Open Common Collaborative Approaches
“There’s a big world out there!”

Alan Rector
Information Management Group / Bio Health Informatics Forum
Department of Computer Science, University of Manchester

rector@cs.man.ac.uk
co-ode-admin@cs.man.ac.uk

www.co-ode.org
protege.stanford.org
www.opengalen.org
www.clinical-escience.org
Open Common Resources

• “In the common”
  – What is needed for the community
  – Things that must be shared to be valuable
  – Things that can only be built in common
    • Things that must serve a wide common
  – Evolution rather than design
    • A means of selection
    • No one owns a natural language
      – Evolves to meet needs
“Open”

- Editorial Management – Who says what’s in it?
- Development - Who does the work?
- Community involvement/“ownership” – who considers it “theirs”?
- Access - Who can use it?
- Control - Who can build on it? Who must acknowledge it?
- Credit - How are the contributors acknowledged?
- Legal ownership/ “Stewardship” - Who guards it?
- Resources / sustainability - Who pays?
Examples

- **Wikipedia**
  - Voluntary contributions; open critique

- **OpenDirectory**
  - Voluntary Editorships; open critique

- **Linux**
  - Prestige model
    - Key champion; widespread tools

- **Bio**
  - The Human Genome at Sanger; The Gene Ontology & MGED, Developmental & Adult Mouse, FlyBase, ZebraFishBase, …
    - Centrally funded cooperative resource
    - Mandated by journals - paid from project/institutional

- **UMLS – Metathesaurus, CUIs + LUIs, MeSH**
  - NCICB – CaBIO/CaBIG/EVS
    - Governmental

- **Standards**
  - HL7, OpenEHR, HTML, XML, RDF, OWL, …
The Semantic Web

• Many definitions and goals. Mine:
  – A facilitator to create virtual communities with shared understanding
  – A common background knowledge resources
  – Distributed, collaborative, not globally consistent
    • “Managed anarchy”
  – Is it a model for biomedicine? Where?
  – Does it provide tools useful to biomedicine?
    • Many more people with relevant skills
    • Standards – XML, RDF, OWL, WSDL, OWL-S, SWRL, …
Google

• NHS now indexes its information with Google
  – Sudden improvement in access
    • Combined with (a little) metadata hopes to be machine processable
  – Open use though proprietary
Public Resources

• Corpora and natural language resources
  – Expensive to build
    • Plagued in medicine by issues of privacy
  – Probably only possible as combination of academic and
Web Speed Response

- Immediate access
- Immediate response
- Developed in response to practical needs
- Collaborative tools and environments
  - Local tools
  - Collaborative methods
  - Repositories
- Effective critique
- Interest
  - Urgent need or love
“Open management” & “Open Source”

• “Open management” – the end
  – The nature of the process and participation
  – Being “in the common”

• “Open Source” – one possible means
  – Many different licenses
    • Viral vs nonViral
      (GNU Public License (GPL) vs everything else)
Tools

• Open development requires available tools
  – Need to be widely distributed
    • People have them anyway – C compilers
    • Companies buy them anyway – Rational Rose
      – Often with cheap readers available
    • Developers make them available
      – Protégé – Protégé/OWL, RACER, FaCT, OpenGALEN
      – Often the most “profitable” means of academic exploitation
    • Standards bodies commission them
    • Gifts
      – GNUEmacs
Research Questions

• Can open communities using XYZ methodology produce useful resources for interoperability?
  – Under what circumstances? With what methodology?
  – Within what constraints?

• How to leverage what exists
  – UMLS? The Semantic Web? Google?
  – OpenGALEN? S-CT? ICD 9/10/11?

• Can open resources be sustained in health
  – If not, why not?
  – Why is bio easier?

• What tools must be built and sustained to foster open collaborative development?