Weaving a Web of data around Linux distributions

March 2010

slauriere@mandriva.com

« Open source software is creating a global software space, with dependencies between projects, is meshing software from many different sources. But we are not meshing the data about the software! »

Henry Story
http://bblfish.net/work/presentations/2007/BOF-6747.pdf

« We can create secure open APIs between Bugzillas "porous federated containers" »

Mark Shuttleworth
http://www.linux-watch.com/news/NS8470376604.html

- Decrease the engineering costs at stake for producing and maintaining a Linux distribution
- Assess accurately the quality of the packages included in Mandriva Linux
- Provide a semantic search engine for users and developers to obtain answers to domain specific problems
- Design and implement a global quality model for OSS covering static code analysis, community analysis, QA metrics, integration tests analysis etc.
- Ease collaboration with upstream projects and with other Linux distributions
- Draw a distributed social network across OSS contributors

- Availability of large amount of structured data and unstructured data covering OSS
- Some existing tools for mining OSS data
- Availability of high performance RDF databases
- Open source NLP frameworks
- Nepomuk KDE semantic desktop
- Ongoing Mandriva research projects related to linked data for software engineering:
    - Helios http://www.helios-platform.org
    - Scribo http://www.scribo.ws
    - Qualipso http://www.qualipso.org
    - Wiki 3.0 http://entrepreneurs.telecom-paristech.fr/revue-de-presse/appel

- Doc4 http://doc4.mandriva.org
    - Data model similar to the Universal Debian Database
    - Support for collaborative semantic annotations
    - Wiki based enhancements
    - Serves RDF/XML representation of packages and related metrics. Next: standardize the package data around a *Unified Package Metadata Initiative*.
- Nepomuk http://nepomuk.kde.org
    - Local metadata database
    - Connection of the Doc4 repository to development tools ("Mylyn for KDE") and to the personal information system
- WebReformatterhttp://code.google.com/p/webreformatter/
    - Homogenous view on heterogeneous data
    - Data enhancement, data integration
    - Easy collaboration through OpenSocial widgets
    - Custom index of domain specific data

- Fetchbugs4me http://blog.fetchbugs4.me/
- Launchpad https://help.launchpad.net/API
- Debian UDD http://udd.debian.org/
- DERI LD2SD http://sw-app.org/pub/seke09-ld2sd.pdf