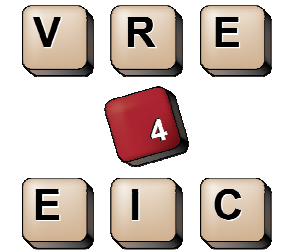A Europe-wide Interoperable Virtual Research Environment
to Empower Multidisciplinary Research Communities and Accelerate Innovation and Collaboration

VRE4EIC View

## Keith G Jeffery Scientific Coordinator ERCIM

Workshop

- Great benefit to VRE4EIC

- Great to listen to the very different experiences and solutions proposed

- Clear that there is much to be done in order to satisfy the various communities

- Hopefully provides a basis for discussion on a way forward to achieve this satisfaction

- VRE4EIC very open to such discussions
  - Already started a few here…….

Requirements and Rich Metadata Solutions
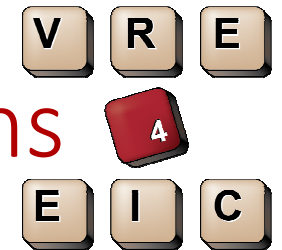
- Versioning (Markus Freudenberg, Simon Dutowski)

- Provenance (Markus Freudenberg, Simon Dutowski)

- Actionable rights (licences etc) (Markus Freudenberg)

- Multiple identifiers (Andrea Perego)

- Agent roles (Andrea Perego)

- Built-in versioning including different formats and related to purpose, persons/organisaions, software…

- Built-in provenance

- supports actionable rights

- Federated identifiers (with relationships between them depending on role and temporal duration)

- Agent roles in linking relations – note person and organisation as base entities because of different legal considerations (liability)

Requirements and Rich Metadata Solutions
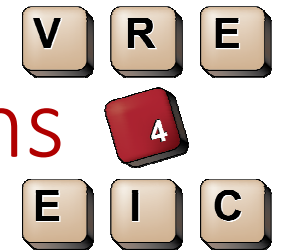
- Relationships between datasets and distributions (linked to versioning and provenance but need to understand e.g. software used) and collections (Makx Dekkers)
- Data citation (Makx Dekkers)

- Structured / detailed contact information (Makx Dekkers)

- Representing access via services/APIs/endpoints (Makx Dekkers)

- Poor geospatial representation (but see GEODCAT-AP) (Andrea Perego)

- Relationships in linking relations (role ad temporal semantics) indicating purpose

- Citation as linking relationship with role and temporal semantics (and linkage to additional information e.g. person, organisation, project…)

- Structured into postal address components and e-addresses – each with role (e.g. skype, email…)

- Services (APIs, endpoints) available for use and can be related to dataset (and software, persons, organisations…)

- Geospatial coordinates represented, and by classification/measurement encoding can represent granularity etc.

Requirements and Rich Metadata Solutions
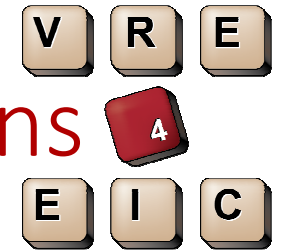
- Referential integrity (e.g. multiple organisations, multiple persons in various roles)

- Functional integrity – attributes not dependent on the primary key (unique ID of the asset)

- Leads to problems in
  - Metadata quality – cannot validate easily – consistency
  - Generating metadata – deduction and induction
  - Use of metadata autonomically for not only access to assets but also constructing workflows etc

- Referential integrity respected

- Functional integrity respected

- Problems not experienced
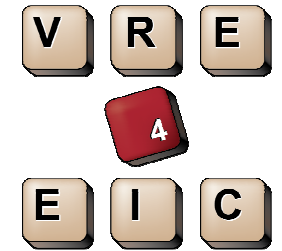
Requirements and Rich Metadata Solutions
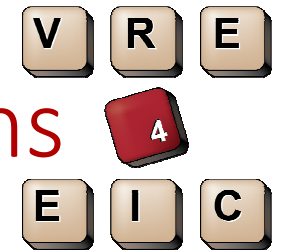
- Public service – needs metadata richer (relating persons, organisations, public services (Jeroem Baltussen)

- Scientific datasets – need much richer metadata than DCAT but domain-specific (Artemis Lavasa, Alejandra Gonzales-Beltram)

- Public service metadata elements covered (with rich linking role and temporal semantics)

- Rich metadata provides a general framework for more detailed scientific metadata using the concepts of classification (to distinguish domain and subdomains), concept of indicator and measurement for parameters

Requirements and Rich Metadata Solutions
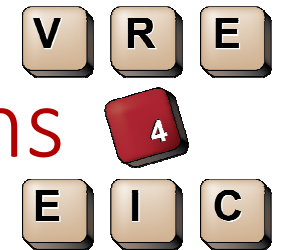
- And additionally:
  - easier to generate queries into the data instance (to get a preview / peek for example) or to obtain parameters for subset query
  - more material for machine learning and improvement;
  - can generate (from mapping) other metadata standards in the research domain and also including e.g. OGD and interoperation across metadata standards (DC, DCAT, CKAN (Prov-O))
  - LOD/SW is not the only way to link data – may need richer semantics
    - RDF needs many triples for temporal semantics and modal semantics
  - Need to back up OGD with relevant research/scientific data (hopefully from which the OGD was generated) – needs detailed link semantics

Conclusion

- We needed rich metadata for the VRE4EIC project
  - Having analysed requirements from many Ris
  - To interoperate many metadata standards
- This is why we chose to use CERIF
  - and why chosen for other projects
- CERIF also provides direct interoperability to the research information systems of various funding organisations and universities
  - Part of research reporting and assessment increasingly demanded by national funding organisations
  - http://www.eurocris.org/cerif/main-features-cerif

Acknowledgment

The VRE4EIC project has received funding
from the European Union's Horizon 2020 research and innovation
programme under grant agreement No 676247

http://www.vre4eic.eu/