

# The Cross-Domain Interoperability Framework (CDIF)

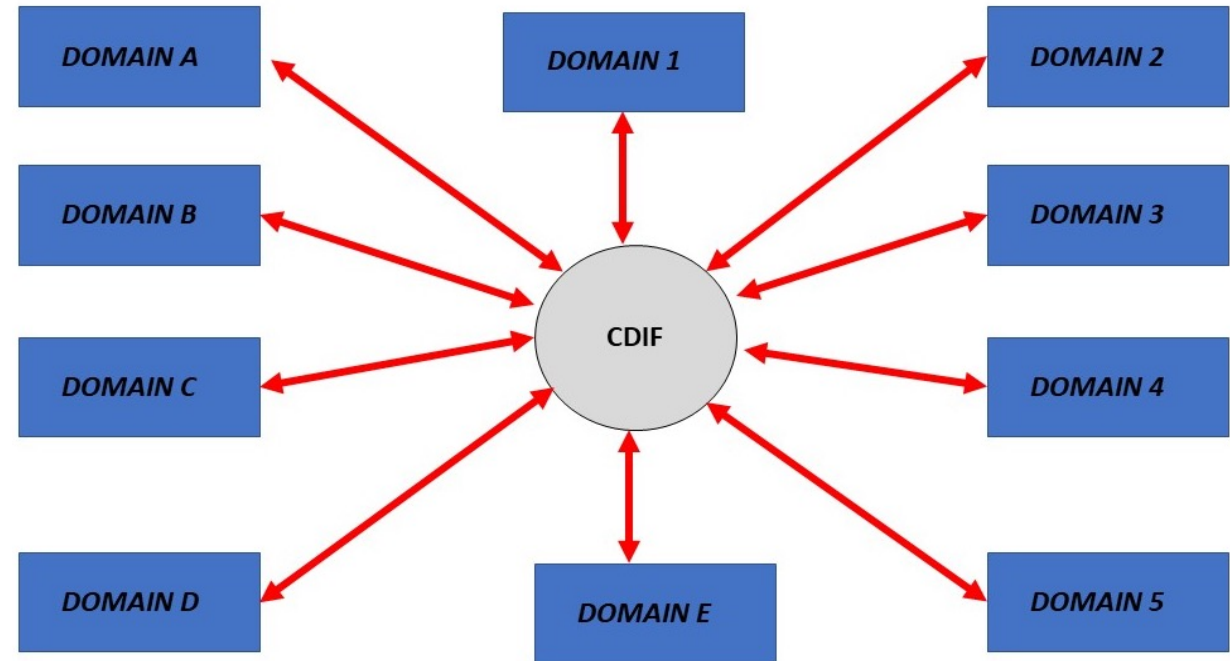
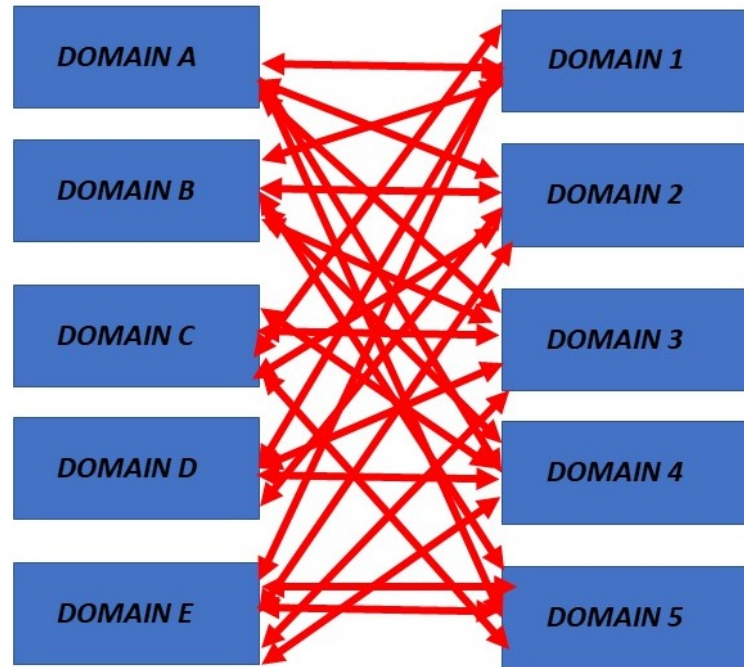
Arofan Gregory

CODATA

# Issues for Cross-Domain Data Use

- Domains have their own standards, models and technology approaches
- These require transformation for use in other domains
- Some metadata can be programmatically transformed – some cannot
  - Semantics/definitional metadata may require human judgement
  - Automation can help support these integration functions
- Cross-domain data use requires granular metadata and rich contextual information
  - To support automatic transformation between formats
  - To clarify how data is collected and processed

# Standards Mappings: Many-to-Many or Many-to-One?



**Note that some information does not require translation – consistent across domains (UoM, time, geography, species, chemicals, etc.)**

# WorldFAIR CDIF

- Addresses Recommendation 4 of *Turning FAIR into Reality*: “**Develop interoperability frameworks for FAIR sharing within disciplines and for interdisciplinary research**”
- CDIF (the “Cross-Domain Interoperability Framework”) is an emerging candidate for this type of effort:
  - Recommendations for the use of a coordinated set of existing standards and technologies for implementing FAIR across domain boundaries
  - It is a primary deliverable from WorldFAIR
  - Intended to be a first draft – CDIF will continue after the project if useful
- Inputs
  - WorldFAIR Case Studies
  - 30 invited experts in a Working Group and an Advisory Group
  - Other related activities (EOSC Future WP 6.3 SP9, EOSC Interoperability Framework, FAIR Impact, etc.)

# What CDIF Is and Is Not

- A set of guidelines for implementers of FAIR-enabled systems for the dissemination and consumption of data and metadata across domain and infrastructure boundaries
- A coordinated set of metadata profiles (and services) to support specific needed functions, based on existing specifications and technologies
- A “lingua franca” for interoperable FAIR implementation in a majority of cases, emphasizing machine-actionability
- A stepwise approach to FAIR implementation, with the lowest possible entry cost, based on current examples of good practice
- CDIF is *not* a new metadata specification (we have lots of those already!)
- CDIF is *not* a cool new technology approach for implementing FAIR (It relies on existing technology and approaches.)

# FAIR Activities and Standards

- Foundational: FAIR Digital Object Framework (**FDOF**)
- Find
  - Discover FAIR resources and explore/evaluate their utility prior to access (coverage, etc.)
  - **Schema.org, DCAT**
- Access
  - Negotiate access to non-public data
  - Enhance efficiencies through automation
  - **ODRL, DPV, DUO**
- Assess/Integrate
  - Understand data structure (**DDI-CDI**)
  - Understand semantics (**SKOS/XKOS, OWL, SSSOM**)
  - Determine origination/context (**PROV-O, I-ADOPT/O&M**)
- Establish connections to higher-level research and management processes
  - Track citation/use
  - Connect to ROI assessments
  - **CERIF, GAMSO**

Integrate climate data from Copernicus ERA5 and air quality data from the European Environmental Agency (EEA) with data from the European Social Survey (ESS)



Climate data

Social scientific survey

European Air Quality Index



**European  
Environment  
Agency**

# Cross-Domain Data Use

- Integrated data
  - ESS is primary data
  - Other sources are secondary and less familiar
  - Data integration at a scientific level requires cross-domain expertise, implemented as transformation/processing
  - Social scientists need the background about the integration of data
- CDIF would support easier exchange between domains to support this type of project
  - Data structures
  - Metadata/Semantics
  - Provenance and processing
  - Services/APIs
- CDIF helps lay the foundation for cross-domain science by maximizing the automated exchange of metadata across domain boundaries