Data Curation at U.Porto: Identifying current practices across disciplinary domains

Cristina Ribeiro – mcr@fe.up.pt
Eugénia Matos Fernandes – efernand@reit.up.pt

IASSIST 2011
Vancouver, June 2 2011
Contents

- U.Porto: a research university
- The U.Porto Information System
- The U.Porto Institutional Repository
- Scientific Data Curation Project
- Dataset Collection
- Data Curation Repository
- State University created the 22nd March 1911
  - Origins dating back to the 18th century
- National Law (RJIES)
  - Law n.º 62/2007 (10th September)
  - Changes HEI governance
- U.PORTO
  - New statutes (DL n.º 96/2009, the 27th April)
  - State Foundation under a private law
- U.PORTO: the largest university in Portugal
  (14 public HEIs)
U. PORTO :: Geographical position

one million inhabitants
U.PORTO :: Schools and Research Units

- **Rectorate/Central Services**

- **14 Schools**
  - School of Architecture
  - School of Fine Arts
  - School of Sciences
  - School of Nutrition and Food Science
  - School of Sport
  - School of Law
  - School of Economics
  - School of Engineering
  - School of Pharmacy
  - School of Arts
  - School of Medicine
  - School of Dental Medicine
  - School of Psychology and Education Science
  - Institute of Biomedical Sciences Abel Salazar
  - Business School

- **69 R&D+i units**
  - 31 assessed as excellent or very good

- **30 Libraries + 12 Museums**

- **Student Support Services**
U.PORTO :: Academic Community

- **Students**
  - 30,898 (total)
  - 1st cycle → 9,647
  - Integrated Master → 12,758
  - Master + 2nd cycle → 5,406
  - Specialization → 4,258
  - PhD + 3rd cycle → 2,828

- **Teachers & researchers**
  - 2,366 → 76% PhD

- **Technical & Administrative staff**
  - 1,654
U.PORTO :: Geographic distribution

Pólo II

Pólo I

Pólo III
Faculdade de Engenharia da Universidade do Porto

Últimas Notícias

- 2º Ciclo de Colóquios "Despertar para a Cidade Porto 2004"
  A Faculdade de Engenharia da Universidade do Porto realiza no dia 29 de Outubro o 2º Ciclo de Colóquios "Despertar para a Cidade Porto 2004". O ciclo de colóquios visa aproximar a universidade e a sociedade, pois os temas abordados são de interesse para todos.

Matrículas

Faculdade de Direito da Universidade do Porto

Últimas Notícias

- Inscrições nas Licenciaturas Via Web (excepto 1º ano, 1º vez)
  Entre 11 de Outubro e 1º de Novembro, pode efetuar as inscrições nas licenciaturas via Web. Fica aberta a inscrição em qualquer posto universitário.

Faculdade de Economia da Universidade do Porto

Últimas Notícias

- Defesa de Tese de Mestrado em Economia
  Defesa de tese a ter lugar no dia 20 de Outubro de 2004, pelo licenciado Daniel Duarte Vieira, tendo como título "Elaboração de Mapas de Preços para a Economia". (Consultar detalhes)

- Novo Working Paper da FEP
  WP nº 118: "Low cost carriers, secondary airports and State aid: an economic assessment of the Charleroi affair". (Consultar detalhes)
Institutional Repository :: Ongoing project

- SIGARRA
- INSTITUTIONAL REPOSITORY
- OPEN REPOSITORY
- THEMATIC REPOSITORY
- SCIENTIFIC DATA REPOSITORY

- Ingest
- Storage
- Preservation
- Access
- Dissemination

Full text
Open access
U.PORTO :: Vision

- Goals:
  - transform U.PORTO into a Research University by increasing the quantity and quality of research work
  - reach excellence in the various levels of higher education, according to international standards
  - reinforce U.PORTO's internationalisation
U.PORTO :: Data Curation Initiative

- Curation of and access to the scientific data generated by researchers
- Short study: February 1 to September 30, 2011
- Expected results
  - Sample of existing datasets
  - Sample of existing curation practices
  - A first impression on researchers’ needs
  - Analysis of datasets from a technical point of view
  - Collection of datasets in a standard repository platform
  - Experimental interrogation of datasets
U.PORTO :: Data Curation Initiative :: Method

- Concluded
  - Interviews with researchers
  - Identification of use cases
  - Collection of datasets

- Ongoing
  - Information modeling for datasets
  - Choice of repository platform

- Future
  - Data curation platform
  - Metadata model
  - Datasets available on chosen platform
Data Curation :: Dataset 1 :: Gravimetry

- The Astronomic Observatory of U.Porto
- Researchers from: Science School
- Data: gravimetry measures
  - Data collected in campaigns at several Portuguese regions
  - Dataset is data from 1 campaign
- Curation
  - Researchers replicate data in personal machines and in CD
  - Data in active use for Ph.D. research
- Use Case
  - Researchers or industry get gravimetry dataset
- Metadata
  - Description used by researchers, has to be enriched for the repository
Data Curation :: Dataset 2 :: Pollution

- The Engineering School, Chemical Department
- Data: pollutant densities
  - Data collected in a Portuguese harbor
  - Dataset is data from 1 set of samples for a single date
- Curation
  - Researchers replicate data in personal machines and in CD
  - Data in active use for a university contract
- Use Case
  - Administrative entity gets results analysis
  - Researchers use data as a time series
- Metadata
  - Description provided by researchers, has to be completed with set-level description
Data Curation :: Dataset 3 :: Materials Mechanics

- The Engineering School, Mechanical Eng. Department
- Data: materials fracture
  - Data collected in experimental setup
  - Dataset is data from 1 experiment
- Curation
  - Group leader keeps copies of files, experimental results and analysis using simulation
  - Some data in active use for research, other not
- Use Case
  - Sharing data with other research groups
- Metadata
  - Some description available in the Excel files, requires set-level description
Data Curation :: Dataset 4 :: Civil Engineering

- The Engineering School, Civil Eng. Department
- Data: sensor data from high-speed railways
  - Data collected in experimental setup
  - Dataset is data from 1 experiment
- Curation
  - Group members keeps files, including experimental results, models and analysis using simulation
  - All data in active use for research
- Use Case
  - Sharing data with research partners
- Metadata
  - Some description communicated by researchers, requires set-level description
Data Curation :: Dataset 5 :: Education Science

- The Psychology and Education Science School
- Data: interviews about basic school experiences
  - Data collected from people in a region
  - Dataset is data from 1 set of interviews
- Curation
  - No concern with curation yet
  - Data in active use for an M.Sc. thesis
- Use Case
  - None identified
- Metadata
  - Possible metadata set from the interview script, has to be completed with set-level description
Data Curation :: Dataset 6 :: Psychology Data

- The Psychology and Education Science School
- Data: sensor data from experiments with people
  - Data collected in lab conditions
  - Dataset is data from 1 experiment
- Curation
  - Group leader keeps copies of files, experimental results and analysis
  - Data provided is not in active use for research
- Use Case
  - Sharing data with other research groups
- Metadata
  - Possible metadata set from the experimental setup, has to be completed with set-level description
Data Curation :: Dataset 7 :: Ecology

- The Sciences School, Biology Department
- Data: geo-referenced species
  - Data collected in selected Portuguese regions
  - Dataset is data from 1 set of samples for a single region
- Curation
  - Researchers are preparing a curation plan in the context of an international project
- Use Case
  - Data exchange in collaborative international projects
- Metadata
  - Description provided by researchers
  - Set-level description according to emerging standards
Data Curation :: Dataset 8 :: Economy

- The Economy School, Econometrics Group
- Data: population statistics
  - Data aggregated from official agencies
  - Dataset is data from 1 aggregation
- Curation
  - Researchers replicate data in personal machines and in CD
  - Data can be generated again from the official databases with stored queries
- Use Case
  - Not identified yet
- Metadata
  - Description provided by researchers, has to be completed with set-level description
Data Curation :: Preliminary conclusions

- Very diverse data
  - Data with very different structure, contents and volume
- Similar use cases in data search
  - Suggests models with common search features
- Dataset-level metadata
  - DCMI- Science Metadata
- U.Porto Data Repository
  - Project encourages the definition of a data curation policy
- Open Access is not yet an issue
  - Project contributes to get researchers confidence on the approach first