

PROJECT TITLE & ACRONYM?

Background

Increasingly Universities throughout Europe are using supporting digital material in the educational process for students, professionals and peers. This ranges from (digital) text books, syllabi, power point presentations, students' posters sessions to streaming video-lectures. The process of educational content creation however, can be considered sub-optimal as many content developers (publishers, lecturers etc.) in different European countries often "re-invent the wheel", developing similar presentations, lectures and syllabi within the same domain. A large existing digital content base is already available and growing throughout Europe, at different educational institutes, in different formats and languages. They can not be (re) used to increase efficiency and quality of digital educational content development for several reasons:

1. Lack of visibility of available content
2. Lack of search ability due to fragmentation of digital content (different institutes, formats etc.)
3. Barriers to access due to multilingualism of content
4. Missing access services to digital materials with assured quality levels

In the scientific discipline of Genetics, many academic institutes throughout Europe are confronted with this problem of fragmented and inaccessible digital content. A substantial amount of digital educational content is available (e.g. the European Genetics Foundation owns more than 280 lectures in streaming video, 120 digital text books, 280 abstracts etc.) and digital content development for education is proliferating. At the same time a culture of sharing knowledge and materials is emerging amongst academic partners in order to improve the process of development of supportive educational materials based on existing high quality materials, potentially speeding-up the efficiency and effectiveness of the process.

Objectives

Against this background the need for a European approach that facilitates multilingual access to and re-use of digital educational content on genetics between distributed (academic) institutes is arising. This project aims to provide a response to this need by improving the quality of digital content through the addition of metadata and demonstrating the added value of a common European approach to enrich and share digital content in a multilingual context. These objectives will be reached by:

1. Carrying out a user needs requirement and identification of available digital educational content in the field of Genetics
2. Developing a European metadata annotation procedure using state of the art tools and techniques
3. Setting-up and applying tools to enable automatic metadata annotation of streaming video
4. Develop a multilingual ontology with conceptual domain descriptions in order to facilitate multilingual cross-border search in distributed enriched content bases
5. Adding metadata (indexes)
6. Carry out a set of real life experiments with students, medical doctors (Oncologists) and lecturers to demonstrate ease of access and enhanced development of digital educational content in the domain of Genetics (e.g. by lecturers)
7. Developing and setting-up a sustainability model with peer-review for quality control of materials to be shared including a business model

The final result will be the enrichment of a critical mass of existing data sets at different institutes applying a metadata annotation procedure and state of the art tools which can be used for the structured development of metadata annotated and reusable digital educational content. Towards the future this will set the basis for a fast growing set of reusable educational content enhancing substantially its development.

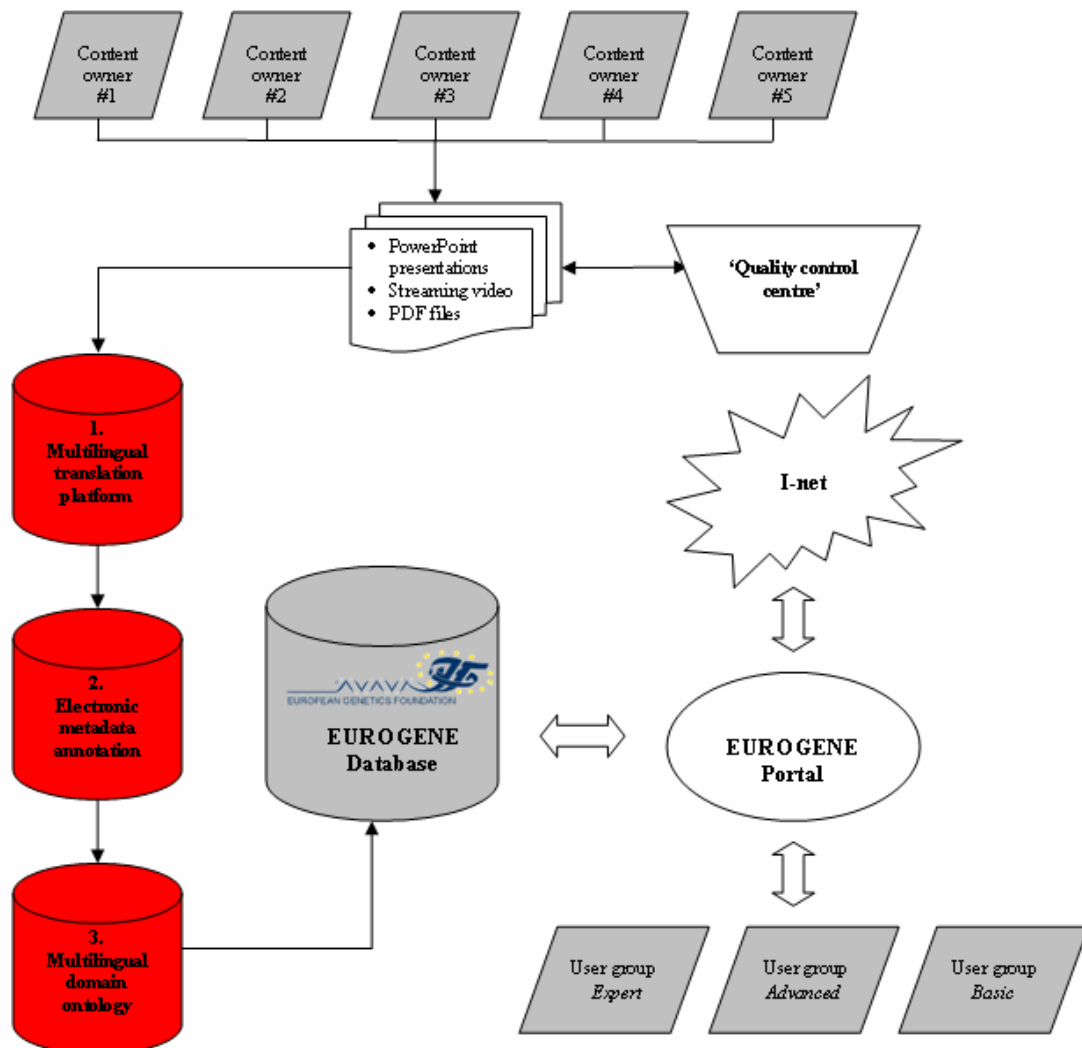


Figure 1: overview of the EUROGENE service structure

Figure 1 shows an overview of the targeted integration of data gathered from content owners. A new portal will be launched specifically to enable access to the database, as well as it being available **through each of the partner sites**. The targeted structure consists of four layers:

1. All content on genetics from local databases available on a single site

All digital educational content on genetics in the current local databases will be available and interoperable in the new database structure. The database will contain digital content on genetics from academic institutes throughout Europe. A new portal will allow users to access this centralized database based on their profile; *expert, advanced or basic domain*

2. 'Quality control centre'

In order to certify the quality of information to be shared, a peer-review is conducted. After this review, raw data is pushed to the next step in the procedure; enrichment. Peer-review is done by young graduates. **Other suggestions?**

3. Enrichment of content on genetics to facilitate data mining

Existing data sets at different institutes will be enriched applying a metadata annotation procedure and state of the art tools which can be used for the structured development of metadata annotated and re-usable digital educational content. Content will be enriched using a multilingual translation platform, electronic metadata annotation and multilingual domain ontology.

4. Portal to access the EUROGENE database

Based on a users profile access is granted to the EUROGENE database content. User groups are divided into three different layers; expert, advanced and basic domain. This profile is a result of a questionnaire that has to be filled in by the user before access to the database is granted.

| | |
|-----------|--|
| Expert: | Lecturers, scientists, medical specialists |
| Advanced: | Medical associations, universities, health foundations representing patients |
| Basic: | Public domain users |

What other user groups should be identified... please add

Besides information from the EUROGENE database, users can search the Internet using a tailored solution provided by MSNsearch (Microsoft) or Google. Search engines will be adapted to specific needs of users searching for educational content on genetics. Non relevant hits will be limited to a minimum.

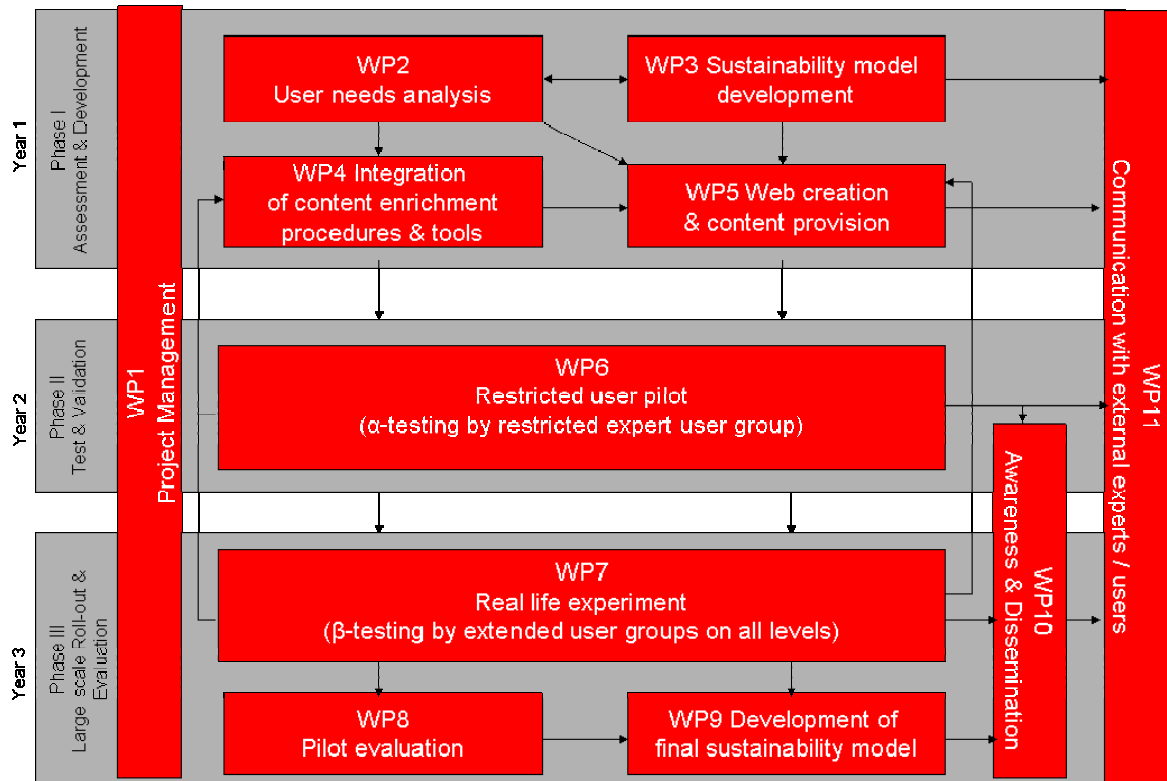


Figure 2: Workplan overview

This content enrichment project will cover a 3 year time period and can be divided in 3 phases:

Phase 1 Assessment & development (m1 – m15)

WP2: User needs analysis (m1 – m4)

- Analyse available information sources of genetics educational content (competing services)
- Analyse user needs in different user groups; what is it they want/ need?
 - o colleague lecturers on genetics (expert access)
 - o students and other academic users (advanced access)
 - o public domain users, such as schools, public awareness campaigns (basic access)
- Functional description: include best available technologies and features based on user needs.

WP3: Sustainability model development (m3 – m6)

- Modelling service concept (provision and use of content, maintenance of database, keeping information up-to-date), basis for future SLA (service level agreement).
- Inventory of potential user groups for service ("sales" forecasting) and exploitation channels (service roll-out to keep it sustainable after 2 years).
- Write initial sustainability plan (to be complemented/ updated during project execution).

WP4: Integration of content enrichment procedures and tools (m3 – m15)

- Develop a dedicated metadata annotation procedure using state of the art tools & techniques
- Setting-up and applying tools to enable automatic metadata annotation of streaming video
- Develop a multilingual ontology with conceptual domain descriptions in order to facilitate multilingual cross-border search in (distributed) enriched content bases.
- Tailor and integrate an MSN or Google based internet and database search engine for dedicated retrieval of genetics related information.
- Integrate tools and procedures in the database structure and search engine

WP5: Web creation and content provision (m6 – m15)

- Select and collect relevant educational content from distributed providers and establish the central database.
- Develop a procedure/ protocol for tagging and indexing of distributed content.
- Create the portal and website, containing database and internet search engines as well as user ID procedures for dedicated content provision
- Establish a back office tool kit for administration and maintenance of portal and content.

Phase 2 Test & concept validation (m16 – m21)

WP6: Restricted user pilot (m16 – m21)

- Provide restricted user group of pilot partners with access to the portal and database (5 to 8 partners, that will be content provider as well as user of each other's content).
- Develop test protocol based on user requirements and functional specifications of the portal and service.
- Test activities and feedback provision based on questionnaires and interviews with users
- Improvement of technical infrastructure or content enrichment procedures, based on user feedback, resulting in validated procedures, systems and tools.
- Continuous provision of new and updated content by all pilot partner.

Maybe a parallel WP might be necessary dealing with technical aspects that may be encountered in the pilot stage (parallel improvements based on initial feedback, e.g. in 2 subsequent α -test of 3 -4 months each).

Phase 3 Large scale roll-out and evaluation (m22 – m36)

WP7: Real life experiment (m22 – m33)

- Launch of β -version of the portal and database, accessible for a large group of users from all three user groups:
 - o colleague lecturers on genetics (expert access)
 - o students and other academic users (advanced access)
 - o public domain users, such as schools, public awareness campaigns (basic access)
- Test and validation activities with on-line feedback provision from different user groups.
- Further sharpening and tailoring services to practical user needs
- Establish back-office with clear procedures and day-to-day administrative services
- Develop Service Level Agreement (SLA) for content provision and service provision by all involved partners.
- Other activities to be added ... *please add*

What are the elements you would like have tested and validated? Refer to project objectives: share and re-use of educational content on genetics and health.

WP8+9: Pilot evaluation & final sustainability model (m34 – m36)

WP8 and WP9 could be integrated into 1 work package, I think.

- Assessment of project's progress against targeted results
- Evaluate pilot outputs and write evaluation report
- Draft final sustainability / business plan based on initial plan and pilot results.

Other activities

During phases two and three of the project potential users for genetics content (colleague institutes and lecturers, students, school associations, etc.) will be informed and encouraged to use the portal. It will be promoted through various marketing methods such as submission to search engines, links, e-mail marketing etc. WP10 contains awareness activities, WP 11 established relationships with peers and other interesting user groups. *Both WP need to be broken down in sub-tasks, depending on your preferences and ideas. Please add!*

1. Technology Providers

- Academic expert partner in Ontology based metadata annotation; e.g. FZI (<http://www.fzi.de/eng/index.php>), Karlsruhe Research Centre for Information Technology: Specialist in development of Knowledge Management Techniques incl. Ontologies, or KMI (<http://kmi.open.ac.uk/index.cfm>), Knowledge Media Institute, Open University Milton Keynes
- Telelingua (<http://www.telelingua.com>), Belgium based expert in voice recognition and Multilinguality Scansoft (Nuance): Voice recognition
- Linguanet (<http://www.linguanet.net/it/home/index.htm>), Polish based expert company in Multilinguality issues
- New-Phenix (<http://www.new-phenix.com/>), France: Supplier of multimedia platform and developer of a tool to index streaming video
- Microsoft (<http://www.microsoft.com>), MSN search technology / .NET technology / SQL technology; or
- Google (<http://www.google.com>), Google search technology

2. Content owners (and users)

- European Genetics Foundation (<http://www.eurogene.org>): Developed a substantial base of streaming video lectures of leading Profs in Europe (280 lectures) that they wish to share but also text books, abstracts etc. Coordinator and initiator of the project
- Department of Medical Genetics University of Turku (they involve a lecturer that is involved in eLearning, has power point presentations for a course in medical genetics in Finnish)
- Life Knowledge Park (<http://www.life.org.uk/>): A group of Genetic Knowledge Centres in the UK that have digital content available that could be of interest for educational content creation. They are responsible for education of cytogenetics and medical genetics in the UK
- Royal Hallamshire Hospital, University of Sheffield, they have contributed to a set of CDs that are published by Henry Stewart Publishing House with 21 lectures by experts in epidemiologic genetics and a manual of genetic statistics
- L'Institut Fédératif des Neurosciences de l'hôpital Pitié-Salpêtrière (www.chups.jussieu.fr/)
- Erasmus University Medical School
- Institut für Medizinische Genetik, Universitätsklinikum Charité
- Nijmegen Medical Centre, Radboud University
- Institute for Bioinformatics GSF - National Research Center for Environment and Health
- INSERM Collège de France
- Division of Medical Genetics, University of Geneva
- IMCB University of Tartu, Estonian Biocentre
- Hopital Paul Brousse
- Birmingham Women's Hospital, Clinical Genetics Unit

3. Users

- Organisation of the European Cancer Institutes (OEI) – An interest group of 26 oncological institutes in Europe that could function as user group (medical doctors) to search and re-use the educational content
- Institut für Medizinische Genetik - Universitätsklinikum Charité; Has no own digital educational content but is interested to participate
- European Cytogeneticists Association (ECA) - Contact Person: Albert Schinzel
- Instituto de Biología y Genética Molecular Contact Person: Juan J Tellería