

## OA to data, data plans and data services



*Access to data is essential for all parts of society: science, policy, decision making, sustainable development, health and welfare.*

- **Bermuda principles 1996**. Made the Human Genome Project a forerunner for OA
- **Berlin Declaration 2003** on OA to scientific knowledge
  - "original scientific research results, raw data and metadata, source materials etc"
- **ICSU Report 2004, CODATA 2009, GEOSS 2005 etc**
  - Managing data and information: Metadata, archiving, interoperability, privacy and security etc
  - Data and information access and dissemination



*Why does it take so long?*

*How can we speed up awareness, incentives, rewards, demands, techniques, regulations, principles etc for cross-border sharing of data?*



## Public data and research

*Public data assembled for purposes of governance, monitoring, surveillance etc of key importance for research.*

*Examples: Mapping and land registration, geological surveys, meteorological and hydrological observations, registers in central statistical offices and boards for health and welfare, hospital biobanks and hospital records, etc.*

- **PSI Directive (EC 2003)** Public Sector Information. Raw data by public bodies should be re-used or integrated into new products and services
- **INSPIRE Directive (EC 2007)** Infrastructure for spatial information in Europe. 34 spatial data themes. Technical implementing rules specified.

*Why is academic access to public data still expensive, restricted and surrounded by complicated procedures and licensing rules?*

*Should the scientific community be invited to the table? Can and should research aspects influence data catchment and content?*

- **Data Protection Directive (EC 1995)** Securing the integrity of individuals while allowing research on health and welfare. Laws and practices vary across countries. Currently under revision.



## Beyond data

*We have “Experimentation, theory, computation and now large and complex data”.*

*But we still largely use stand alone, home made, vaguely documented software and algorithms for mining the data!?*

Integrated open source and open development software – an example:

- **Bioconductor:** Provides integrated tools for the analysis and comprehension of high-throughput genomic data (microarrays, variants, sequence data, annotations, high-throughput assays). Bioconductor uses the **R** statistical programming language, and is open source and open development. It has two releases each year, 516 software packages and an active user community.
- **R** is a free software environment for statistical computing and graphics. Easy prototyping of new computational methods. Packages together software components and documentation. Common object-oriented framework. Supports statistical simulation and modeling activities and data and model visualization capabilities. Is a basis for research in parallel statistical computing. Is under active development by a dedicated international team of researchers

Feb 25th 2010 | From The Economist print edition

