**NBD(NIST Big Data) Requirements WG Use Case Template Aug 11 2013**

|  |  |  |  |
| --- | --- | --- | --- |
| **Use Case Title** | | The ‘Discinnet process’, metadata <-> big data global experiment | |
| **Vertical (area)** | | Multidisciplinary / Natural sciences | |
| **Author/Company/Email** | | P. Journeau / Discinnet Labs / [phjourneau@discinnet.org](mailto:phjourneau@discinnet.org) | |
| **Actors/Stakeholders and their roles and responsibilities** | | Actors Richeact, Discinnet Labs and I4 OpenResearch fund France/Europe. American equivalent pending. Richeact is fundamental R&D epistemology, Discinnet Labs applied in web 2.0 [www.discinnet.org](http://www.discinnet.org), I4 non-profit warrant. | |
| **Goals** | | Richeact scientific goal is to reach predictive interdisciplinary model of research fields’ behavior (with related meta-grammar). Experimentation through global sharing of now multidisciplinary, later interdisciplinary Discinnet process/web mapping and new scientific collaborative communication and publication system. Expected sharp impact to reducing uncertainty and time between theoretical, applied, technology R&D steps. | |
| **Use Case Description** | | Currently 35 clusters started, close to 100 awaiting more resources and potentially much more open for creation, administration and animation by research communities. Examples range from optics, cosmology, materials, microalgae, health to applied maths, computation, rubber and other chemical products/issues.  How does a typical case currently work:   * A researcher or group wants to see how a research field is faring and in a minute defines the field on Discinnet as a ‘cluster’ * Then it takes another 5 to 10 mn to parameter the first/main dimensions, mainly measurement units and categories, but possibly later on some variable limited time for more dimensions * Cluster then may be filled either by doctoral students or reviewing researchers and/or communities/researchers for projects/progress   Already significant value but now needs to be disseminated and advertised although maximal value to come from interdisciplinary/projective next version. Value is to detect quickly a paper/project of interest for its results and next step is trajectory of the field under types of interactions from diverse levels of oracles (subjects/objects) + from interdisciplinary context. | |
| **Current**  **Solutions** | **Compute(System)** | | Currently on OVH servers (mix shared + dedicated) |
| **Storage** | | OVH |
| **Networking** | | To be implemented with desired integration with others |
| **Software** | | Current version with Symfony-PHP, Linux, MySQL |
| **Big Data  Characteristics** | **Data Source (distributed/centralized)** | | Currently centralized, soon distributed per country and even per hosting institution interested by own platform |
| **Volume (size)** | | Not significant : this is a metadata base, not big data |
| **Velocity**  **(e.g. real time)** | | Real time |
| **Variety**  **(multiple datasets, mashup)** | | Link to Big data still to be established in a Meta<->Big relationship not yet implemented (with experimental databases and already 1st level related metadata) |
| **Variability (rate of change)** | | Currently Real time, for further multiple locations and distributed architectures, periodic (such as nightly) |
| **Big Data Science (collection, curation,**  **analysis,**  **action)** | **Veracity (Robustness Issues, semantics)** | | Methods to detect overall consistency, holes, errors, misstatements, known but mostly to be implemented |
| **Visualization** | | Multidimensional (hypercube) |
| **Data Quality (syntax)** | | A priori correct (directly human captured) with sets of checking + evaluation processes partly implemented |
| **Data Types** | | ‘cluster displays’ (image), vectors, categories, PDFs |
| **Data Analytics** | |  |
| **Big Data Specific Challenges (Gaps)** | | Our goal is to contribute to Big 2 Metadata challenge by systematic reconciling between metadata from many complexity levels with ongoing input from researchers from ongoing research process.  Current relationship with Richeact is to reach the interdisciplinary model, using meta-grammar itself to be experimented and its extent fully proven to bridge efficiently the gap between as remote complexity levels as semantic and most elementary (big) signals. Example with cosmological models versus many levels of intermediary models (particles, gases, galactic, nuclear, geometries). Others with computational versus semantic levels. | |
| **Big Data Specific Challenges in Mobility** | | Appropriate graphic interface power | |
| **Security & Privacy**  **Requirements** | | Several levels already available and others planned, up to physical access keys and isolated servers. Optional anonymity, usual protected exchanges | |
| **Highlight issues for generalizing this use case (e.g. for ref. architecture)** | | Through 2011-2013, we have shown on [www.discinnet.org](http://www.discinnet.org) that all kinds of research fields could easily get into Discinnet type of mapping, yet developing and filling a cluster requires time and/or dedicated workers. | |
| **More Information (URLs)** | | On [www.discinnet.org](http://www.discinnet.org) the already started or starting clusters can be watched in one click on ‘cluster’ (field) title and even more detail is available through free registration (more resource available when registering as researcher (publications) or pending (doctoral student)  Maximum level of detail is free for contributing researchers in order to protect communities but available to external observers for symbolic fee: all suggestions for improvements and better sharing welcome.  We are particularly open to provide and support experimental appropriation by doctoral schools to build and study the past and future behavior of clusters in Earth sciences, Cosmology, Water, Health, Computation, Energy/Batteries, Climate models, Space, etc.. | |
| **Note:** <additional comments>: We are open to facilitate wide appropriation of both global, regional and local versions of the platform (for instance by research institutions, publishers, networks with desirable maximal data sharing for the greatest benefit of advancement of science. | | | |

**Note: No proprietary or confidential information should be included**