



National Data Service and the Research Data Eco-System

Landscape Analysis
by WayMark Systems

*Support from the National Science Foundation is deeply appreciated:
NSF-VOSS EAGER 0956472, "Stakeholder Alignment in Socio-Technical Systems,"
NSF OCI RAPID 1229928, "Stakeholder Alignment for EarthCube,"
NSF GEO-SciSIP-STS-OCI-INSPIRE 1249607, "Enabling Transformation in the Social
Sciences, Geosciences, and Cyberinfrastructure,"
NSF I-CORPS 1313562 "Stakeholder Alignment for Public-Private Partnerships"*



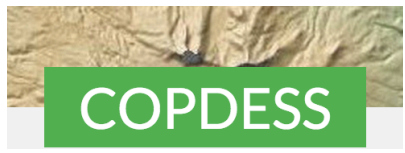
Extreme Science and Engineering
Discovery Environment



RESEARCH DATA ALLIANCE



PROTEIN DATA BANK



The data

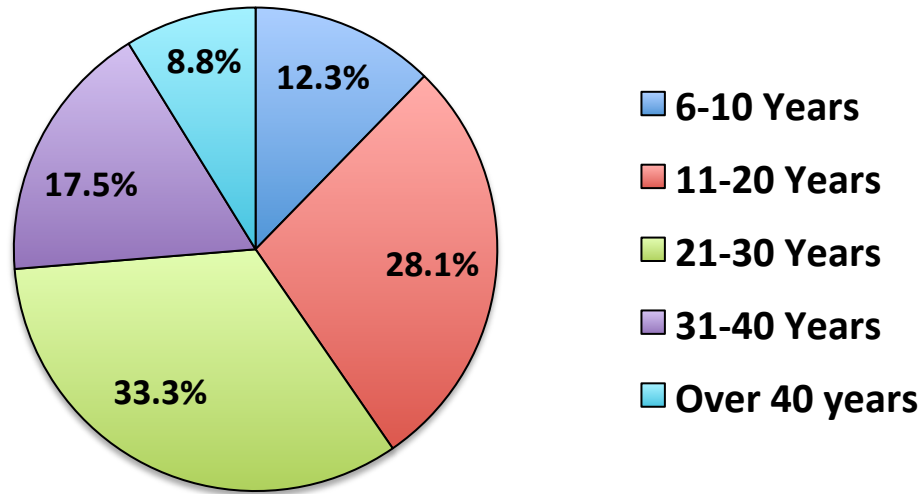
Introduction:

- Survey of individuals on the distribution list of the National Data Service (n=305) with data from 57 responses reported here (response rate of 19%).
- Respondents select initiatives and consortia with which they are familiar
- Data collection continuing – so results are not final and additional analysis needed.

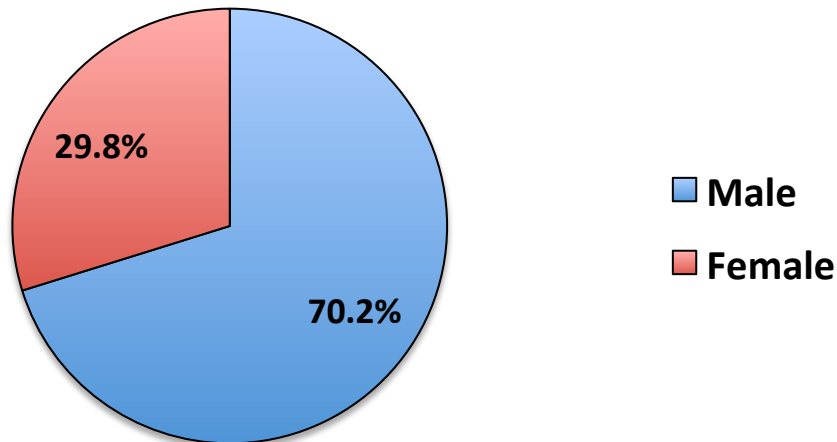
Reported responses:

- The landscape analysis first reports initiatives and consortia selected
- Among those selected: How large are the challenges facing this initiative?
 - 0=*none or minimal*; 1=*minor*; 2=*moderate*; 3=*major*; 4=*massive*; 9=*don't know/not applicable*
- Among those selected: How important is the success of this initiative to your work?
 - 0=*Not important at all*; 1=*minor importance*; 2=*moderate importance*; 3=*major importance*; 4=*absolutely essential*; 9=*don't know/not applicable*
- The one initiative or consortia that is the most important to your work right now and the most important “must have” from this initiative or consortia

Respondent profile



Years in Primary Area of Professional Experience



Gender

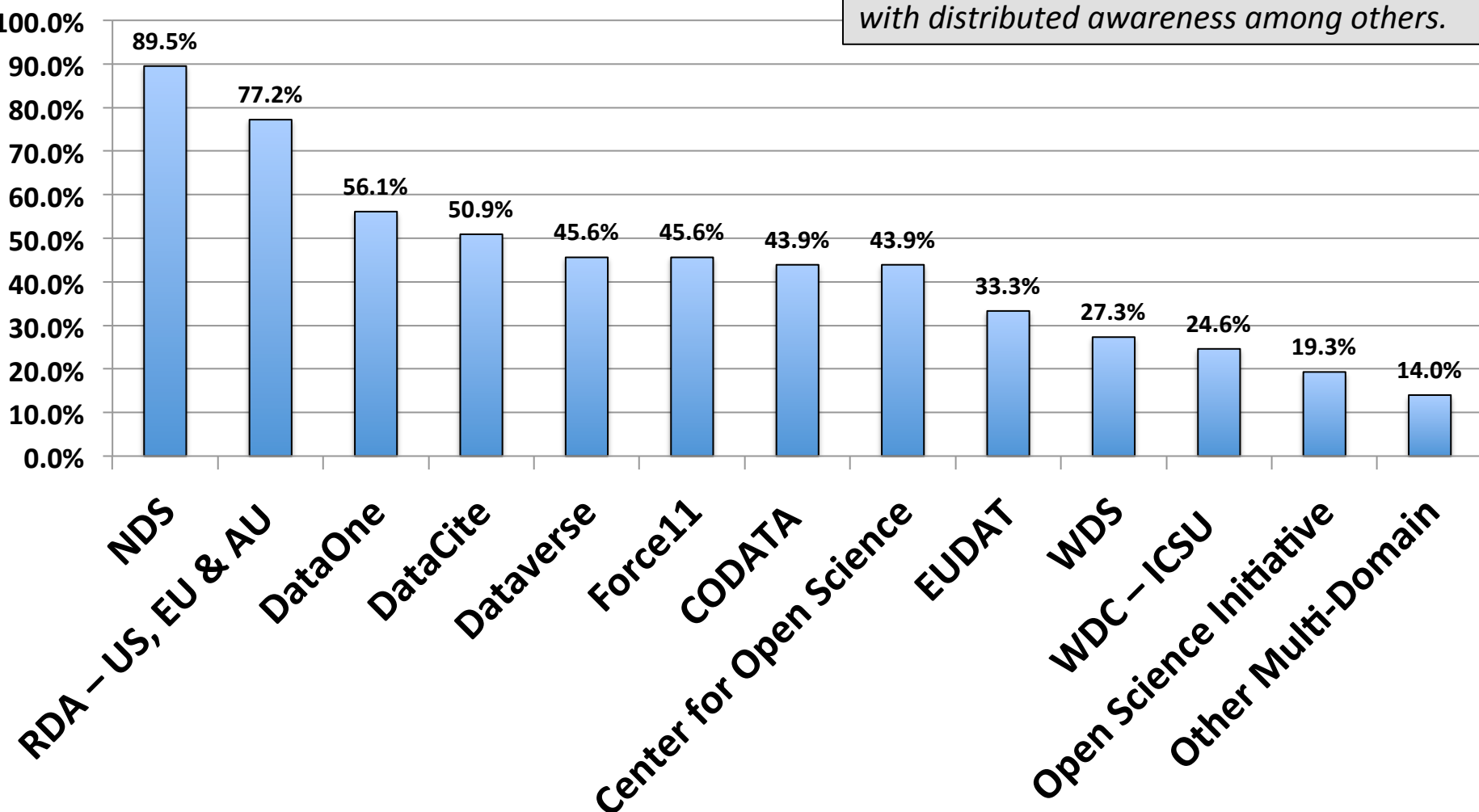
Familiar with Initiatives and Consortia



Painting by David Kupferman

Familiar with Multi-Domain Initiatives and Consortia

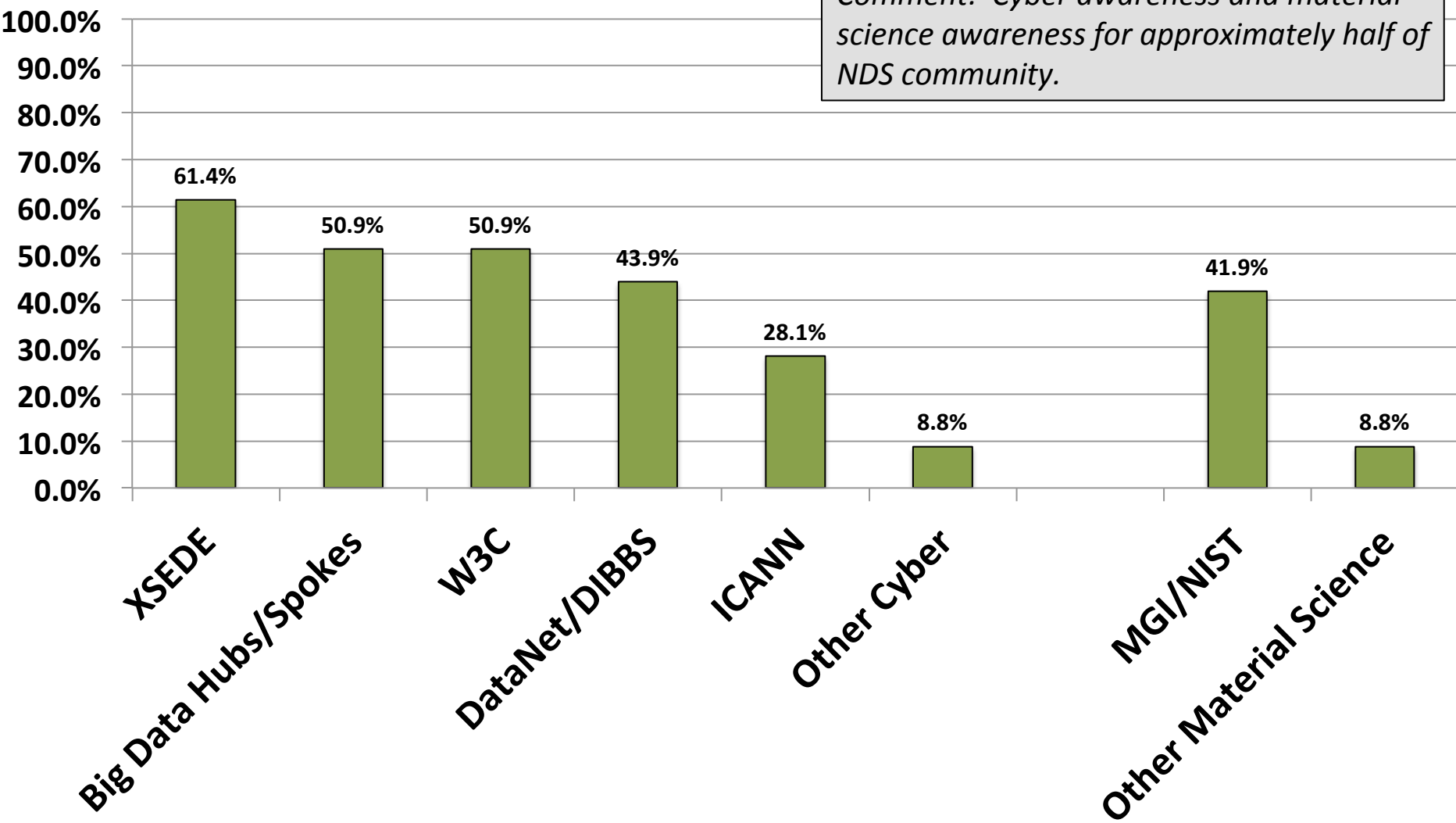
Comment: NDS and RDA as leading pair, with distributed awareness among others.



Other Multi-Domain: Arctic and Antarctic Data consortia, Crossref, Dryad, Figshare, Globus, Open Science Grid, Zenodo

Familiar with Cyberinfrastructure and Material Science Initiatives

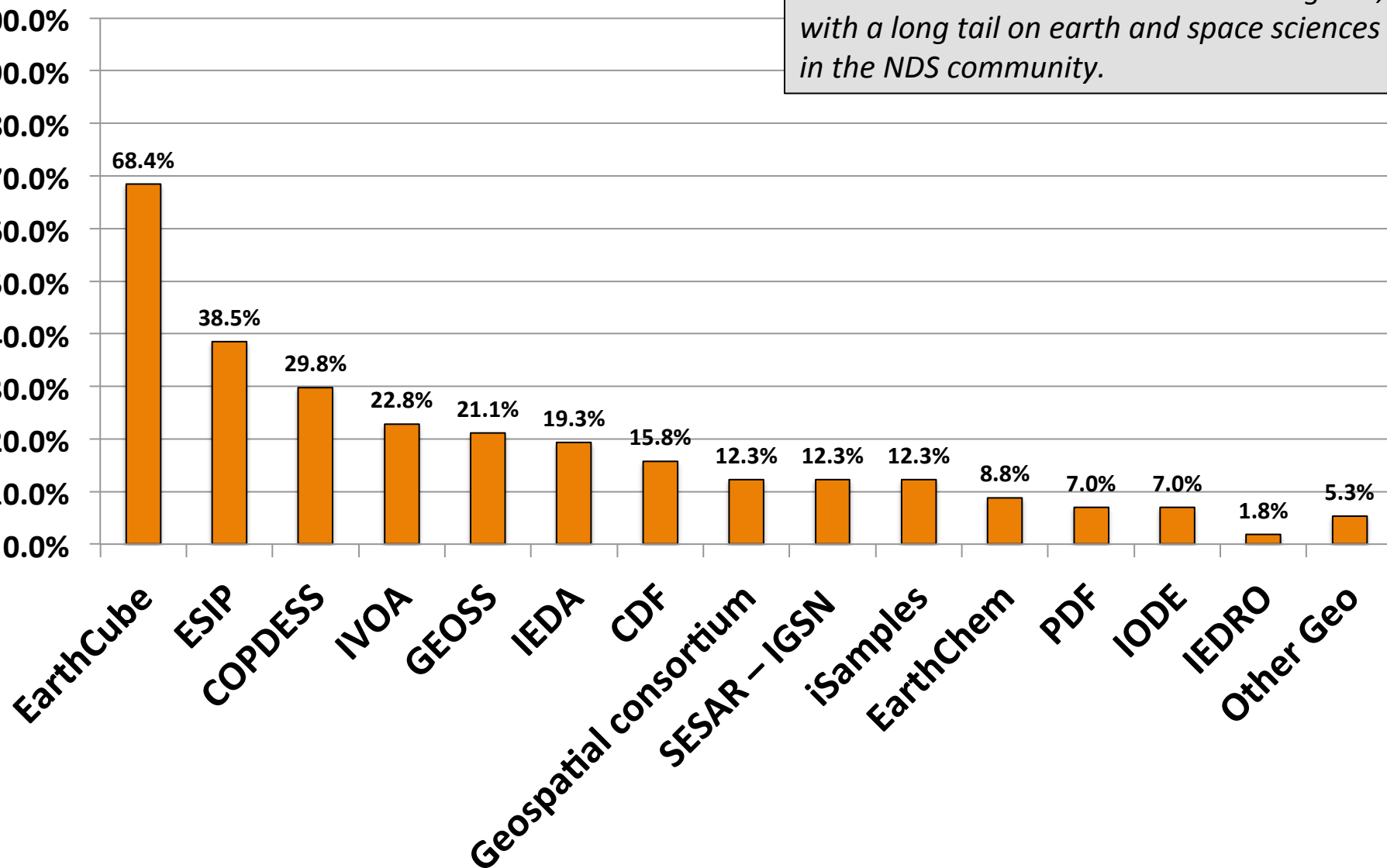
Comment: Cyber awareness and material science awareness for approximately half of NDS community.



Other Cyberinfrastructure: DPN, EGI, EGF, IETF, Globus, OGC, “probably many others”
Other Materials Science: CHIMAD, LBL projects, Materials Data Facility, PRISMS,

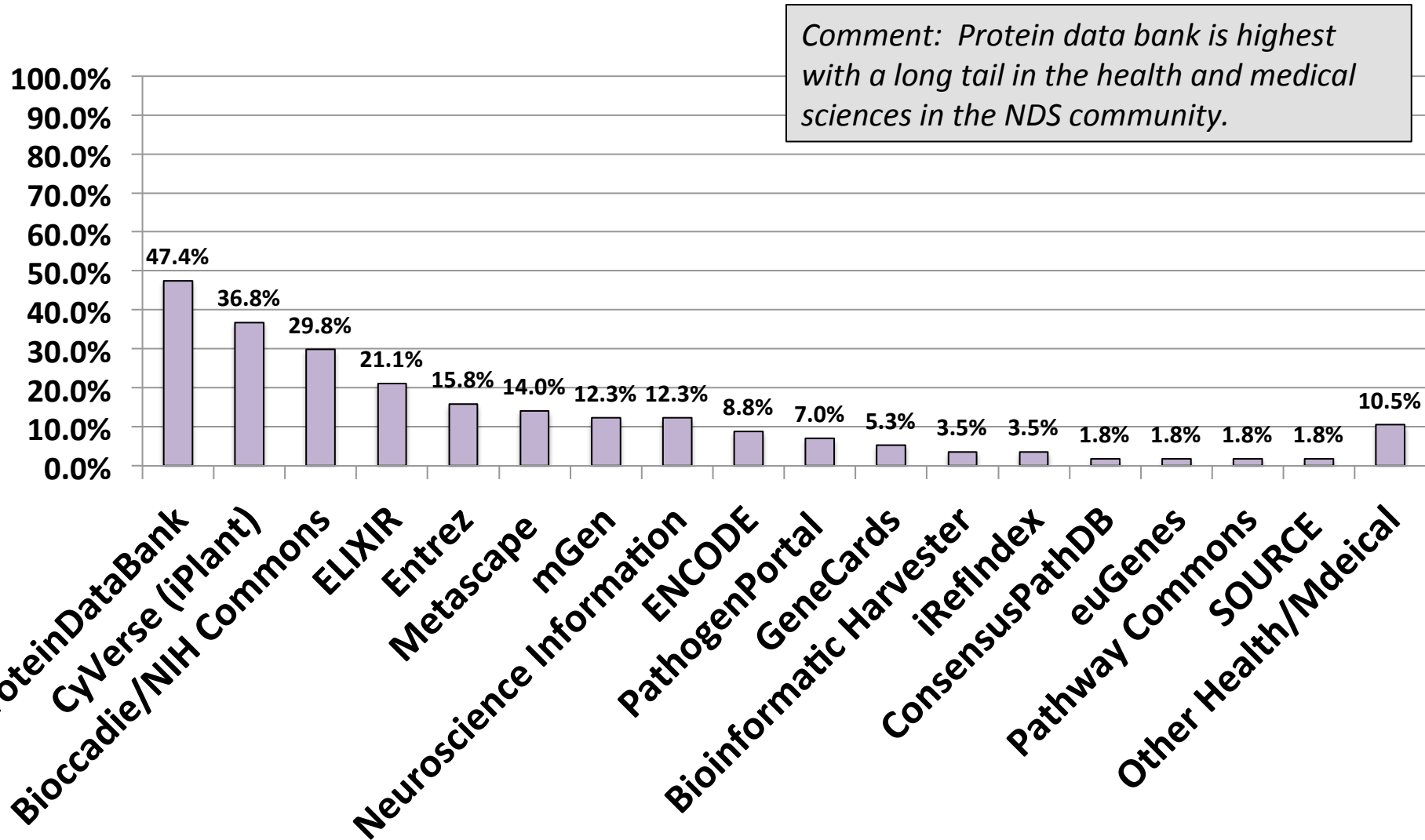
Familiar with Earth & Space Science Initiatives

Comment: EarthCube awareness is highest, with a long tail on earth and space sciences in the NDS community.




Other Earth/Space Science: Earth System Grid Federation, IRIS, OneGeology

Familiar with Biological/Health/Medical Initiatives

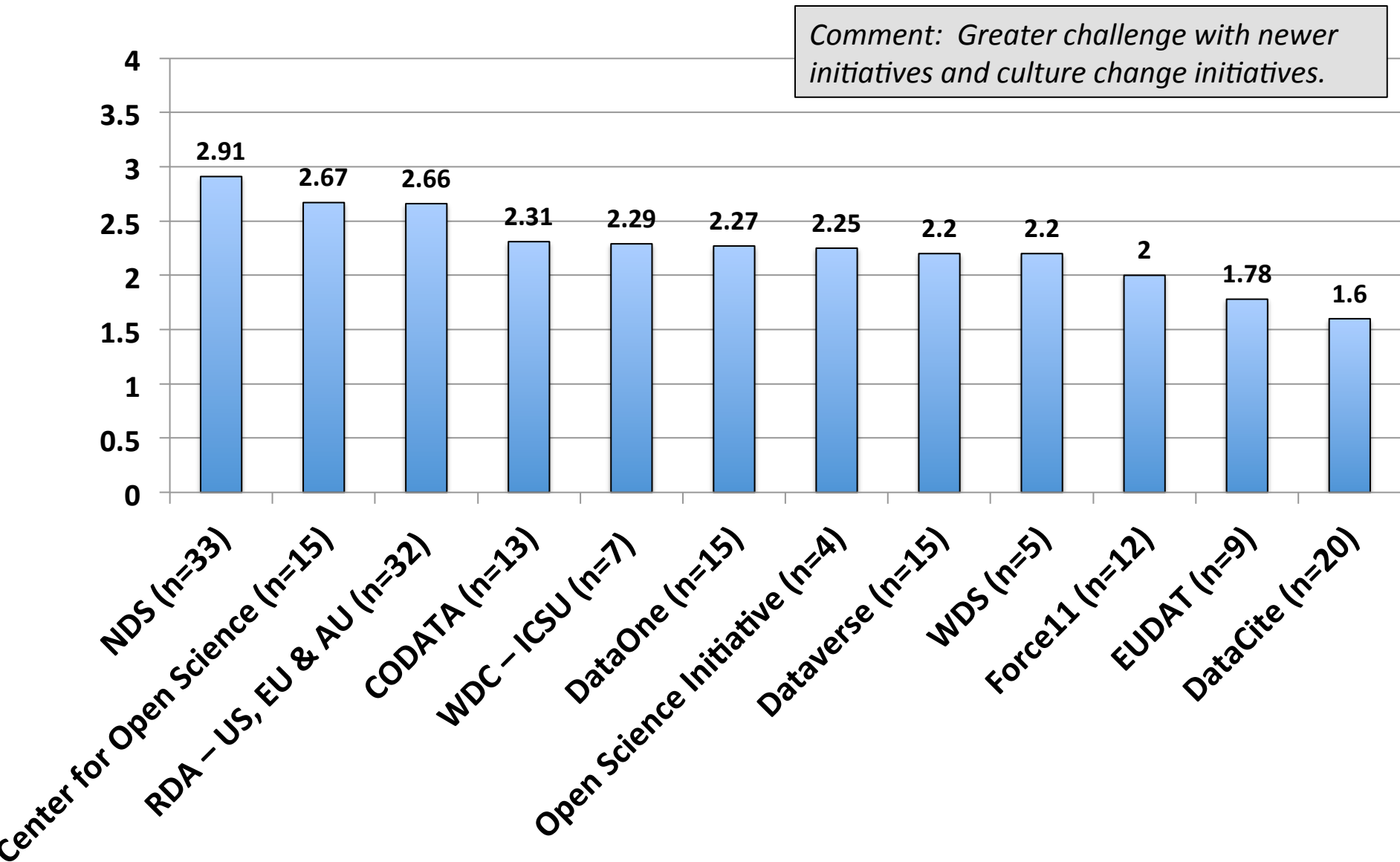


No response: BioGraph, BioGPS, Human Epigenome Atlas, MetaBase, MOPED, Nowomics
Other Health/Medical: BDDS Center, CGHub, Eukaryotic Linear Motif resource, European Bioinformatics Institute, Nucleic Acids Database, PSI, SBGrid, SwissProt, Wormbase



How large are the challenges facing this initiative?

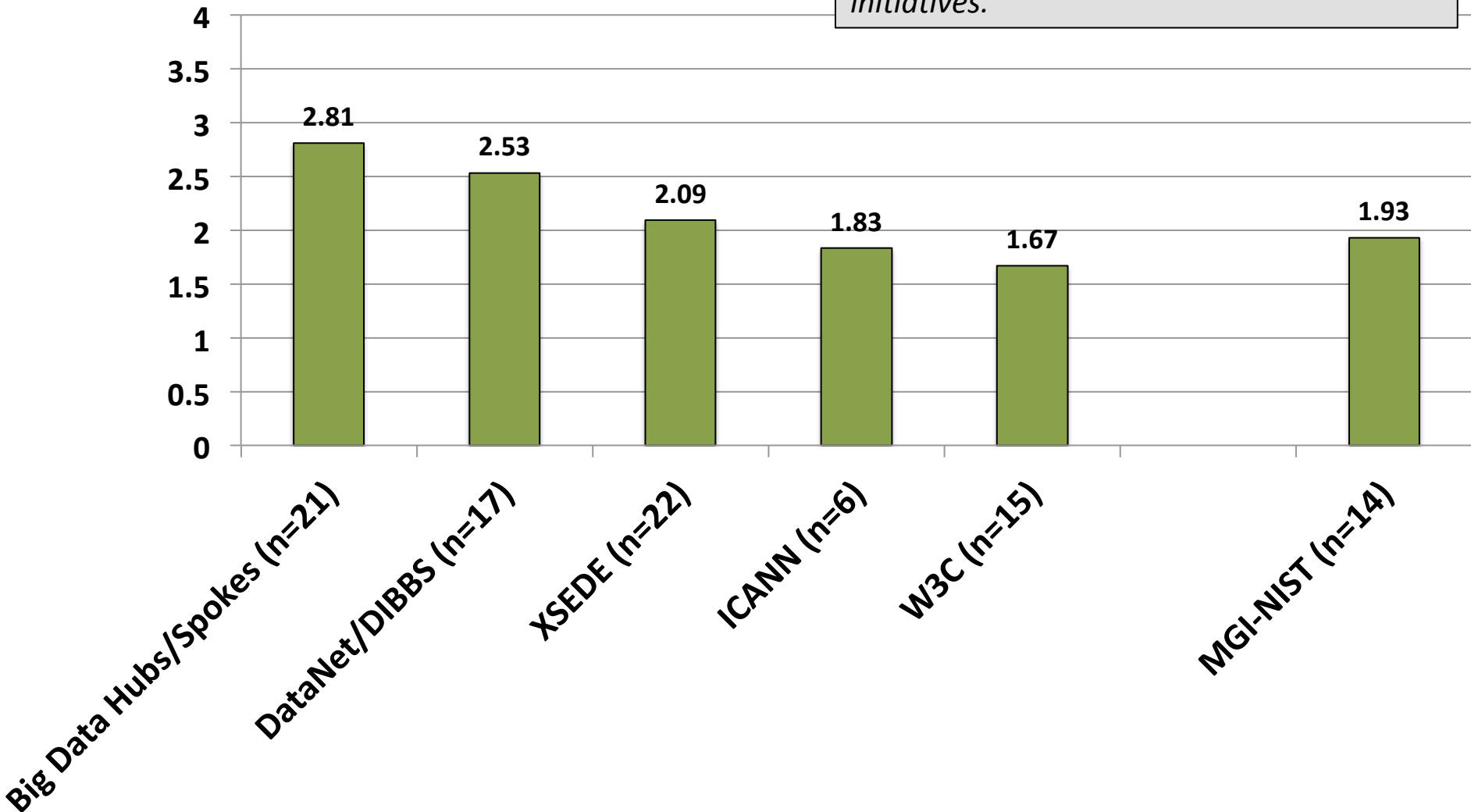
Size of Challenges Facing Multi-Domain Initiatives and Consortia



0=none or minimal; 1=minor; 2=moderate; 3=major; 4=massive; missing=don't know/not applicable

Size of Challenges Facing Multi-Domain Initiatives and Consortia

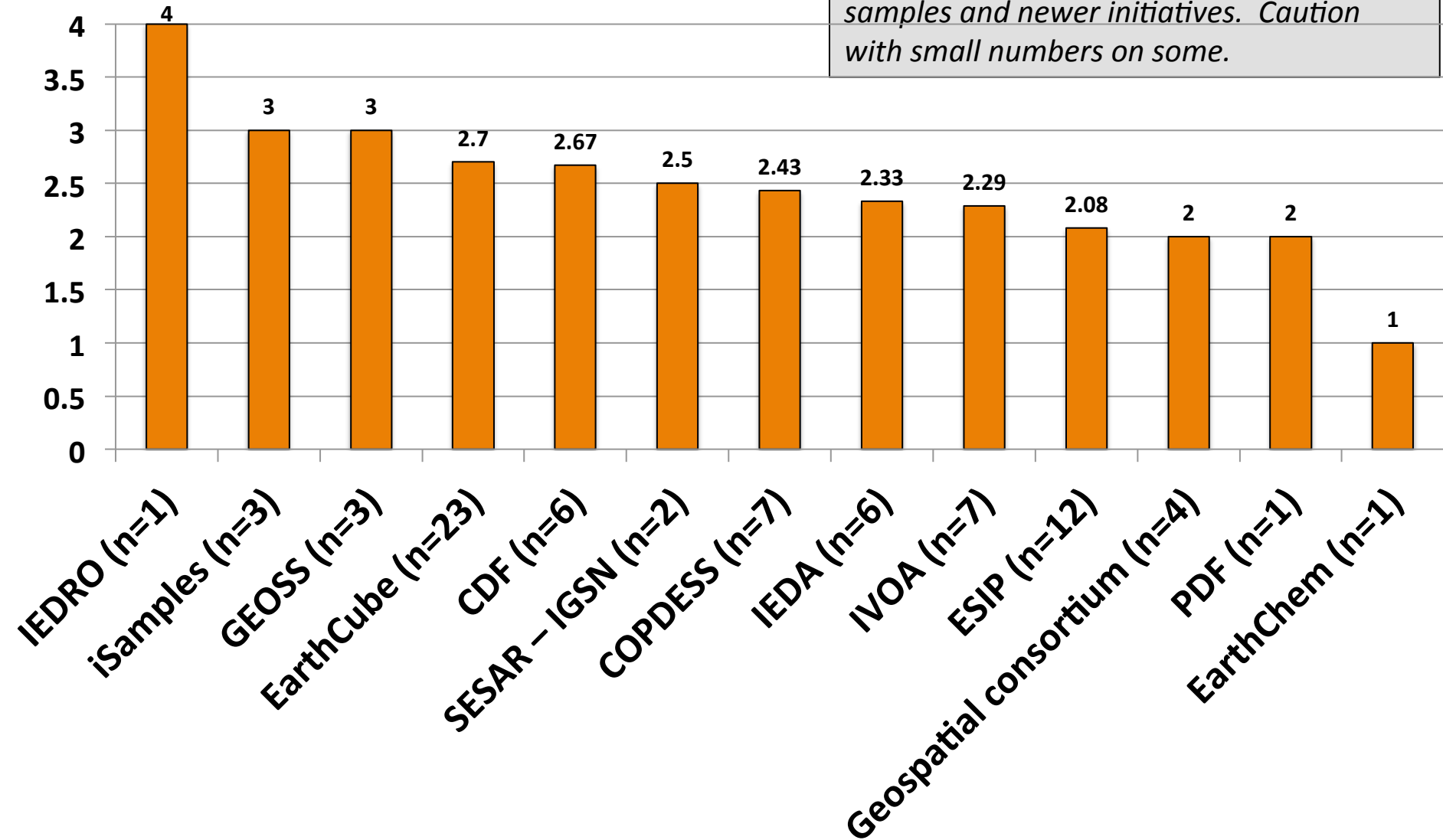
Comment: Greater challenge with newer initiatives.



0=none or minimal; 1=minor; 2=moderate; 3=major; 4=massive; missing=don't know/not applicable

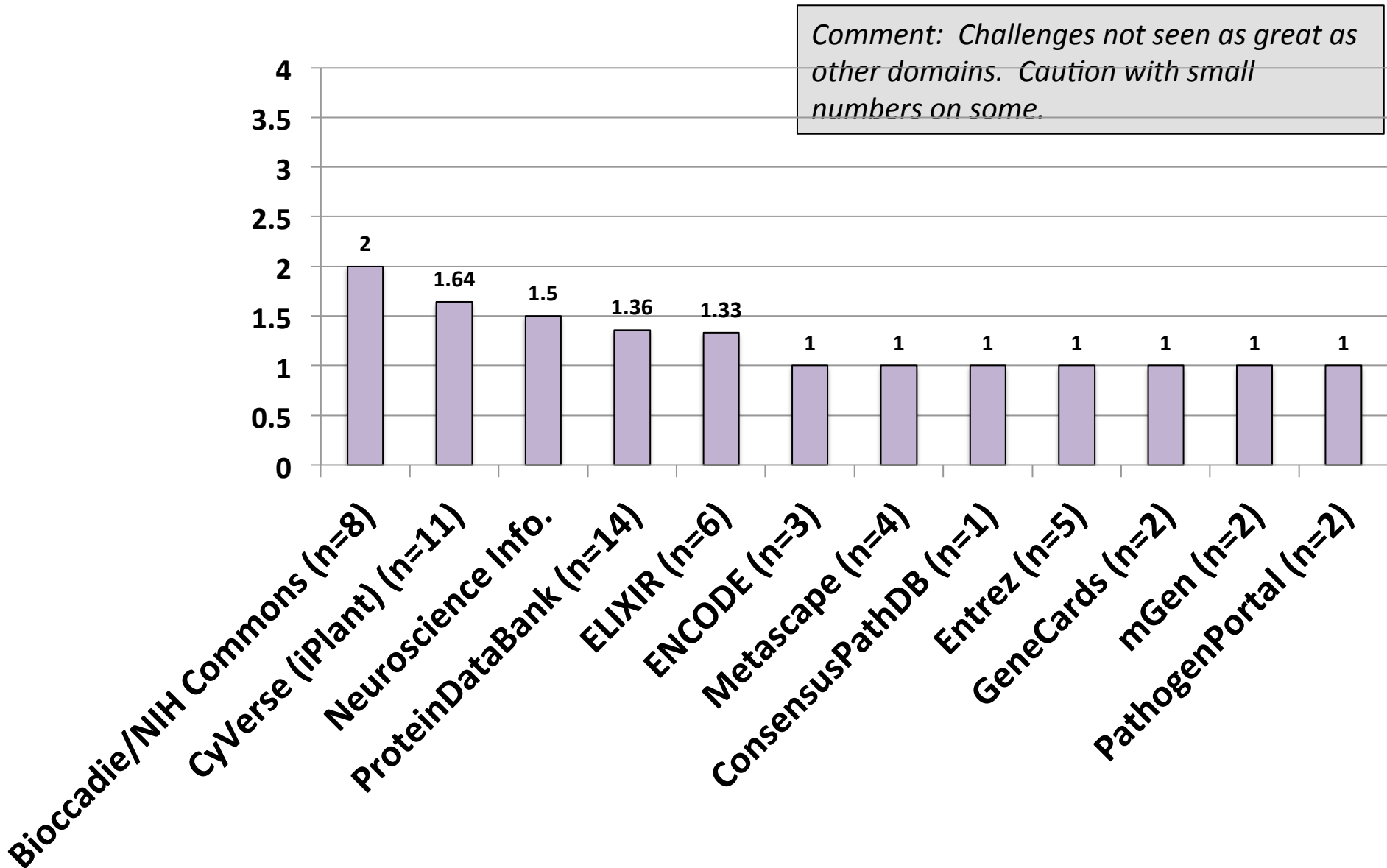
Size of Challenges Facing Earth & Space Science Initiatives

Comment: Greater challenges with physical samples and newer initiatives. Caution with small numbers on some.



0=none or minimal; 1=minor; 2=moderate; 3=major; 4=massive; missing=don't know/not applicable

Size of Challenges Facing Biological/Health/Medical Initiatives

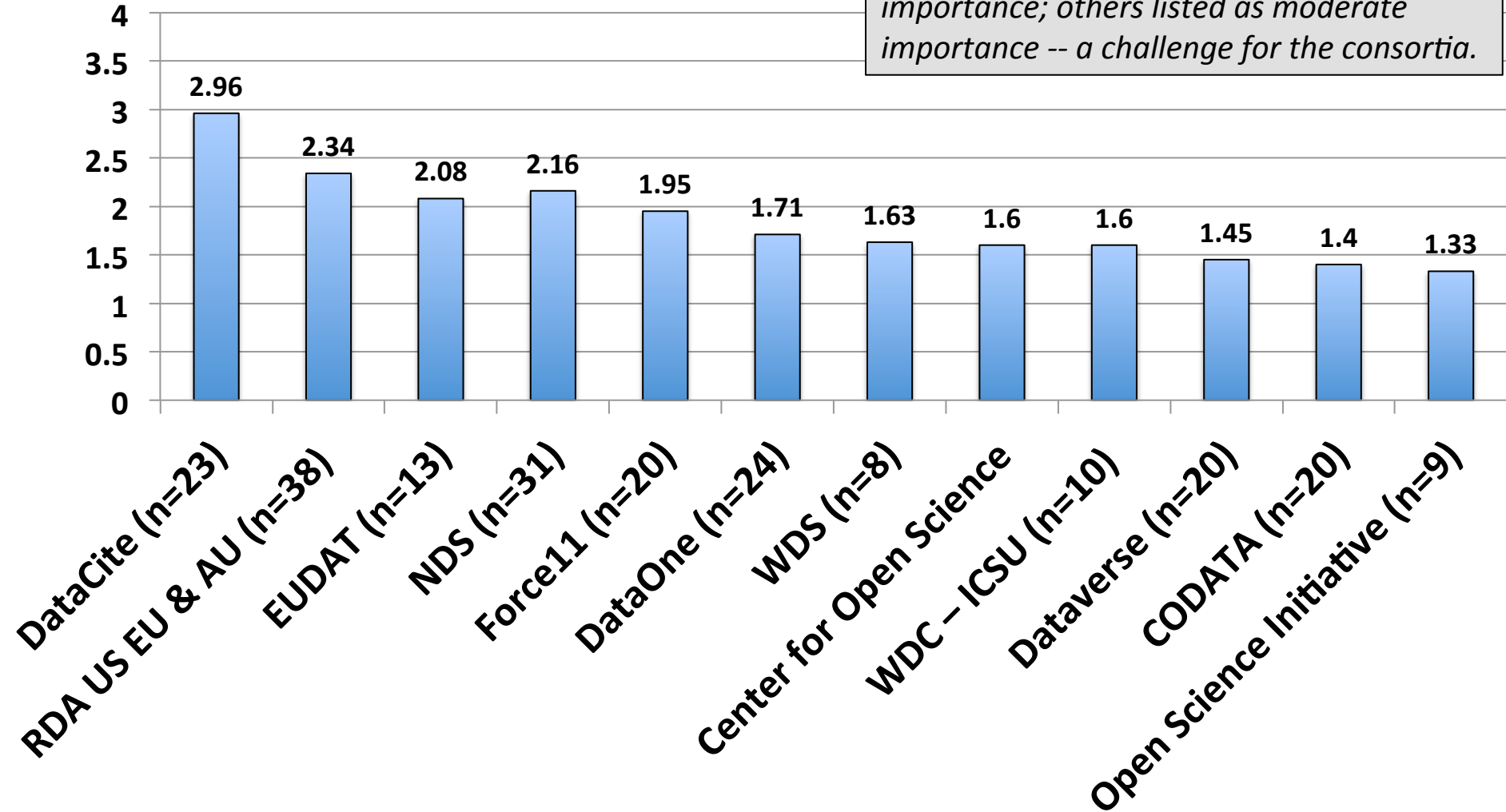


0=none or minimal; 1=minor; 2=moderate; 3=major; 4=massive; missing=don't know/not applicable

How important is the success of this initiative
to your work?

Importance to Your Work of Success by Multi-Domain Initiatives and Consortia

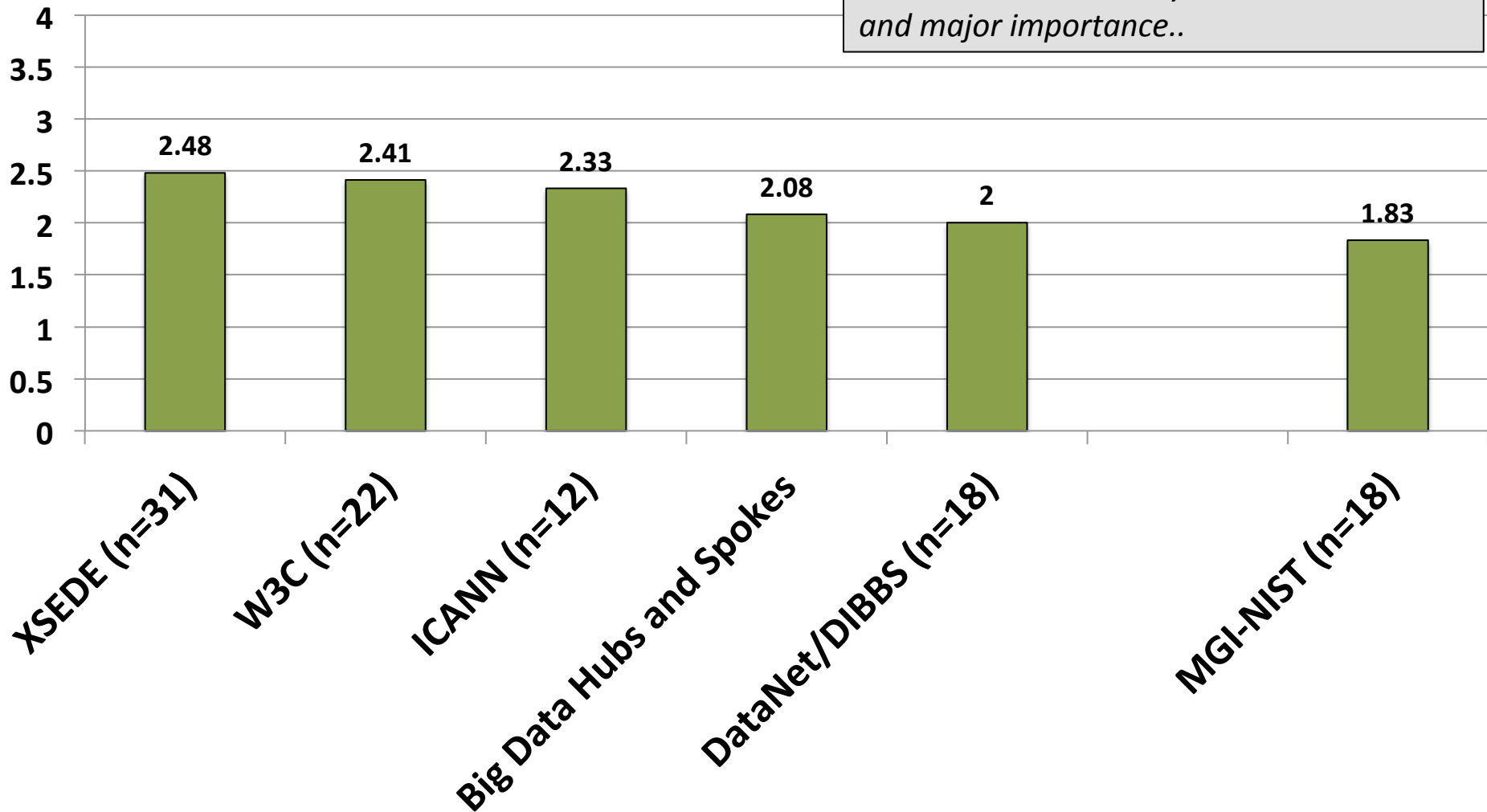
Comment: DataCite is of major importance; others listed as moderate importance -- a challenge for the consortia.



0=Not important at all; 1=minor importance; 2=moderate importance; 3=major importance; 4=absolutely essential; missing=don't know/not applicable

Importance to Your Work of Success by Cyberinfrastructure and Material Science Initiatives

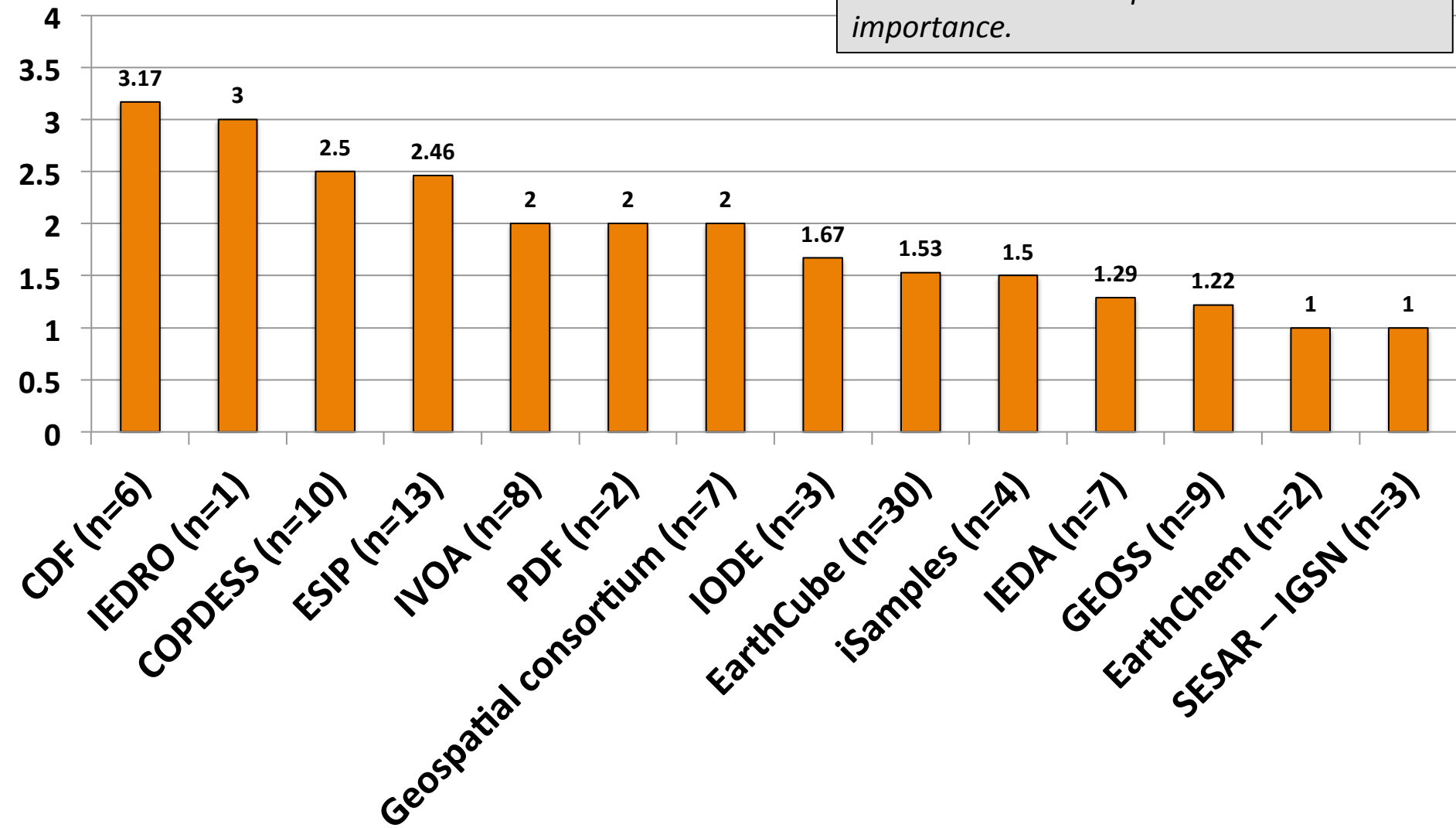
Comment: Consistently between moderate and major importance..



0=Not important at all; 1=minor importance; 2=moderate importance; 3=major importance; 4=absolutely essential; missing=don't know/not applicable

Importance to Your Work of Success by Earth & Space Science Initiatives

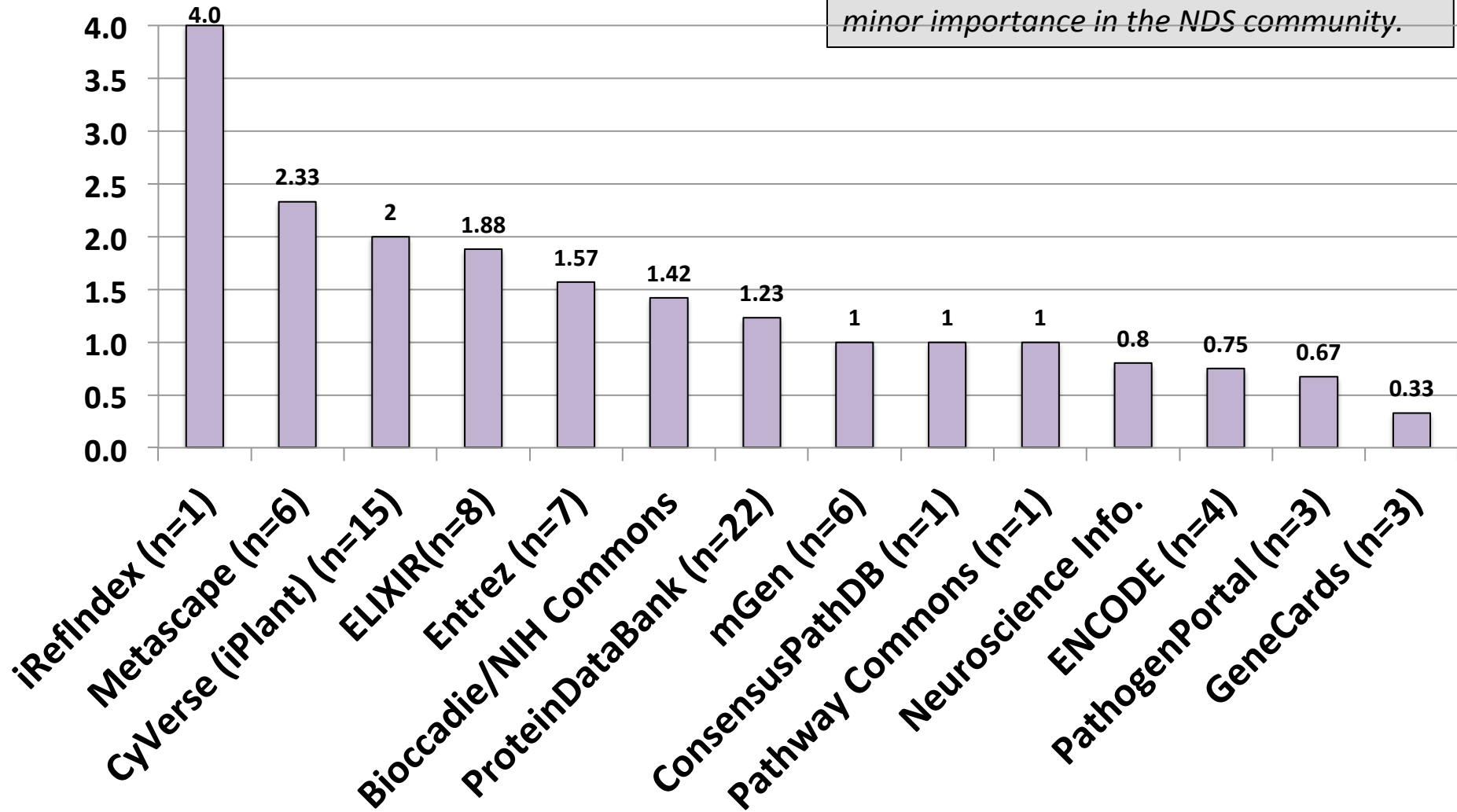
Comment: A broad spectrum in indicated importance.



0=Not important at all; 1=minor importance; 2=moderate importance; 3=major importance; 4=absolutely essential; missing=don't know/not applicable

Importance to Your Work of Success by Biological/Health/Medical Initiatives

Comment: Most listed as moderate to minor importance in the NDS community.



0=Not important at all; 1=minor importance; 2=moderate importance; 3=major importance; 4=absolutely essential; missing=don't know/not applicable

Most Important Initiative or Consortia
and one “must have”

Most Important Initiative or Consortia and one “must have”

Multi-Domain Initiatives and Consortia (selected response)

- Center for Open Science
 - To continue the Open Science Framework
 - Planning on using the OSF REST API for creating custom dashboards for projects
- CODATA
 - Authoritative recognition
- Dataverse
 - Interoperate with other systems
- DataCite
 - Unencumbered ability of authors to directly obtain DOIs for their data sets.
 - Consistent element definitions
 - Standardizing how the metadata fields are completed for DOIs by different repositories
- DataOne
 - Expansion of the DataONE federation to include more environmental data repositories.
- NDS
 - Unified web service access to virtual datasets. Not just files.
 - Common architecture to deploy tools on local clouds
 - NASA involvement in NDS
 - Tools to facilitate interoperability across domains / standards / data models
- RDA
 - Good platform for connecting to an interdisciplinary, international, multi-stakeholder community concerned with Research Data Management issues.
 - Must be a bottom-up activity with a strong top-down support!
 - Near production-ready data services that are implementable at the campus level.

Most Important Initiative or Consortia and one “must have” (cont.)

- Other Multi-Domain Initiatives
 - Reliable cross-connections of DOIs between Crossref, Datacite and other data providers
 - Tighter (provenance enabling) integration with data generating research instruments and equipment
 - Technology for seamlessly sharing computing resources for High Throughput data processing
 - Identify workflows

Cyberinfrastructure Initiatives and Consortia (selected responses)

- Big Data Hubs and Spokes
 - Building a sustainable model for spokes/hubs
 - Partners with applications that require big data solutions
- DataNet/DIBBS
 - Long-term sustainability of the scientific data of their customers
- W3C
 - Browser standards updates - declarative formats for human computer interactions
- XSEDE
 - Access to HPC resources in the form of both hardware and people who understand HPC
 - Deliver on the promise of campus bridging
 - More support. They're underfunded
 - Access to SC resources and advanced user support
 - Data curation and preservation
- Other Cyberinfrastructure
 - Data discovery

Most Important Initiative or Consortia and one “must have” (cont.)

Material Science Initiatives and Consortia (selected responses)

- Materials Genome Initiative/NIST
 - Discovery, access, and interoperability across materials science data resources
 - I am an experimentalist: therefore computational work that clearly lays out what kind of materials should be made for some kind of application would be great
 - Alloy chemistry and property data along with heat treatment protocol followed
 - Definition of a standardized approach to creating interoperable domain metadata in MatSci subfields

Earth & Space Science Initiatives and Consortia(selected responses)

- EarthCube
 - Explicit long-term commitment & support from NSF
 - Restructured organization to facilitate convergence on system design
 - Alignment of fields and disciplines around the open sharing and reuse of data, models, and software
 - Useful products and services, including datasets and tools
- Geospatial Consortium
 - Geospatial semantic interoperability
- Federation of Earth Science Information Partners (ESIP)
 - Support for education and outreach
 - A definition of roles and responsibilities for earth science data stewardship

Most Important Initiative or Consortia and one “must have” (cont.)

Biological/Health/Medical Initiatives and Consortia (selected responses)

- Bioccadie/NIH Commons
 - Tighter collaboration with RDA
 - To settle on a common cloud environment and data access modalities
- Cyverse (formerly iPlant)
 - Fully stated long-term sustainability model
 - Sustainability/longevity
 - Institutional buy-in
- mGen
 - Timely access to data
- Other Biological/Health/Medical
 - improved software and APIs
 - Integration of data and software infrastructure

Comment: Among 44 respondents there are over 20 individual consortia identified as most important by people on the NDS distribution list, with “must haves” that are distinct value propositions for each.

Sample landscape analysis format from EarthCube

Research Domain	Organizational Domain									
	Professional Organizations and Societies			Umbrella Organization Networks		Data Facilities		Research Infrastructure		Long Term Projects (5+ years)
	Upper Atmosphere					NASA DAACs Unidata NGDC (NCEI)		EISCAT AMISR		UCAR/NCAR
	Lower Atmosphere					Unidata NCDC (NCEI) NASA DAACs		NEON Eslri Living Atlas Byrd Polar and Climate Research Center CyberGIS Center (U. Illinois) Polar Geospatial Center		UCAR/NCAR
	Cryosphere					PGC NASA Cryosphere Science Research Portal		Eslri Living Atlas CIRES IARC Polar Geospatial Center		BCube
	Oceans					IEDA BCO-DMO CCHDO NODC (NCEI) R2R/ UNOLS		Scripps GDC IODP NCDC (NCEI) NGDC (NCEI)		SCOPE HOT OceanOBs NCEAS
	Earth Surface					IEDA NGDC (NCEI) USGS EROS State Geo. Surveys CUAHSI WDC Neotoma CLIVAR		NAMSS ASP@LDEO ASP@UTIG NYS GIS		LTER BCube NCEAS
	Deep Earth					IRIS DMC USGS NAMSS Scripps GDC IODP		ASP@LDEO ASP@UTIG NGDC (NCEI)		UNAVCO IRIS DMC WHOI

Key to Abbreviations

Multi-Domain Initiatives and Consortia:

- Center for Open Science
- CODATA
- Dataverse
- DataCite
- DataOne
- EUDAT
- Force11
- National Data Service (NDS)
- Open Science Initiative
- Research Data Alliance (RDA) – US, EU, and AU
- World Data Center (WDC) – ICSU
- World Data System (WDS)
- Other Multi-Domain

Cyberinfrastructure Initiatives and Consortia

- Big Data Hubs and Transformational Science/Spokes (NSF)
- DataNet/DIBBS
- Extreme Science and Engineering Discovery Environment (XSEDE)
- Internet Corporation for Assigned Names and Numbers (ICANN)
- World Wide Web Consortium (W3C)
- Other Cyberinfrastructure

Material Science Initiatives and Consortia

- Materials Genome Initiative (MGI)/National Institute of Standards & Technology (NIST)
- Other Material Science

Key to Abbreviations (cont.)

Earth and Space Science Initiatives and Consortia:

- Coalition on Publishing Data in the Earth and Space Sciences (COPDESS)
- Council of Data Facilities (CDF)
- Earth Science Information Partners (ESIP)
- EarthChem
- EarthCube
- Federation of Earth Science Information Partners (ESIP)
- Geospatial consortium
- Global Earth Observation System of Systems (GEOSS)
- Interdisciplinary Earth Data Alliance (IEDA)
- International Environmental Data Rescue Organization (IEDRO)

- International Oceanographic Data and Information Exchange (IODE)
- International Virtual Observatory Alliance (IVOA)
- iSamples
- Polar Data Forum (PDF)
- System for Earth Sample Registration (SESAR) – IGSN
- Other Earth & Space Science

Biological, Health and Medical Initiatives and Consortia:

- Biocaddie/NIH Commons
- BioGraph (University of Antwerp, Vlaams Instituut voor Biotechnologie)
- BioGPS (The Scripps Research Institute)
- Bioinformatic Harvester (Karlsruhe Institute of Technology)
- ConsensusPathDB

Key to Abbreviations (cont.)

Biological, Health, and Medical Initiatives and Consortia (cont.)

- CyVerse (formerly iPlant)
- ELIXIR
- Encyclopedia of DNA Elements (ENCODE)
- Entrez (National Center for Biotechnology Information)
- euGenes (Indiana University)
- GeneCards (Weizmann Institute)
- Human Epigenome Atlas
- iRefIndex (BIND, BioGRID, CORUM, DIP, HPRD, InnateDB, IntAct, MatrixDB, MINT, MPact, MPIDB, MPPI and OPHID)
- MetaBase (KOBIC)
- Metascape (NCBI (Entrez, OMIM, ClinVar), GO, KEGG, MSiGDB, UniProt, Protein Atlas, Ensembl, JAX, DrugBank, NHGRI-EBI, DDG2P)
- mGen (GenBank, Refseq, EMBL and DDBJ)
- MOPED (Seattle Children's Research Institute)
- Neuroscience Information Framework (University of California, San Diego)
- Nowomics
- PathogenPortal (National Institute of Allergy and Infectious Diseases (NIAID))
- Pathway Commons (Memorial Sloan-Kettering Cancer Center and University of Toronto)
- ProteinDataBank
- SOURCE (Stanford University)
- Other Health/Medical

Appreciation

NDS Consortium Steering Committee

- Alan Blatecky, RTI International, Visiting Fellow
- Mercè Crosas, Director of Data Science IQSS, Harvard University
- Ted Habermann, Director of Earth Science, The HDF Group
- Robert J. Hanisch, Director, Office of Data and Informatics Material Measurement Laboratory, National Institute of Standards and Technology
- Carole Palmer, Associate Dean for Research Information School, University of Washington
- Ed Seidel, National Center for Supercomputing Applications, Director University of Illinois Urbana-Champaign
- Anita de Waard, VP for Research Data Services, Elsevier Publishing

Ex Officio

- John Towns, NDS Interim Director
- Christine Kirkpatrick (chair TAC), San Diego Supercomputing Center
- Joel Cutcher-Gershenfeld, facilitator, Brandeis University



Stakeholder Alignment Collaboration:

Karen Baker, UIUC

Nick Berente, University of Georgia

Dorothy Carter, University of Georgia

Joel Cutcher-Gershenfeld, Brandeis University

Leslie DeChurch, Georgia Tech University

Courtney Flint, Utah State University

Gabriel Gershenfeld, LA Dodgers

Michael Haberman, UIUC

John L. King, University of Michigan

Eric Knight, University of Sydney

Barbara Lawrence, UCLA

Chris Lenhardt, RENCi

Spenser Lewis, Draper Labs

Matt Mayernik, NCAR/UCAR

Charles McElroy, Case Western Reserve

Barbara Mittleman, NIH (retired)

Mark Nolan, Carnegie Mellon University

Dechying Ruengvisesh, UIUC

Namchul Shin, Pace University

Cheryl Thompson, UIUC

Susan Winter, University of Maryland

Ilya Zaslavsky, UCSD

With past contributions from:

Burcu Bolukbasi, University of Kansas

Noshir Contractor, Northwestern University

Cathy Davidson, City University of New York

Pablo Lopez, UIUC

Ethan Masella, Brandeis University

Melanie Radik, Brandeis University