理的方式、标准、参考体系也各不相同，给环境大数据的快速形成与综合应用提出了挑战。

近年来，各部门不断更新采用现代化管理手段，主要有地理信息系统、全球定位系统、遥感技术和计算机的普及应用等，不仅提高了管理水平，也积累了海量的流域管理基础数据。虽然很多数据已经集中到了“数据中心”，但数据难以真正流动起来，环境信息化存在体制机制不顺，基础设施和系统建设分散，应用“烟囱”和数据“孤岛”林立，业务协同和信息资源开发利用水平低，综合支撑和公众服务能力弱等突出问题，难以适应和满足新时期生态环境保护工作需求。

2015年4月，“水十条”提出“完善流域协作机制，流域上下游各级政府、各部门之间要加强协调配合、定期会商，实施联合监测、联合执法、应急联动、信息共享”、“深化重点流域污染防治，研究建立流域水生态环境功能分区管理体系”、“完成全国排污许可证管理信息平台建设”等要求。2016年生态环境部成立水环境管理司，体现了国家以水环境问题为导向，统筹开展水环境精细化管理的决心，理顺了内部职责和业务关系，提高了工作效率。

然而，与国外发达国家相比，目前我国流域水环境管理缺乏全国本底数据支撑，开展顶层设计及宏观政策制定缺乏强有力的数据支撑；亟需开发智能化的水环境管理大数据库和信息平台来提升流域水环境精细化管理的综合水平，促进数字环保向智慧环保的转变，为我国流域水环境精细化管理提供技术保障。

（三）标准规范完善的必要性

大数据标准体系由七个类别的标准组成，分别为：基础标准、数据标准、技术标准、平台和工具标准、管理标准、安全和隐私标准、行业应用标准。

（1）基础标准

为整个标准体系提供包括总则、术语、参考模型等基础性标准。

（2）数据标准

该类标准主要针对底层数据相关要素进行规范。包括数据资源和数据交换共享两部分，其中数据资源包括元数据、数据元素、数据字典和数据目录等，数据交换共享包括数据交易和数据开放共享相关标准。

（3）技术标准

该类标准主要针对大数据相关技术进行规范。包括大数据集描述及评估、大数据处理生命周期技术、大数据开放与互操作、面向领域的大数据技术四类标准。其中，大数据集描述及评估标准主要针对多样化、差异化、异构异质的不同类型数据建立标准的度量方法，以衡量数据质量，同时研究标准化的方法对多模态的数据进行归一化处理，并根据我国国情，制定相应的开放数据标准，以促进政府数据资源的建设。大数据处理生命周期技术标准主要针对大数据产生到其使用终止这一过程的关键技术进行标准制定，包括数据产生、数据获取、数据存储、数据分析、数据展现、数据安全与隐私管理等阶段的标准制定。大数据开放与互操作标准主要针对不同功能层次功能系统之间的互联与互操作机制、不同技术架构系统之间的互操作机制、同质系统之间的互操作机制的标准化进行研制。面向领域的大数据技术标准主要针对电力行业、医疗行业、电子政务等领域或行业的共性且专用的大数据技术标准进行研制。

（4）平台和工具标准

该类标准主要针对大数据相关平台和工具进行规范，包括系统级产品和工具级产品两类，其中系统级产品包括(实时计算产品(流处理)，数据仓库产品(OLTP)，数据集市产品(OLAP)，数据挖掘产品，全文检索产品，非结构化数据存储检索产品，图计算和图检索产品等)；工具级产品包括平台基础设施、预处理类产品、存储类产品、分布式计算工具、数据库产品、应用分析智能工具、平台管理工具类产品的技术、功能、接口等进行规范。相应的测试规范针对相关产品和平台给出测试方法和要求。

（5）管理标准

管理标准作为数据标准的支撑体系，贯穿于数据生命周期的各个阶段。该部分主要是数据管理、运维管理和评估三个层次进行规范。其中数据管理标准主要包括数据管理能力模型、数据资产管理以及大数据生命周期中处理过程的管理规范；运维管理主要包含大数据系统管理及相关产品等方面的运维及服务等方面的标准；评估标准包括设计大数据解决方案评估、数据管理能力成熟度评估等。

（6）安全和隐私标准

数据安全和隐私保护作为数据标准体系的重要部分，贯穿于整个数据生命周期的各个阶段。大数据应用场景下，大数据的4V特性导致大数据安全标准除了关注传统的数据安全和

系统安全外，还应在基础软件安全、交易服务安全、数据分类分级、安全风险控制、电子货币安全、个人信息安全、安全能力成熟度等方向进行规范。

(7) 行业应用标准

行业应用类标准主要是针对大数据为各个行业所能提供的服务角度出发制定的规范。该类标准指的是各领域根据其领域特性产生的专用数据标准，包括工业、政务、服务等领域。

根据大数据标准体系框架，整理分析已发布、已报批、已立项、已申报、在研以及拟研制的大数据相关国家标准 104 项。

通过对现有大数据国家标准进行分析可以看出：

(1) 在数据资源方面，我国已具备一定的标准基础，相关国家标准在大数据领域下同一样适用。下一步需要根据大数据技术、产业现状适时修订已有国家标准，保证标准紧跟技术、产业发展；同时推进相关数据资源标准的推广与应用，保证标准的落地实施。

(2) 在交换共享方面，依托全国信标委大数据标准工作组，已经开展了相关标准的研制工作，发布数据交易国家标准 2 项，在研开放共享国家标准 3 项。下一步需要推进相关数据开放共享国家标准的报批工作，同时围绕国家对于政府数据开放共享的任务要求，加快适用于政府数据开放共享的国家标准研制。

(3) 在数据管理方面，《数据管理能力成熟度评估模型》(GB/T 36073-2018) 作为我国首个数据管理领域的国家标准已经发布。下一步急需从数据管理能力评估方法的角度开展标准化研究，落实标准在产业中的应用，推动整个行业数据管理能力的提升。

(4) 在平台和工具标准方面，目前已立项《信息技术大数据系统通用规范》(计划号：20171082-T-469) 等 5 项国家标准，标准范围覆盖大数据通用系统、大数据存储与处理系统、大数据分析系统。下一步需要围绕大数据系统功能模块，完善数据收集、数据访问等功能模块的相关标准研制，同时围绕国家标准，开展大数据系统产品的标准符合性测试评估工作。

(5) 在行业大数据领域，主要开展了参考架构等基础标准的制定。

总体来说，目前我国在大数据领域在基础术语、数据资源、交换共享、数据管理、大数据系统产品、工业大数据等方面已开展了国家标准研制工作。但在环境保护流域管理分支中

并无专项的国家标准的发布和研究，因此，编写制定流域水环境大数据存储与交换的相关规范势在必行。

（四）效益促进作用的必要性

由于我国的流域水环境管理制度采取分职能设计、分部门管理、分地区实施，同时各体系大多实行封闭监管模式，彼此之间管理信息不能有效共享，数据、凭证传递不及时，各方数据常有出入，造成流域水环境管理协调难度加大，信息共享不足导致资金使用效率不高。面对这些问题，流域水环境管理大数据标准体系的深入应用在信息采集、信息资源利用、跨地域管理等方面有效地提高了资金使用效率：

（1）流域水环境大数据提高信息采集效率。通过流域水环境大数据标准的建设，原本存在于不同部门的信息比对将成为可能，对数据的加密脱敏技术，将解决出于信息保密原因造成的信息共享困难，国家和各省市各级水环境管理大数据中心将有序归集数据资源，解决由于信息系统建设权属和数据存储方式造成的共享难题，从根本上解决信息的有效采集。

（2）流域水环境大数据解决资源利用难题。利用多源异构数据处理技术，不同部门采集的数据不一致，数据源存储格式不一的水环境信息资源可以得到共享。同时，来自各部门系统的水环境信息可以得到良好的比对和识别，分别储存和统一使用。数据的比对和分析，可以彻底解决因为信息分散而导致的行政管理问题，确实提高资金的使用效率。

（3）流域水环境大数据解决跨地域问题。流域水环境大数据的应用将有助于利用信息化手段消除因南北东西各个流域行政管理差异造成的数据问题，同时有助于建立统一的水体信息画像，彻底解决跨地域信息不统一的问题。

为落实党中央、国务院决策部署和部党组要求，充分运用大数据、云计算等现代信息技术手段，全面提高生态环境保护综合决策、监管治理和公共服务水平，加快转变环境管理方式和工作方式，制定本标准规范。

本规范通过分析国家流域各职能部门现有数据结构形式、数据通讯方式及信息管理模式，创新式的建立信息与数据共享机制，采用统一的互访接口，将多源数据资料进行汇总组织，形成一套完整的流域数据库，实现信息资源的整合和有限共享，提高综合管理决策的科学性和高效性。建立健全国家流域信息共享的保障机制、政策保障机制、资金保障机制、安全保障机制等。在流域综合管理组织框架中明确流域信息共享和信息管理的责任部门和规章制度，

明确流域信息的综合采集、存储、传输与交换、管理、应用等各项流程。

综上，目前我国已有的大数据标准在数据交换共享、可视化、系统集成等方面尚无法满足需求，为填补行业应用空白，亟待研究建立流域水生态环境数据的收集、传输、整合、交换和集成通用标准，以保障平台数据来源的可靠性，系统与模块一体化集成的可行性以及可扩展性，为平台开发提供有力的技术支撑。

三、主要工作过程

规范编制组主要开展了以下调查和研究工作：

1、成立编制组及工作分工

2017 年，水专项课题申报成功后，课题承担单位北京师范大学，联合协作单位中科宇图科技股份有限公司、北控水务（中国）投资有限公司、山东省生态环境规划研究院等单位在第一时间组成标准编制组和工作团队，并明确分工。其中，北京师范大学、北控水务（中国）投资有限公司、山东省生态环境规划研究院主要负责规范审核和立项申请；中科宇图科技股份有限公司主要负责规范的编制，对规范条款的体系性、科学性、严谨性和可行性进行分析研究，解释和修订征求意见单位的相关问题。

2、流域水环境大数据存储或交换规范的调查研究

通过资料调研和专家咨询的方式，对有关大数据技术的国家标准和行业标准进行了调查研究。大数据存储标准有国家 2019 年发布的《信息技术 大数据存储与处理系统功能要求》（GB/T 37722），2019 年发布的《信息技术 云计算 分布式块存储系统总体技术要求》（GB/T 37737），2019 年发布的《信息安全技术 信息系统安全等级保护基本要求》（GB/T 22239），交换标准有国家 2014 年发布的《环境空间数据交换技术规范》（HJ 726），2014 年发布的《环境信息交换技术规范》（HJ 727）。

《信息技术大数据存储与处理系统功能要求》规定了大数据存储与处理系统的功能要求适用于指导大数据存储与处理系统的设计、开发和应用部署。《信息技术 云计算 分布式块存储系统总体技术要求》规定了分布式存储系统资源管理功能要求、系统管理功能要求、可扩展要求、兼容性要求和安全性要求。《信息安全技术 信息系统安全等级保护基本要求》规定了对信息系统安全等级保护状况进行安全测试评估的要求。《环境空间数据交换技术规范》规定了环境空间数据交换信息 XML 描述、交换模型、交换基础、交换流程、数据交换质量要求和安全要求等内容。

综上，现有规范从信息学和环境领域对数据的存储和交换进行规范，并未制定流域水环

境管理领域的数据存储和交换，由此，结合上述标准在数据存储、交换、安全等方面的要求，在流域水环境大数据管理规范上为本文件编制提供研究依据。

3、确定规范编制思路

2019年1月至3月，标准编制组研究确定了流域水环境大数据存储与交换规范编制思路、主要要点，起草了标准总体框架和主要思路，明确了任务分工、工作计划，开展了规范制定专题研究。

4、明确规范编制要求

2019年5月21日，以北京师范大学为课题主持单位，整合生态环境部信息中心、中国环境科学研究院、中科宇图科技股份有限公司、北京澜途集思科技有限公司、中节能咨询有限公司、山东省生态环境规划研究院和北控水务（中国）投资有限公司等单位，形成“政、产、学、研、用”一体化的科研团队，在大数据平台构建、水环境多模集合模拟、业务应用、软件研发以及应用示范方面开展了大量研究，积累了丰富的成果，具有良好的工作基础。

标准编制包括立项、起草、审核、批准和发布等环节。立项申请由申报单位提出标准立项申请；立项审查主要依托中国环境科学学会（以下简称“学会”）分技术领域“标准专家组”进行立项审查；标准起草由申报单位成立标准起草工作组，开展标准编制；初稿征求意见由学会秘书处向会员单位及相关单位征求意见；送审稿审核则在意见处理完成后形成，由学会组织标准会议审查；最终审核通过报批公示，准备发布。

5、规范初稿起草

在前期调研和研讨的基础上，2019年7月，标准编制组起草完成了《流域水环境管理大数据存储与交换规范》初稿。

2019年7月22日、2019年8月29日、2019年9月23日，标准编制组三次召开专家咨询与研讨会，邀请中国环境科学研究院、生态环境部信息中心、中国环境监测总站、北京师范大学等单位专家与会并提出意见，修改完善了《流域水环境管理大数据存储与交换规范》初稿。

6、规范初审会

2019年9月在北京召开“流域水环境大数据信息挖掘平台构建”技术指南标准簇自审会，对包括“存储与交换规范”在内的五项技术指南进行自审。专家组听取了主编单位针对“流域水环境大数据信息挖掘平台构建”技术指南标准簇中五项指南规范的编制说明和草案内容，经询问，讨论，形成以下意见：一、编制单位调查研究充分，提交的五项技术指南说

明和指南材料齐全，结构合理，内容详实，完整。二、标准簇定位准确，编制原则科学合理，具有可操作性，该标准簇明确了流域水环境管理大数据存储与交换规范的数据分类、存储要求及交换要求等内容，规范了流域水环境大数据平台的存储与交换的开发与使用；基于业务管理需求，理清数据内容、数据种类与流域水环境的关系，为平台开发及数据交换共享奠定基础；对在建或要建的信息化系统产生约束，为系统建设及验收依据和准则；流域水环境管理大数据采集、存储、共享、利用、更新和维护等环节的工作内容，包括基本原则、主体责任、数据资源使用权限等，充分发挥标准规范在支撑业务管理等方面的作用；规范了各类工作中的数据交换、数据传输、数据存储标准、文件存放目录规范等，促进各相关系统间的互联互通。适用于对我国境内流域水环境管理大数据业务的存储与交换进行规范。同时，专家组提出意见如下：（1）进一步明确提供哪些智慧化多功能的信息服务。（2）进一步实现流域水生态环境数据收集传输、整合、交换和集成通用目标。（3）建议规范按照生态环境部规范绿色通道项目进行管理。

2020年8月召开标准立项会，中国环境科学学会组织专家通过视频会议对《流域水环境管理大数据存储与交换规范》（以下简称为“存储与交换规范”）进行了立项评审，会议决定“存储与交换规范”根据专家意见修改后立项。

7、规范专家咨询会

2020年9月28日，编制组组织召开流域水环境大数据信息挖掘平台构建关键技术专家研讨会，根据专家意见对《流域水环境管理大数据存储与交换规范》征求意见稿进行了修改完善。

四、国内外相关标准研究

大数据存储与交换技术在各行各业都有着很广泛应用，如在天文观测，科学计算，生物数据处理相关，气象观测等领域。

在美国环保局内部业务应用阶段早已经实现大数据存储与交换技术，美国环保局通过建立环境数据传输与交换系统、环境数据仓库，实现了更加复杂的与各联邦政府间数据共享、业务协同的高级应用。美国在开始进行信息化建设时，为了满足环保工作各项实际业务的需求，美国环保局开发建设了数据库和信息系统，然而系统之间存在互不兼容，难于共享的问题。美国环保局没有依靠分别建立传输网络和系统，而是依靠统一的中央数据交换系统，实现了对不同业务系统的数据传输和交换，提供单一类型数据的申报和信息交换的窗口，为美

国环保局各级部门，州级环保部门、部落或者其他环保社团提供服务。企业通过此系统快速传输和交换环境数据。中央数据交换系统已经广泛应用，加拿大、墨西哥等都加入，实现了跨境的数据信息交换，成为了美国环保局数据信息采集、传输和共享交换最重要的基础设施。

近年来，我国在大数据存储与交换技术领域也取得了长足发展。中国科学院刘应波针对在使用红外望远镜和大型太阳望远镜过程中海量观测数据需要存储的应用场景，分析了目前行业中常见的分布式文件系统，在此基础上设计了一种面向天文观测的海量分布式存储系统。其主要思想为分布式、高性能、并行、分片等。通过数据多点并行读写的方式提高了数据读写能力。东华大学过汇卿提出了一种适用于射电天文处理上层应用的分布式数据存储策略 DDSS(Distributed Data Storage Strategy)，设计了分布式数据存储策略的系统框架。并且设计了混合分片列式存储方法，在保留列存储查询优势的同时提升了数据导入的速度。然后，通过维护基于相对位置映射的元数据来快速读取包含大量数据的天文阵列数据，显著地提升了天文处理应用底层数据读写的吞吐量。

四川联合大学展之宏通过对单驱动器磁盘系统和多驱动器配置构成的磁盘阵列系统存储性能对比研究发现，在解决动态大容量影像数据实时存储时，使用多驱动器配置的磁盘阵列系统能有效解决单磁盘驱动使用中效率低下的问题。为了降低读写文件过程中的 CPU 占用问题，应尽可能使用具有异步 I/O 功能的 SCSI 磁盘，同时应采用多个 SCSI 通道以提升 I/O 性能。北京航天飞行控制中心李彤华针对某飞控中心数据存储系统的实际需求，设计并实现了一套基于 SAN 存储架构的高可用实时可扩展集群存储方案，通过软硬件结合的方式，实现了冗余数据的路径、故障切换、实时存储和查询。华中科技大学的刘超针对传统的关系数据存储系统性能不足、容错性差，无法适应海量非结构化数据管理的问题，提出一种高性能、高可用非关系型存储管理机制。设计了良好的用户访问服务接口，通过高效的一致性哈希算法支持数据分发到多个存储节点；采用可配置的数据副本机制改善存储系统的可用性；提出查询故障处理机制，用以提升存储系统的容错性，避免节点失效导致服务中断问题。通过实验，证明了新的存储系统的并发访问请求能力和传统的文件系统、关系型数据库相比性能有很大提升。

在卫星遥感领域，我国构建了自主的海量卫星遥感数据实时存储与交换技术架构。硬件架构体分为消息分发服务器、实时交换服务器、应用处理服务器、内存数据库、关系型数据库、共享文件系统以及前端显示工作站。海量卫星遥感数据实时存储与交换过程中存在着大量的数据传输，对数据传输的速率、安全、稳定、低延迟等都有很高要求，所以使用万兆网

络作为数据传输的通道。数据实时存储主要流程为：实时交换（写）服务器接收外部数据进行调制解调，通过数据切分将数据分块。通过万兆网络将非结构化数据存入共享文件系统，将结构化数据存入关系型数据库。数据实时交换包括两个层次：对于需要持久化存储与高速交换的数据，设计了一种基于共享文件系统和文件编目消息推送的海量卫星遥感数据实时存储与交换机制。

同时，国内外相关组织机构在大数据标准化方面已经开展了很多工作，但整体尚处于研制阶段。国外方面，自 2014 年，ISO/IEC JTC1（第一联合技术委员会）、ITU（国际电信联盟）、NIST（美国国家标准和技术研究院）等国外标准化组织先后成立大数据标准化研究工作组，研制大数据定义、相关术语、需求等方面的标准。目前整体工作进展为：数据管理与交换方面，ISO/IEC JTC1 SC2 数据管理和交换分技术委员会，致力于数据域定义、数据类型和数据结构以及相关语义等标准，用于持久存储、并发访问、并发更新和交换数据的语音、服务和协议等标准以及构造、组织和注册元数据及共享和互操作相关的其他信息资源的方法、语言服务和协议等标准的研制，目前正在编制国际标准《SQL 对多组数组的支持》、《数据集注册元模型》和《数据源注册元模型》并开展 SQL 对多态表功能的支持、SQL 对 JSON 的支持和 SQL 对多维数组的支持等课题研究。

在基础性标准方面，ISO/IEC JTC1 开展了参考架构和术语、大数据标准化需求等研究，目前正在研制 information technology - big data - overview and vocabulary, information technology - big data reference architecture. NIST 在形成产业界、学术界和政府等不同领域共识性的定义、术语、安全参考体系架构和技术路线图等方面进行研究，目前完成了大数据定义、大数据分类、大数据参考架构调研白皮书、大数据参考架构和大数据技术路线图等 V1.0 版本，正在进行 V2.0 版本的工作。

在大数据需求标准化方面，ITU 开展了针对大数据的物联网具体需求和能力要求、基于云计算的大数据需求和能力、大数据交换需求和框架、大数据即业务的功能架构等标准课题研究工作，目前基于云计算的大数据需求和能力标准已发布，对大数据的定义、大数据特性、大数据功能、大数据与云计算的关系、从电信角度看基于云计算的大数据好处、大数据需求、能力要求、用户案例及应用场景等内容进行了描述。其他标准正在研制过程中。此外，NIST 在数据分析技术互操作性、可移植性、可用性和扩展性需求和安全基础设施需求方面开展了大量研究工作，目前完成了大数据用例和需求、大数据安全和隐私需求等成果的 V1.0 版本。

表 1 标准研制情况

序号	标准号	标准名称	状态
1	GB/T 35295- -2017	信息技术大数据术语	发布
2	GB/T 35589-2017	信息技术大数据技术参考模型	发布
3	GB/T 34952-2017	多媒体数据语义描述要求	发布
4	GB/T 34945- -2017	信息技术数据溯源描述模型	发布
5	GB/T 35294-2017	信息技术科学数据引用	发布
6	GB/T 36073-2018	数据管理能力成熟度评估模型	发布
7	20141200-T-469	信息技术数据交易平台交易数据描述	报批
8	20141201-T-469	信息技术数据交易平台通用功能要求	通过评审
9	20141203-T-469	信息技术数据质量评价指标	报批
10	20141204-T-469	信息技术通用数据导入接口规范	报批
11	20160597-T- 469	信息技术大数据分析系统基本功能要求.	征求意见
12	20160598-T- 469	信息技术大数据存储与处理平台技术要求	草案
13	20171083-T- 469	信息技术大数据基于参考架构下的接口框架	草案框架
14	20171082-T-469	信息技术大数据分类指南	草案框架
15	20171082-T-469	信息技术大数据系统通用规范	草案
16	20171081-T-469	信息技术大数据存储与处理系统功能测试规范 范	草案框架
17	20171065-T-469	信息技术大数据分析系统功能测试规范	草案框架
18	20171066-T-469	信息技术大数据面向应用的基础计算平台基 本性能要求	草案框架
19	20171067-T-469	信息技术大数据开放共享第 1 部分：总则	草案
20	20171068-T-469	信息技术大数据开放共享第 2 部分：政府数据 开放共享基本要求	草案
21	20171069-T-469	信息技术大数据开放共享第 3 部分：开放程度 评价	草案

22	20173818-T-469	信息技术大数据系统运维和管理功能要求	草案框架
23	20173819-T-469	信息技术大数据工业应用参考架构	草案框架
24	20173820-T-469	信息技术大数据产品要素基本要求	草案框架

国内大数据标准化方面也尚处于起步阶段。在基础性标准方面，正在研制大数据术语、技术参考模型、参考架构等；在数据资源方面，我国已经研制了一些相关数据标准，适用于大数据，目前正处于推广应用与完善阶段。在交换共享方面，目前研制了《信息技术 数据交易服务平台 交易数据描述》、《信息技术 数据交易服务平台 通用功能要求》、《环境空间数据交换技术规范》、《环境信息交换技术规范》4项交易类的国家标准，尚缺乏交易流程、交易数据管理方面的标准。技术标准方面，包括数据访问、数据质量、大数据安全三类，其中，数据访问类集中在数据导入和数据库相关标准，但尚缺乏分析、可视化类标准；数据质量类的多项标准正在研究过程中；大数据安全类中大数据的安全框架、隐私和访问控制等标准正在研制。大数据平台和工具标准方面，目前较多数据库、非结构化数据管理产品类标准正在研制，缺乏大数据系统级相关产品的标准。在生态环境行业应用领域，也存在空白，大数据标准在数据存储、交换共享、可视化、系统集成等方面尚无法满足需求。

在此基础上，现有存储与交换标准的规定大体可以分为两类：一类国家通用性标准，只规定了相关要求，未细化具体内容；另一类是行业标准涵盖不够全面。第一类标准与本标准之间不存在交叉，在系统实现时，两个标准都执行即可，国标也是本标准的基本架构。因此不存在标准兼容性的问题。例如在执行《信息技术 大数据存储与处理系统功能要求》（GB/T 37722）中关于数据存储要求时，同时可以执行本标准存储的相关规定。再如《环境空间数据交换技术规范》（HJ 726-2014），明确了交换的空间数据的格式、交换模型、安全认证，但不影响本规范关于流域水环境空间数据的存储和交换，二者可同时执行。第二类是在行业标准中，存在行业关联性弱和内容不够明确等问题。现行的标准规范中，暂无对流域水环境结合数据存储和交换的规则，本规范结合流域水环境的业务分类及数据特征，细化了存储规则和交换规则，对流域水环境系统的交流沟通起到了规范作用。

据此，本标准未采用国际标准内容，主要参考国内相关标准，在国内属首次制定。

五、标准制定的原则和依据

（一）标准制定原则

本标准内容以全国信标委大数据标准工作组、JTC1 WG9 大数据工作组和 NIST 的研究成果为基础，以目前业界主要大数据企业的主流大数据产品为依据进行研制。参考《信息技术大数据参考架构》等大数据在研标准，研究制定相应的基本功能要求，保障大数据平台存储与交换的应用能力。

为使标准能够适应我国的实际需求，对我国流域水环境管理大数据存储与交换的研制和使用具有指导性和适用性，在标准制定过程中，主要遵循了如下几个原则：

1. 实用性

实用性原则是标准研究中最重要的基本原则，存储与交换标准编写规范是否实用，关系到各类业务系统中数据交换共享标准的制定，进而影响各类数据资源的整合利用，因此标准在编制时认真分析研究了现实状况，充分在标准中体现实用性。

2. 开放性

本标准作为一个推荐性标准，为了尽可能地覆盖现有的数据现状，在编制标准时，积极引入先进的模式和前沿技术，同时充分考虑了目前的需求和技术水平，使标准能够根据技术和行业发展，以及市场需求的变化不断进行扩充和完善。

3. 通用性

标准制定的目的是规范各类数据存储与交换标准的制定，因此非常注重通用性原则，主要对各类数据的格式、质量、要求、存储、交换等通用指标进行规定。

（二）标准主要技术内容的依据

本规范编制中主要参考以下相关标准：

GB/T 18793 信息技术可扩展置标语言（XML） 1.0

GB/T 37722 信息技术 大数据存储与处理系统功能要求

GB/T 37737 信息技术 云计算 分布式块存储系统总体技术要求

GB/T 22239-2008 信息安全技术 信息系统安全等级保护基本要求

HJ/T 416 环境信息术语

HJ 726-2014 环境空间数据交换技术规范

HJ 727-2014 环境信息交换技术规范

六、标准主要内容和情况说明

(一) 标准主要内容

本规范规定了流域水环境信息系统集成过程中所涉及到的相关技术要求,包括数据信息存储、数据信息使用、数据信息交换和系统集成中的交换规则,对于跨地区、跨部门、跨机构的水生态环境信息数据存储和交换问题的解决具有重要的意义。

本规范根据流域水环境业务系统的所涵盖的数据内容,从数据分类、数据特性、数据组织、数据存储、数据交换等方面进行的数据存储与交换的分类与说明。主要包括四部分内容:

第一部分:流域水环境管理大数据存储与交换规范文档的范围、规范性引用文件、术语和定义、缩略语。此部分内容主要说明该规范所依据的文件、参考的标准以及流域水环境和大数据的相关术语、缩略语,为使用者理解后文内容提供说明性指导。

第二部分:流域水环境管理大数据存储与交换规范的总体框架,此部分介绍了流域水环境管理数据的业务分类和类型分类。从数据形式上明确了业务数据的分类标准,从水环境业务组织性上,分别对水环境空间数据和非空间数据进行了细分,然后将每种组织结构对应的数据类型做了概化的存储要求说明。流域水环境大数据组成结构如下图所示:

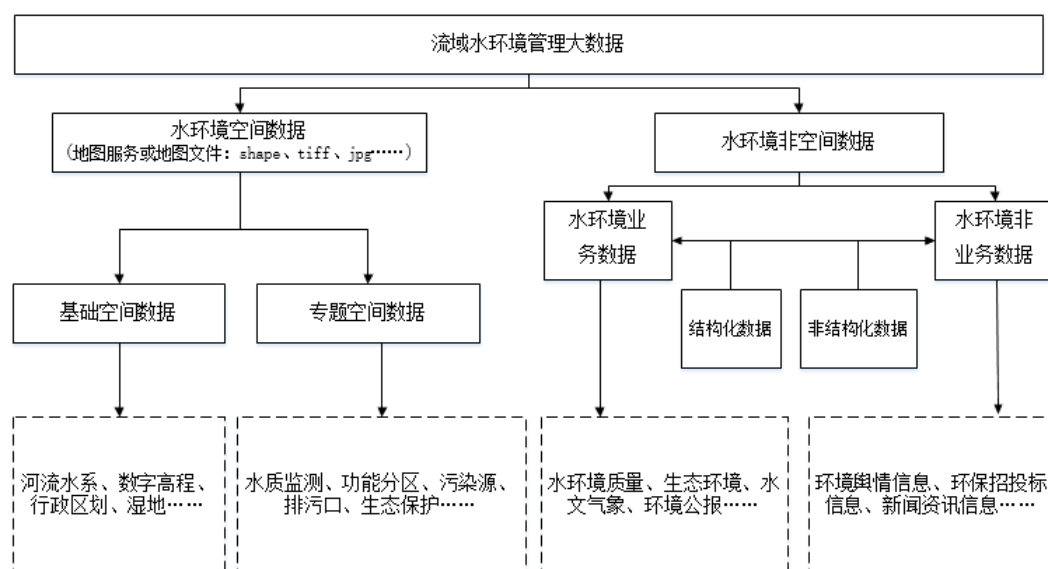


图1 水环境管理大数据组成

流域水环境管理大数据从数据形式上分为水环境空间数据和水环境非空间数据,水环境空间数据包含了基础空间数据和专题空间数据;水环境非空间数据包含了水环境业务数据和水环境非业务数据,且两类数据都可以以结构化和非结构化的形式进行存储。

第三部分: 流域水环境管理大数据存储, 说明了流域水环境管理大数据存储的体系结构、存储模式、存储要求以及流域水环境资源的存储要求, 重点介绍了依托于环境管理职能和业务、数据类型如何进行分类存储, 并对关系型和非关系型数据的存储细化了结构要求。流域水环境管理大数据存储体系结构图如下图所示:

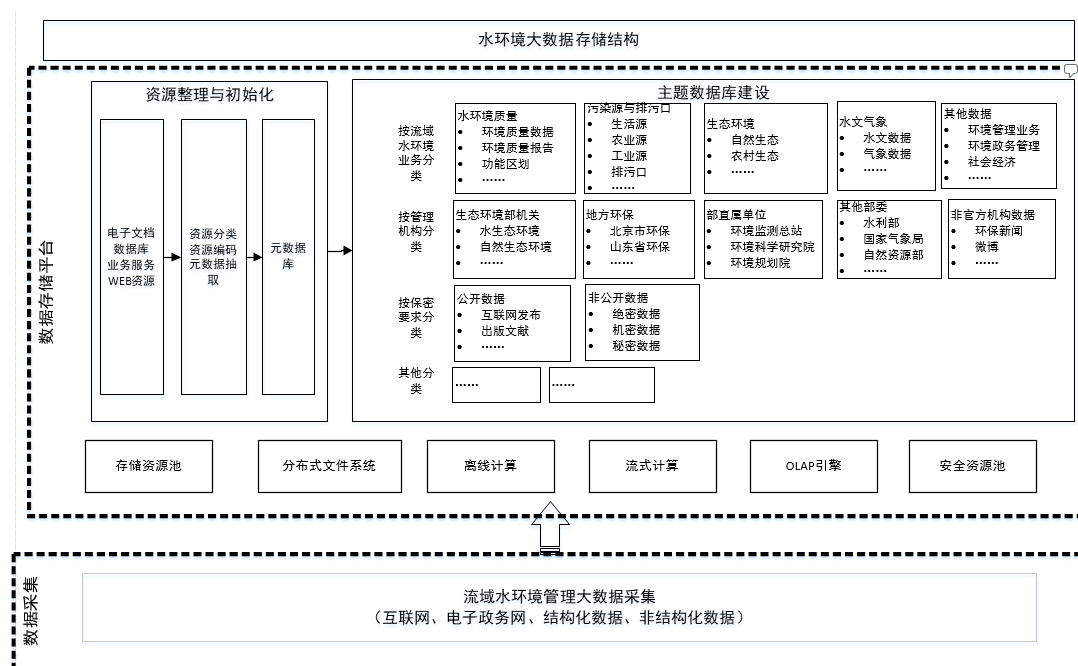


图2 流域水环境管理大数据存储体系结构图

流域水环境管理大数据存储系统以物联网、互联网、专网和政务网等来源的数据为基础, 通过离线计算、流式计算等处理技术对数据进行整理与初始化, 并根据流域水环境管理的业务分主题进行数据库建设。

流域水环境管理大数据的存储模式包含: 分布式存储、集中式存储、内存存储。根据书环境管理数据的大小、来源、用途、适用频次等介绍了各类数据宜适用的存储方式。

流域水环境管理数据以资源存储包形式进行存储结构设计, 包括资源目录的内容对象 (及其相关资源) 的结构化数据及非机构化的存储。实现这两类数据的逻辑存储结构即为资源目录存储结构。

如果文件中涉及媒体对象相关数据文件的具体存储位置时, 该存储位置可以是当前实体文件存储位置的相对路径, 也可以是其他路径。

示例:当前非关系型资源文件存储为“/BssePath/ObjectInfo.xml”(“BssePath”为存储文件的基础存储路径)。该文件中涉及的媒体对象相关数据文件可存储于同一路径下,其表达形式为XML。

第四部分:流域水环境管理大数据交换,说明了数据交换的交换总则、交换模型、网络及安全要求、交换流程以及空间交换的要求参照。此部分重点介绍并引入了资源包作为数据交换的基础结构,规定了资源包的组织结构和命名要求,并对资源包最常用且最基本的调用接口、调用定义等做出了定义和说明。

交换资源包结构如下:

资源信息包用以封装资源对象及其所包含的单一媒体对象(及相关数据文件)信息,其结构如下:

```
Object_Local_Code.zip/  
  
    mimetype  
  
    CONTAINER/  
  
        doc_rdf_instance.xml(实体文件)  
  
        dataFile/  
  
            [media_type_label]/  
  
                某一媒体类型数据文件
```

“Object_Local_Code.zip”为资源对象的本地标识符;交换信息包以“Object_Local_Code.zip”命名,采用ZIP压缩格式对信息包进行压缩。

“mimetype”资源的类型。

对于流域水环境管理的空间数据,主要从空间数据交换模型、交换的格式要求、交换的流程以及交换后数据质量的要求几方面进行规定,从而使流域水环境空间数据的交换能够在完整且完备的目标下有规可依,有迹可寻。

(二) 关键技术说明

本标准的重点在于说明大数据业务系统中各类数据的存储与交换的要求。

本标准基于业务管理需求，根据数据的业务属性、数据格式等分类别进行存储与交换的要求与说明：

(1) 根据流域水环境管理的数据内容，结合环境数据交换规范、大数据存储规范等已有标准规范，根据水环境管理业务数据的内容特征，分别规定了基础空间数据、专题空间数据、水环境业务数据、水环境非业务数据；

(2) 流域水环境大数据业务系统基于大数据生态技术构建了水环境管理业务，大数据存储体系提供大数据的分布式存储管理，涵盖多种存储方式，包括分布式文件存储、分布式结构化数据存储、分布式列式存储、分布式图数据存储。

下面几种成熟和常用的存储模式，实现不同类型数据的存储：

a) 大数据生态模式

目前主流的大数据生态模型多为 Hadoop 生态模式或 Spark 生态模式；

Hadoop 框架的核心是 HDFS 和 MapReduce。其中 HDFS 是分布式文件系统，MapReduce 是分布式数据处理模型和执行环境。

b) 关系型数据库；

对于结构化数据的存储，较多的采用关系型数据库。当前主流的关系型数据库有 Oracle、DB2、PostgreSQL、Microsoft SQL Server、Microsoft Access、MySQL 等。

c) 内存存储；

常用的内存存储有：关系型内存数据库；键值对内存数据库；传统数据库的内存数据库引擎；具有持久化功能的内存库，如 redis、couchbase。

(三) 专利情况说明

标准编制过程中，在国家知识产权局官网等渠道进行了相关专利的检索，经过对标准范围内相关关键词信息的检索，未检索到与本标准核心内容相近的专利，本标准编写及相关参与单位也未依据本标准相关的成果进行专利申请。

七、标准实施的环境效益与经济技术分析

(一) 环境效益分析

本规范能够指导地方对于流域水环境相关数据的统一规范存储和使用，消除同构、异构数据间的交流沟通，提升水环境管理部门依赖于数据的管理能效，进而快速对水环境应急事件或其他管理职能做出响应。

本规范实施后，有利于管理部门统一“十一五”、“十二五”的研究成果，建立基于数据的统一管理标准和应用，推动流域的水环境质量改善，为“十四五”期间实现治理能力现代化提供科技支撑。

（二）经济技术效益分析

本规范经济效益主要体现在社会稳定、环境优美所带来的环境管理成本下降、经济的可持续增长。数据存储与交换标准将减少管理部门对数据交流的投资，也将减少或者免除因数据规整所需的资金，节约系统开发时间，带来间接的经济效益。良好的信息系统支持将增加城市魅力，降低突发环境污染事件造成的经济损失，从而拉动投资，激活经济活力。

流域水环境管理大数据存储与交换规范可带来直接经济效益和间接经济效益，主要体现在以下几个方面：

一是以大数据技术为支撑，建立统一的大数据存储与交换规范，支持国家生态环境大数据建设。用海量生态环境大数据筑起“绿水青山”的数据堡垒，通过快速、实时的数据存储与交换，及时掌握流域水环境状况，提前预防和应对环境风险事故，减轻水环境污染对人民健康和国民经济造成的损失，为保护和改善生态环境质量提供技术标准支撑。

二是通过信息化手段提升流域生态环境监管能力，尤其是流域水环境监管工作能效，有限的人力资源得到更合理的利用，并使管理达到最优化，节约人力物力。通过综合采用大数据、云计算、模型分析等现代化技术手段，利用“标准化+大数据技术”对提高行业管理的监管效率、规范行业数据，从而促进环境监管的自动化和智能化，提升管理质量，减少人力、物力及时间成本，提高生态环境保护相关政府职能部门的办事效率，为构建新型流域监管模型和风险防控提供数据基础支撑。

三是提升风险防范能力和综合决策能力，带来间接经济效益。本规范的实施将有助于强化生态环境部门的风险防范能力，通过科技手段使得日常监管力度增加，促进环境风险管理能力的提升，能够最大限度的防范和降低突发环境污染事件造成的经济损失，带来较大的间接经济效益。

八、标准实施建议

（一）加强互联网、物联网技术衔接

当下互联网、物联网，每3年对互联网、物联网技术定位进行评判，确定新的技术对数据存储与交换的影响升级。

（二）加强与流域水环境资源目录的衔接

当前数据存储业务目录基本结构依照生态环境部业务管理设计，“十四五”规划中，生态文明建设占据重要的位置和组成，为了更好的完成生态文明的建设，生态管理部门业务及组织的调整能够影响到资源目录的组成，因此业务数据的存储和交换应当加强与流域水环境资源目录的衔接。

（三）适时开展规范实施评估

建议每五年对已经制定的流域大数据存储与交换规范开展实施评估，根据信息化技术和业务管理部门职能调整来决定是否停止或执行流域水环境管理大数据存储和交换规范。

（四）规范贯彻措施

宣贯方式主要包括提供宣贯资料、举办培训班和召开宣贯会。

标准项目筹备：依托中国环境科学学会平台，邀请相关单位参与标准研制，组织标准项目筹备会议。

过程宣传：研制过程重要节点事件和会议宣传。

标准解读：每项标准由起草组提供标准解读文章，由中国环境科学学会统一对外发布。

标准培训：根据会议需求调查情况，组织线上和线下培训。

标准宣贯及技术研讨会：学会联合分会或牵头单位组织标准宣贯和技术研讨会。

（五）其他情况说明

本规范适用于水环境管理信息调用、处理以及水环境管理相关信息系统建设的指导工作。

本规范为首次发布。

本规范为指导性标准。