

WDS Implementation Projects

Wim Hugo

Contributions from Michael Diepenbroek, Rorie Edmunds, Mustapha Mokrane, Yasuhiro Murayama

March 2014 | Updated and Extended November 2014

Draft for Discussion

# The Role of WDS in Global Research Infrastructure

WDS has to establish functioning infrastructure and services to support its objectives and its mandate. WDS has positioned itself to be a critical component of emerging Global Research Data Infrastructure (GRDI), specifically aiming to provide services that enable

1. Aggregated Meta-Data based on the collections offered by members,
2. A Knowledge Network that leverage information contained in the aggregated meta-data, as well as providing links to authoritative registries,
3. Maintain or support a registry of Trusted Digital Repositories, which includes the WDS membership as a sub-set. Central to such a registry would be a uniform and consistent quality assessment of data centers and their repositories, aligned with WDS, Data Seal of Approval, ISO 16363, and similar authorities or standards.

WDS needs to establish funding for these services, for the following reasons:

1. Components of a Global Research Data Infrastructure have been funded in the past as projects within research programmes, but these have a finite lifetime – typically 3-5 years, after which the services are often lost. WDS must be in a position to continue to operate these components as and when required, or establish it with a longer-term funding source;
2. More critically, at present there are few systems that are built on the services and components that WDS and similar organisations offer – but this is likely to change radically in years to come. The risk to systems developers and third-party investors is greatly increased by uncertainty about the future availability of foundational components and services.

This creates a need for a sustainable baseline operation within the WDS - in essence, guaranteeing the availability of registry services that will allow third parties to develop value-added services and applications. A similar need exists for complementary services and components supplied by other organisations.

# A Vision for Global Research Data Infrastructure

WDS is supporting a conceptual design, or vision, of interlinked foundational Global Research Infrastructure that is sustainable, scalable, and distributed amongst leading organisations, data centers, and initiatives. To achieve this, one has to identify the major actors (people, institutions, research outputs, and systems) that contribute to the conduct of science today, and then implement a minimum ***Knowledge Network*** that will maximize the return on investment for all (or most) of these actors.

A central characteristic of the Knowledge Network is that it should allow contributions from many sources, including, at least conceptually, from non-traditional sources such as project websites, social media, and citizen science.

WDS foresees that such a Knowledge Network will draw on the example of Linked Open Data, and re-use as many services, components, standards, and existing capacity as is possible. WDS has a role in cementing the missing relationships in the network, and in providing components (nodes) that are not yet available.

The Knowledge Network is shown schematically in Figure 1.

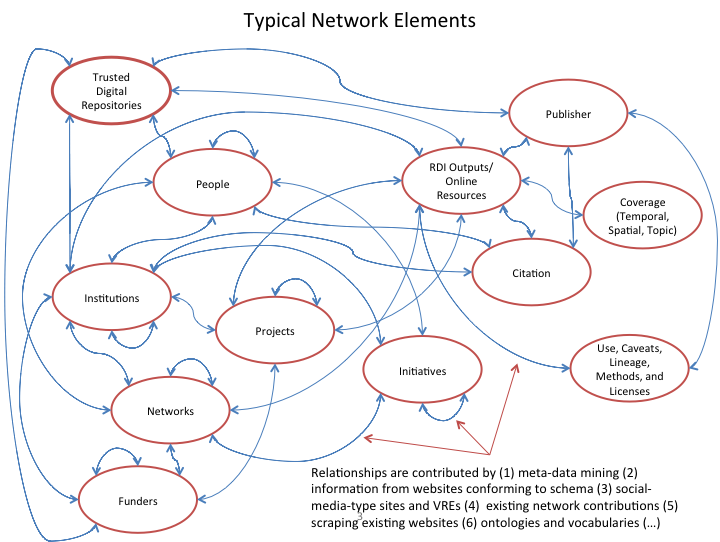


Figure 1: Elements of the Knowledge Network and Implied Relationships.

# Design Considerations

The Knowledge Network, in its implementation, must satisfy a number of considerations, some of which were mentioned already:

* **Scalability**: This is required in two ways:
  + Limiting the extent to which institutions, projects, or other initiatives need to process and commit content in bulk – by rather relying on *many, near-real time contributions* from all participants;
  + Allowing for a *distributed resource*, while avoiding fragmentation: At present, this distribution is a reality already, but it is not possible for humans or machines to access and exploit the network in a predictable, uniform way. Centralisation or a monolithic solution will not work, for a number of practical, political, and technical reasons.
* **Sustainability**: This applies to technical, governance, and financial elements of implementation.
  + Technical sustainability pivots on standards, scalability, and open, extensible software.
  + Sustainable governance should include community support and buy-in, quality assurance, and elements of trust.
  + Financial sustainability is a prerequisite, given the extent to which other investment and infrastructure will rely in future on the Knowledge Network. There is no apparent long-term funder for the entire network; hence, the practical solution must be a community consensus on the loose couplings between sustainably funded components.
* **Diversity**: A portfolio of initiatives, projects, and established interests all contribute to and/ or own parts of the network. There is no incentive to recreate these components, hence the only sensible approach is the maximum re-use within the Network.

## Twitter as an Example

Twitter is a perfect example of the usability that is required from the Knowledge Network. It must certainly be possible to contribute snippets of information to the network as easily as it is to tweet. The hash tags associated with content in TwitterSpace equate to DOIs for a number of actors, systems, and topics in the Knowledge Network – slightly more formalized, but probably not limited to them. The Knowledge Network has the added advantage of predetermined relationships between ‘hash tags’ – more about that below.

* **Ease of Use**: A significant hurdle to current use of research infrastructure and services is the weight of technical knowledge required: it should be easy and simple to contribute to and exploit the network. See box for some ideas.
* **Registries**: Any Knowledge Network will rely on essentially a ‘Registry of Registries’ – a sub-set of the Linked Open Data universe – with roles assigned to the contents of these registries.
* **Relationships**: While many institutions or initiatives contribute the elements of the Knowledge Network (the ‘nodes’), very few, if any, store the relationships between them explicitly.

# More About Relationships

A central aspect of the Knowledge Network, as depicted in Figure 1, is the relationships between the ‘nodes’ or elements of the network. As an example, ‘**People**’ can have the following standard relationships:

* To other **People**, by virtue of *Collaboration*, *Association*, etc,;
* To **Trusted Digital Repositories**, by *Depositing* **Research Outputs**;
* To their own **Research Outputs**, through *Authorship*;
* Indirectly to **Coverages**, by virtue of *Interest* in **Topics**, **Locations**, and **Time Periods**;
* By being *Hosted* or *Employed* by **Institutions**;
* and so on.

Many of these relationships are implicit in meta-data, but cannot easily be extracted by others. For example, the fact that two people are co-authors of a data set implies collaboration, but this cannot easily be determined from meta-data. The Knowledge Network mines meta-data, amongst other sources, for these relationships, but it can also be contributed from many other sources. At a minimum, simple interfaces should be available to commit meta-data, originating in a number of domain standards, to the Knowledge Network as and when it is created/ updated.

# Leveraging Existing Infrastructure

Several elements of the Knowledge Network exists already, for example DataCITE assisting with the allocation of DOIs to some research outputs (data specifically), ORCID assigning permanent identifiers to researchers, and so on. Figure 2 identifies these existing components – provisionally, at this point.

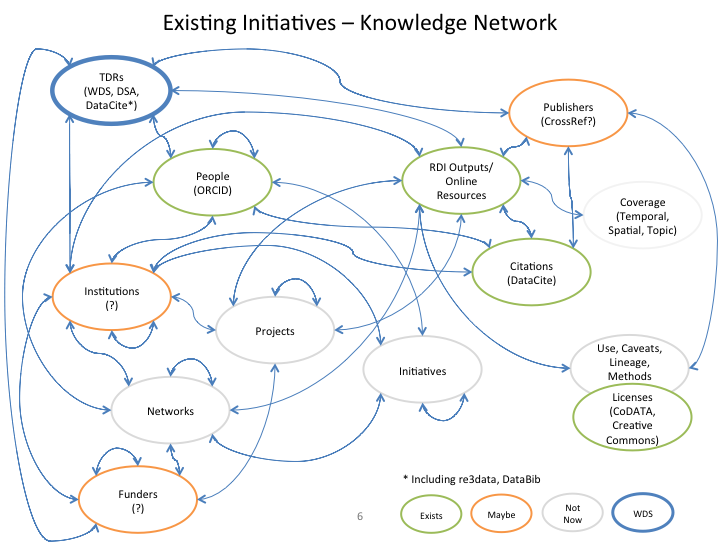


Figure 2: Assignment of Existing Initiatives to Knowledge Network Components

Specific actions that can form part of an initial phase of implementation include

1. Collaboration, with DataCite, on leveraging the merger between re3data, DataBib, and DataCite to establish a Registry of Trusted Digital Repositories. This is a sub-set of the registry envisaged by DataCite, or an independent extension of the registry that aggregates quality and certification properties (WDS, Data Seal of Approval, ISO 16363, etc.), managed by WDS.
2. Inclusion of ORCID as a major registry of permanent identifiers for researchers.
3. DOIs for citable data sets, as provided by DataCite.
4. Relationships, mined form WDS and possibly DataCite meta-data, supplemented by information from the WDS membership.

The Knowledge Network can grow in subsequent phases in three ways:

* Extending to other elements (topic coverages, institutions, publishers, funders).
* Inclusion of repositories that do not necessarily form part of the WDS membership.
* Contribution of relationships that do not originate from meta-data.

All of these are desirable extensions, and improve the utility of the Knowledge Network for users.

# Practical Implementation by WDS in the Short Term

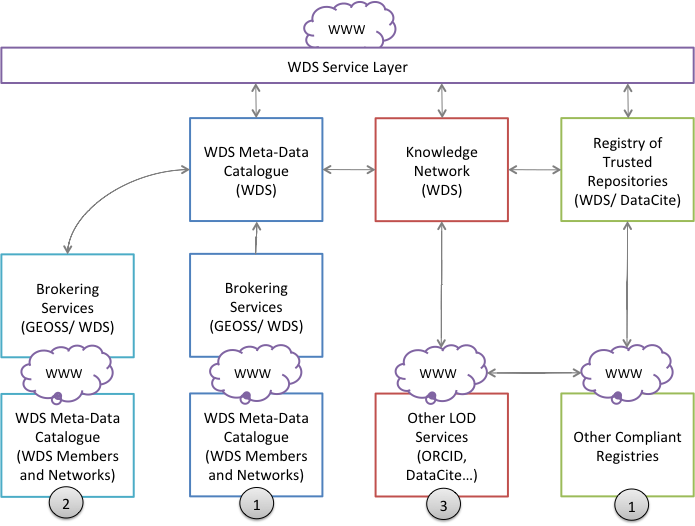
WDS will be pursuing the following goals in the next year:

1. Development of a service and standards specification – iteratively as part of an initial infrastructure establishment (see details below). This comprises a Meta-Data aggregator, Membership database, and KNwledge Network based on meta-data from membership.
2. Establishment of a registry of Trusted Digital Repositories in collaboration with DataCite (including re3data and DataBib), and adding elements denoting certification and quality criteria. These elements form part of an effort by WDS and Data Seal of Approval to align their certification procedures, and may in time extend to others – including ISO 16363.

Both of these goals require a two-pronged approach:

1. Practical measures to implement working infrastructure by engaging a ‘coalition of the willing and able’;
2. Referencing and peer review within a wider interest group or working group arrangement, all of which is either being pursued or in process of being pursued within RDA.

In summary, the WDS will be establishing an infrastructure that looks something like this:



Note the following:

1. WDS needs to establish two service infrastructures as a first priority:
   1. A meta-data harvesting arrangement, that draws partly on the GEOSS Broker, which assists with the synchronization and maintenance of a WDS aggregated meta-data catalogue. This catalogue spans a variety of protocols and meta-data standards, across multiple scientific disciplines.
   2. A registry of trusted repositories, probably in collaboration with an established service (re3data, DataCite). This registry, if it also offers a DOI for each repository, begins to form a critical part of the Linked Open Data services – not only used by our own Knowledge Network, but also by other services. ‘Trusted’ could be a metric or composite property, rather than a prerequisite.
2. The second priority is the establishment of a brokered access to data services referenced in meta-data. This is a major focus of GEOSS (for Earth and Environmental Observation data), and WDS will be able to apply this brokering service at least in part to provide access to data services. WDS will have to look for collaboration and funding opportunities to extend the capabilities of this service.
3. Third priority is to establish and operationalize our Knowledge Network. This can be done in two phases:
   1. A working implementation that is not directly based on WDS aggregated meta-data and repository registry, and aims to provide a proof of concept and integration test;
   2. An operationalisation phase, utilizing meta-data from the WDS aggregated repository and from the register of trusted repositories.

# Timelines and Releases

## Release 1

Release 1 implements a basic knowledge network that allows updates to the triple database from selected meta-data sources (Pangaea Dublin Core, and DataCite Meta-Data Schema), either automatically through a harvest mechanism, or manually by indicating individual files.

Triples are stored in a triple database to be selected by the developer to suit their skills, experience, and development environment. Requirements include

* Future scalability (possibly even cloud-based);
* Open source, in all probability;
* Java or Python-based, to align with the typical research infrastructure software;
* Support for SPARQL queries;
* Support for REST interfaces.

Possibly use a cloud-based service such as <http://docs.dydra.com/dydra> -

Queries are based on SPARQL, but not all end users will be able to construct their own SPARQL queries. For this reason, the query interface must also support pre-defined, named queries with limited parameterization.

Limited visualization needs to be implemented utilizing named queries. These examples highlight the application of the knowledge network in some of its typical use cases.

Release 1 will be completed in the first release cycle, ending in March 2015.

## Release 2

Release 2 will add the following functionality:

1. Linked Open Data – resolving as many named labels as possible to their permanent identifiers as possible.
2. Additional visualizations and named queries can be added.

Target for Release 2 is not yet defined.

# Annexure A: Estimates of Implementation Resources and Timelines

Estimates of development time required for establishment of initial WDS component infrastructure, together with potential collaborations and status of funding options:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **#** | **Infrastructure Component** | **Resource Estimate** | **Contributors** | **Proposals/ Resources** | **Risks** |
| 1 | Meta-Data Catalogue | 6 person-months | Pangaea  ESSI Lab  SAEON  NICT? |  | In-kind contributions  Dedicated time  Technology Transfer |
| 2 | Registry of Trusted Repositories | 4 person-months | Pangaea  IPO  SAEON  DataCite  Re3data |  | In-kind contributions  Dedicated time  DataCite Membership |
| 3 | Knowledge Network | 6 person-months | SAEON  NICT  JISC? | Microsoft? | In-kind contributions  Dedicated time  Technology Transfer |
| 4 | Data Brokering | 6 person-months | SAEON  ESSI Lab | SAEON/ESSI Lab | In-kind contributions  Dedicated time |

Recommendations/ Notes:

1. NICT has offered in-kind contribution of a systems developer – priority should be the Meta-Data Catalogue and Knowledge Network;
2. Formal implementation programme: systems engineering and project liaison on a bi-weekly cycle.
3. SC Members to assist with identification of funding opportunities and in-kind contributions.
4. These are relatively small integration projects and largely rely on existing applications and methods. The risk is not is R&D, but in technology transfer and integration testing. Hence the implementation team needs to be small.

These components will require a hosting organization each, or need to be hosted by the IPO. Depending on the decision in this respect, the development and delivery processes will be different.

Irrespective of the decision, the maintenance and support requirements have to included in the long-term costs. For third parties to become reliant on the infrastructure, this horizon should be in the order of 3-5 years.

Estimate of maintenance and support requirements: 1.5 resources per annum, available full time and part-time.

# Annexure B: Architecture, and Service and Content Specifications

## B.1 Architecture Notes

The Knowledge Network should be implemented as a set of services, capable of operating as RESTful or traditional web services, and using a generic and flexible content specification that allows a wide variety of knowledge objects to be stored.

The most appropriate candidate for the content backbone of the information architecture is RDF[[1]](#footnote-1). RDF allows the creation of relational triples describing almost any information, and is by design already a directed graph, forming the basis of a Knowledge Network.

A triple has a structure of subject-predicate-object. RDF allows two types of information to be stored generically because of flexibility in the role of the predicate and object:

|  |  |  |  |
| --- | --- | --- | --- |
| **Type of Information** | **Subject** | **Predicate** | **Object** |
| Relationship between Elements | URI (DOI, PID, …) | Named Relationship | URI |
| Property of an Element | URI | Property | String (Value) |

RDF can be serialized in a number of ways, but the most appropriate options remain XML and JSON.

The style of the services can be modeled on typical OGC examples – benefiting from wide familiarity within the Earth and Environmental Observation Community, and serves an additional purpose as a checklist for completeness of the Knowledge Network service specification.

## B.2 Content Standard

The Content Standard needs to be aligned with an object-relational model, largely derived from the schema in Figure 2. This implies that we have

* A collection of independent objects (people, institutions, topics, research outputs), some of which may be specializations (funder is a special case of institution);
* Properties allocated to the objects, such as the names of people or institutions, or synonyms for topics, derived from thesauri, controlled vocabularies, or ontologies. Some of these property schema are already well established in themselves (for example the ‘Friend of a Friend’ – FOAF – schema for describing people and their relationships), and some will have to be designed (very pertinently, the schema for description of Trusted Digital Repositories);
* Standardised values (‘lookup lists’) for some of these properties. This will be required for simplification – for example the academic qualification of a person, or the format of location coordinates for an institution;
* Critically important – the specification of the standard relationships that can be defined between objects.

Note that the triple store will have to persist additional (transactional) information with each triple – information such as the date of creation and modification, versioning, origin/ user account, and so on.

In addition, the triple store will have to allow for the resolution of objects and subjects to their linked open data permanent identifiers – these are almost always stored somewhere else in the web.

## B.3 Services

The services will fall into the following categories:

1. **Capabilities Statement**: this is equivalent to the typical ‘GetCapabilities’ request in OGC standards, and conveys information about the capabilities and characteristics of a Knowledge Network registry. The Knowledge Network can be distributed, and not all registries will have the same capabilities or scope.
2. **Content Request**: retrieving information, offered as serialized directed graph (RDF triples), with a subset of objects, their properties, and relationships. A query language is required to convey the content request, and RDF triple stores already have standardized options in this regard – for example SPARQL. One of the challenges is the relatively poor penetration of SPARQL amongst non-IT professionals, hence we foresee a need to define a number of **Named Content Requests** that are listed as part of a Capabilities Statement – allowing non-technical users to request content in a simplified way. The Named Content Request retrieves a predefined SPARQL query on the service endpoint, and substitutes it on behalf of the requestor.
3. **Transactional**: Any registry can allow additions and corrections to its triple store, and one or more services are required to accomplish this.

The content request will require additional parameters to specify the output format or mime type. It is conceivable that in addition to RDF serialisations, users may also want the content rendered as comma-delimited text, network graph images, or any number of similar derivations.

## B.4 Client Applications and Interfaces

One of the major impediments to widespread use of RDF and its query languages is technical: it is not intuitively useful to non-IT professionals. Hence both the query facilities (above) and the rendering of results should to some extent be pre-packaged for the wider users community.

SAEON has started experimenting with some options for rendering, and we propose that some standard renderings be offered, either as embeddable scripts, or as rendered images that are requested on refresh – in this way, operation is quite similar to the widely implemented OGC Web Map Services (WMS) standard for geospatial data rendering.

We have compiled a list of examples, and these are presented in the table. The examples are based on the InfoVis library, open source software developed by Nicolas Garcia Belmonte[[2]](#footnote-2), but other options are also available.

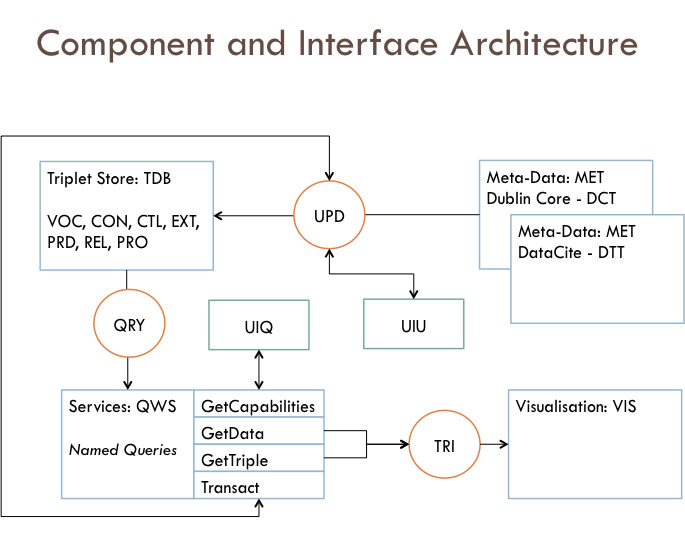
|  |  |  |  |
| --- | --- | --- | --- |
| **Style** | **Example** | **Application** | **URL** |
| TreeMap |  | Relative Importance of a Property or Count of Named Relationships for a filtered sub-set of objects. Example: Compare institutions based on dataset DOI count. | <http://adca.dirisa.org/test/treemap-example> |
| Sunburst |  | Relationships and weight of relationship between filtered subset of objects, as well as a property of the object. Example: Number of Collaborations between institutions, with Institutional author count. | <http://adca.dirisa.org/test/copy_of_treemap-example> |
| SpaceTree |  | Relationships, especially a hierarchical or tree structure, as a filtered subset of objects. Each node can represent an additional property as a colour. Navigable – clicking a node requests additional information and makes the node the root. | <http://adca.dirisa.org/test/copy2_of_treemap-example> |
| HyperTree |  | A hierarchical or tree structure, as a filtered subset of objects. Each node can represent an additional property as a node size, and an additional property as a colour. Relationship weight can be represented as a line thickness or a colour. Navigable – clicking a node requests additional information and makes the node the root. | <http://adca.dirisa.org/test/copy3_of_treemap-example> |
| Animated Pie Chart |  | A filtered subset of nodes, representing a property or relationship count as slice size, and categorized in any number of classes. | <http://adca.dirisa.org/test/copy4_of_treemap-example> |
| Composite Tree |  | Combines tree and pie chart views, and can represent a large number of properties and relationships. Navigable - clicking a node issues new requests and makes the node the root of the view. Can represent a classified property of a node, as well as relationships and weight of relationships to other nodes. | <http://adca.dirisa.org/test/copy5_of_treemap-example> |
| Proportional Hierarchy |  | Represents a tree view, as a filtered subset of nodes. Each node can be coloured and sized, representing two properties or relationship counts, and the tree is navigable – clicking a node makes the node the root. | <http://adca.dirisa.org/test/copy6_of_treemap-example> |
| Large Networks |  | Simplified view of large networks, with limited property or relationship count information to reduce complexity. Navigable, and useful for exploring the Knowledge Network. | <http://adca.dirisa.org/test/copy7_of_treemap-example> |

# Annexure C: Unified Approach to Certification – Maturity of Digital Repositories

To be finalized at RDA 4 with Lesley and Ingrid Dillo ?

# Annexure D: Detailed Specifications

## Concept Diagram



## CAS: Components, Actors, Systems

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **ID** | **Description** | **Reference/ Specification** |
| CAS-01 | Acquisition Service | A service to accept a triple in an agreed schema for inclusion into the Triple Store (TDB) | TAS |
| CAS-02 | Triple Store | A Triple Store (Triple Database) storing a directed graph | TDB |
| CAS-03 | Meta-Data Record | Sources of triples from meta-data | MET |
| CAS-04 | Queries | Querying services for the Triple Store | QRY |

## TDB: Triple Database

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **ID** | **Description** | **Reference/ Specification** |
| TDB-01 | Schema | Schema needs to include basic triple elements: object, predicate, subject, properties |  |
| TDB-02 | Standard Values | Objects/ Subjects, Predicates, Properties have standardized lookup vocabularies that are determined by WDS | [VOC](#_VOC:_Vocabularies) |
| TDB-03 | Services | Services are required for updating the Triple Store and for querying the triple store | [UPD](#_UPD:_Update_Process)  [QRY](#_QRY:_Query_Process) |
| TDB-04 | Extended Schema | Each object/ subject has additional properties in the triple store relating to its permanent identifiers in the web. This does not replace, but supplements the original object/ subject label. | [EXT](#_EXT:_Extended_Schema) |
| TBD-05 | Control Schema | This contains information pertaining to management and control of the triple entry | [CTL](#_CTL:_Control_and) |
| TBD-06 | Technology | Many options - <http://en.wikipedia.org/wiki/Triplestore> |  |

## UPD: Update Process

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **ID** | **Description** | **Reference/ Specification** |
| UPD-01 | Parsing MetaData | Meta-data needs to be parsed into triples, using the appropriate vocabularies, based on supported meta-data schema. | [MET](#_MET:_Meta-Data_Schema) |

## QRY: Query Process

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **ID** | **Description** | **Reference/ Specification** |
| QRY-01 | Query Language | SPARQL is a typical and preferred query language interface for a Triple Store. |  |
| QRY-02 | Named Queries | There are named queries that are maintained as service parameters in the query component and available to external actors (users, machines) as web services | [QWS](#_QWS:_Query_Web) |

## MET: Meta-Data Schema

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **ID** | **Description** | **Reference/ Specification** |
| MET-01 | Dublin Core | Most data sets harvested by WDS directly will be presented as Dublin Core. A mapping is provided between Dublin Core and the Triple Schema. | [DCT](#_DCT:_Dublin_Core) |
| MET-02 | DataCite | Some WDS members will provide meta-data via DataCite. A mapping is provided between DataCite Schema and Triple Store. | [DTT](#_DTT:_DataCite_to) |

## VOC: Vocabularies

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **ID** | **Description** | **Reference/ Specification** |
| VOC-01 | Object/ Subject | Objects and Subjects are derived from the WDS Conceptual Model for scope of research activity and outputs. | [CON](#_CON:_Conceptual_Model) |
| VOC-02 | Predicates | Predicates define standard relationships between objects and subjects | [PRD](#_PRD:_Predicates_1)  [REL](#_REL:_Relationships_Between) |
| VOC-03 | Properties | Each specific object has a standardized property structure | [PRO](#_PRO:_Properties_1) |

## CON: Conceptual Model - Vocabulary

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **ID** | **Description** | **Reference/ Specification** |
| CON-01 | TDR | Trusted Digital Repository | [PRO](#_PRO:_Properties_2) |
| CON-02 | PPL | People: researchers, data centre managers, meta-data authors, research output authors |
| CON-03 | RDO | Research and Development outputs: data sets, journal articles, code, reports, … |
| CON-04 | PUB | Publishers – open access or traditional data and article journals |
| CON-05 | COV | Coverage: Spatial, Temporal, and Topic coverages |
| CON-06 | CIT | A citation object – either a text citation, handle, persistent identifier, or DOI |
| CON-07 | IST | Institutions – research facilities, universities, academies, … |
| CON-08 | PRJ | Projects and programmes: grant-funded projects that relate multiple institutions, funders, and people. |
| CON-09 | NET | Networks: associations and collaborations between people, institutions, and funders |
| CON-10 | FUN | Funders of grants and research programmes. |
| CON-11 | INI | Initiatives: Similar to PRJ, possibly, but not always funded |
| CON-12 | USE | A usage object, containing data such as lineage, protocols, license, caveats, and more. |

## CTL: Control and Management Schema

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **ID** | **Description** | **Reference/ Specification** |
| CTL-01 | Accession Data | Information on   1. Accession data 2. Source 3. IP |  |
| CTL-02 | Modification Data | Information on   1. Modification Date 2. Source 3. Type of Change 4. IP |  |
| CTL-03 | States | Triple States   1. Public – no QA 2. Public – QA 3. Marked as Private |  |

## EXT: Extended Schema Identifier Lookups

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **ID** | **Description** | **Reference/ Specification** |
| CON-01 | TDR | Trusted Digital Repository: WDS to establish its own registry in collaboration with <http://service.re3data.org/> - API is not yet available but we have requested one. |  |
| CON-02 | PPL | Researchers: ORCID | ORCID |
| CON-03 | RDO | Research Output Types: WDS to establish its own registry |  |
| CON-04 | PUB | Publishers: URI |  |
| CON-05 | COV | Spatial: EPSG 4326/ GeoNames  Temporal: To be determined  Topic: To be determined – multiple vocabularies and ontologies | GeoNames |
| CON-06 | CIT | DataCite: DOIs | DataCite |
| CON-07 | IST | Institutions: To be determined |  |
| CON-08 | PRJ | Projects: To be determined |  |
| CON-09 | NET | Networks: URI |  |
| CON-10 | FUN | Funders: URI |  |
| CON-11 | INI | Initiatives: URI |  |
| CON-12 | USE | Licenses: Creative Commons  Protocols/ Methodology: URI | Creative Commons |

## PRD: Predicates

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **ID** | **Description** | **Reference/ Specification** |
| PRD-01 | Default | The default predicate is ‘Is Linked To’ |  |
| PRD-02 | Specific predicates | The list of specific predicates is shown below. |  |

|  |
| --- |
| List of applicable predicates: |
| collaborates with |
| covered by |
| covers |
| deposit in |
| funded by |
| funds |
| hosts |
| involved in |
| is deposited in |
| is described by |
| Is harvested by |
| is interested in |
| is linked to |
| is used by |
| links to |
| member of |
| operated by |
| operates this |
| participate in |
| points to |
| produced at |
| produced by |
| produced this |
| published from |
| publishes in |
| refers to |
| supported by |
| supports |
| used by |
| uses |
| were produced by |

## REL: Relationships Between Objects/ Subjects and Predicates

See table below (in two sections)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | TDR | PPL | RDO | PUB | COV | CIT |
| TDR | Is harvested by | is used by | hosts this | links to | - | is linked to |
| PPL | deposit in | collaborates with | produced this | uses | is interested in | is linked to |
| RDO | is deposited in | were produced by | is linked to | published in | covers | is described by |
| PUB | links to | uses | refers to |  | covers | is linked to |
| COV | - | is linked to | covered by | is linked to | is linked to | is linked to |
| CIT | links to | is linked to | applies to | points to | covers | is linked to |
| IST | operates this | hosts | produced this | publishes in | covers | is linked to |
| PRJ | is linked to | is linked to | produced this | publishes in | covers | is linked to |
| NET | is linked to | is linked to | produced this | publishes in | covers | is linked to |
| FUN | is linked to | is linked to | funds | is linked to | covers | is linked to |
| INI | is linked to | is linked to | produced this | publishes in | covers | is linked to |
| USE | is linked to | is linked to | links to | is linked to | covers | is linked to |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | IST | PRJ | NET | FUN | INI | USE |
| TDR | operated by | supports | supports | supported by | supports | is linked to |
| PPL | hosted by | participate in | participate in | funded by | participate in | is linked to |
| RDO | produced at | produced by | produced by | funded by | produced by | is linked to |
| PUB | published from | used by | used by | used by | used by | is linked to |
| COV | covered by | covered by | covered by | covered by | covered by | linked to |
| CIT | is linked to | produced by | produced by | funded by | produced by | is linked to |
| IST | is linked to | involved in | member of | funded by | member of | is linked to |
| PRJ | is linked to | is linked to | is linked to | funded by | is linked to | is linked to |
| NET | is linked to | is linked to | is linked to | funded by | is linked to | is linked to |
| FUN | is linked to | funds | funds | is linked to | funds | is linked to |
| INI | is linked to | is linked to | is linked to | funded by | is linked to | is linked to |
| USE | is linked to | is linked to | is linked to | is linked to | is linked to | is linked to |

## DCT: Dublin Core to Triple Store

TBC: WH to provide as soon as possible.

## DTT: DataCite to Triple Store

TBC: WH to provide as soon as possible.

## QWS: Query Web Services

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **ID** | **Description** | **Reference/ Specification** |
| QWS-01 | GetCapabilities | Obtain service information including a list of named services. Model on OGC GetCapabilities statement or OAI |  |
| QWS-02 | GetData | Invokes a named service or a SPARQL Query as a service parameter. Output is RDF or RDFa, or JSON | [NSV](#_NVS:_Named_Services) |
| QWS-03 | GetTriple | Obtain a triple as RDF, RDFa or JSON, parameter is a triple UID |  |
| QWS-04 | Transact | Update, delete, change state of triple, parameter is a triple UID and input is a triple expressed as RDF, RDFa, or JSON. |  |

## PRO: Properties

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **ID** | **Description** | **Reference/ Specification** |
| PRO-01 | TDR-P | Name  WDS Accreditation  DSA Accreditation  ISO Accreditation  WDS URL (if WDS Member) |  |
| PRO-02 | PPL-P | Name |  |
| PRO-03 | RDO-P | Title  Type (of research output)  URL (to meta-data)  URL (to resource service or download point)  MIME (MIME type of resource endpoint) | ROT |
| PRO-04 | PUB-P | Name  Code (from authoritative list)  Website URL |  |
| PRO-05 | COV-P | Spatial Coverage: Bounding Box, Point, Geographical descriptor or label  Temporal coverage: Start Date and End Date  Topic Coverage: One or more keywords |  |
| PRO-06 | CIT-P | Citation Text |  |
| PRO-07 | IST-P | Name |  |
| PRO-08 | PRJ-P | Name |  |
| PRO-09 | NET-P | Name |  |
| PRO-10 | FUN-P | Name |  |
| CON-11 | INI-P | Name |  |
| CON-12 | USE-P | License: Text Value |  |

In addition, each item usually has generic properties:

1. PID, URI or Website URL

## TRI: Triple Package

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **ID** | **Description** | **Reference/ Specification** |
| TRI-01 | Triple Package | Contains information on one or more triples retrieved through query. |  |
| TRI-02 | Supported Formats | Supported formats –  RDF  RDFa  JSON |  |
| TRI-03 | Aggregate Format | Format: Unique Combinations of  *Object | Subject | Predicate*  Filtered by  *Date | Any Subject/ Object | Any Predicate | Any Status*  Additional Values  *Count of Entries | Average of Entries* |  |

## UIQ: User Interface - Queries

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **ID** | **Description** | **Reference/ Specification** |
| UIQ-01 | User Interface | User interface to assist web-based formulation of SPARQL queries. Users to select aggregation and filtering from available objects, subjects, and predicates |  |

## UIU: User Interface - Updates

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **ID** | **Description** | **Reference/ Specification** |
| UIU-01 | User Interface | User interface to assist with web-based processing of one or more meta-data files (local, or web-addressable XML file/ service). |  |

## NVS: Named Services (Queries)

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **ID** | **Description** | **Reference/ Specification** |
| NVS-01 | Lists | Invokes GetData Service  ?verb=GetList  &entity=’TDR|PPL|…’  **Returns**:  List of entities matching criteria | CON |
| NVS-02 | Relationships | Invokes GetData Service  ?verb=GetRelationships  &object=<Some label or PID>  **Returns**:  List of relationships and counts matching criteria |  |
| NVS-03 | Relations | Invokes GetData Service  ?verb=GetRelations  &object=<Some label or PID>  &predicate=’collaborates with | covered by | …’  **Returns**:  List of relationships, subjects, and counts matching criteria |  |

1. Or one of its derivatives – RDFa, etc. [↑](#footnote-ref-1)
2. <http://philogb.github.io/jit/> [↑](#footnote-ref-2)