Connecting to data sets through SciVerse Applications

IJsbrand Jan Aalbersberg
Elsevier, S&T Journals
Access vs. Importance

(Researchers, N = 3824; study by Publishing Research Consortium, 2010)

- Research articles: Highly important AND easily accessible
- Data sets/Data models/Algorithms & Programs: Important but NOT easily accessible

Mean Importance of access vs. Ease of access (Easy or Fairly Easy)

- Clinical guidelines
- Prof/Trade publications
- Reference works
- Books/Monographs
- Conference proceedings
- Doctoral theses/dissertations
- Historical archives/public records
- Market Research reports
- Patent information
- Technical info.

ELSEVIER Building Insights. Breaking Boundaries™
Location and Linking

Where would you be willing to submit your research data?

- Digital archive of organisation
- Publisher (data centre) of your discipline
- External web service (e.g. Google Base)
- I do not want to submit digital research data
- Other

Do you think it is useful to link underlying research data with formal literature?

- Yes 85%
- No 15%

(Researchers, N=1202; study by EU PARSE.Insight, 2009)
STM / ALPSP Position on Data Sets

“... believe that, as a general principle, data sets, raw data outputs of research, and sets or subsets of that data should wherever possible be made freely accessible to other scholars...”

(Statement from STM & ALPSP, June 2006)

“... Raw research data should be made freely available to all researchers. Publishers encourage the public posting of the raw data outputs of research. ....”

(Brussels Declaration on STM Publishing, November 2007)
What does this mean?

- Elsevier does *not* “copyright” raw datasets that are submitted with an article
  - Such datasets can be made freely available by author
  - Journals could request authors to submit data (elsewhere)

- SciVerse ScienceDirect supports the discoverability of datasets through connecting them with the formal research article
  - If supported by scientific community and journal editors
  - By linking from entity, article or through application
What does this mean?

- Elsevier does *not* “copyright” raw datasets that are submitted with an article
  - Such datasets can be made freely available by author
  - Journals could request authors to submit data (elsewhere)

- SciVerse *ScienceDirect supports the discoverability of datasets through connecting* them with the formal research article
  - If supported by scientific community and journal editors
  - By linking from *entity, article* or through *application*
Upon irradiation at 446 nm, the chromophore switches to the cis form, and in the folded state, it is trapped in a metastable state fold [33] and [34] (Fig. 1). In the dark, the trans folded state is regained in seconds. Several studies have shown that the N-terminal section (residues 1–25) of PYP undergoes an almost complete unfolding upon irradiation, exposing a hydrophobic patch on the protein. The N-terminally truncated protein (Δ1–25) is still fully functional, albeit with a slower photocycle.

Fig. 1.
Structures of PYP under various conditions. (a) NMR solution structure of dark-adapted Δ25PYP (PDB code 1XFN). (b) X-ray crystal structure of dark-adapted full-length wild-type PYP (the N-terminal crystallographic domain, PDB code 1NWZ). (c) NMR solution structure of light-adapted Δ25PYP (PDB code 1XFO).

Our approach to using PYP as a photoswitch aims to use the folded state of PYP to sterically prevent interaction, which is then allowed to occur in the flexible light-adapted form. It has been suggested that such a mechanism may explain how PYP interacts with its putative partner protein in vivo. We wish to explore the possibility of designing a genetically encoded photoswitchable DNA binding protein by fusing PYP to the prototypical leucine-zipper-type DNA binding protein GCN4-bZIP. The bZIP domain

- Automatically or author-tagged
- Automatically linked to data:
  - External
  - Selectable
  - Personal preference
Linking from an article

- Author-posted at repository
- Automatic connection between Elsevier and repository
Dataset linking: how does it work?

- Based on image-based linking
- SD article asks for a “data set image” from CCDC
- If data available, CCDC shows image with link to CCDC
- If no data available, CCDC shows 1-pixel transparent image (which is de-facto invisible for the user)
Dataset integration: how does it work?

USER → SD Article → SD Server → articles

USER → SD Article → CCDC Server → data sets

USER → SD Article → SV app → Research Data Server → data sets
SciVerse Applications: What is it?

- Integrates data and articles
- Puts data in the proper context
- Discovers remote data

- Open development platform
- For partners and developers
- Can be customized per customer
SciVerse Applications: What is it?

- Integrates data and articles
- Puts data in the proper context
- Discovers remote data

- Open development platform
- For partners and developers
- Can be customized per customer
Pulling data in from Protein Data Bank

- Author-tagged
- Data from PDB
- 3D Visualisation
  - Select
  - Zoom
  - Rotate
- All inside article
Pulling data in from Gene Data Bank

- Author-tagged
- NCBI gene viewer
- Functionality to drill down deeply into gene information
- All inside article
- PANGAEA-hosted
- Author-posted data
- Real-time match
- Full Google Maps functionality
- Links to PANGAEA
- All inside article

ARTICLE
Displaying Protein Interactions (MINT)

- Curated information
- MINT-created application
- Functionality to drill down deeply into protein interactions
- All inside article
Exoplanets+

- **Apps for Science Challenge**
- Extracts exoplanets (extrasolar planets) from articles and displays rich data alongside
- Displays data from:
  - Exoplanets.org
  - Exoplanet.eu
  - Visual Exoplanet
  - NStED
  - SIMBAD Astronomical DB
Species+

- *Apps for Science Challenge*
- Extracts species from articles and displays rich data alongside article
- Displays data from:
  - Encyclopedia of Life
  - Discover Life
  - Global Diversity Information Utility
1. Elsevier supports freely accessible data sets

2. SciVerse ScienceDirect enables wide discoverability of related data sets at data repositories through entity / article linking or article interoperability

3. SciVerse Applications enable data sets to be further exposed, and be put into the context of the article, further increasing value of article and data sets

4. Elsevier invites data repositories to collaborate!
Summary (2)

1. Elsevier supports freely accessible data sets

2. SciVerse ScienceDirect enables wide discoverability of related data sets at data repositories through entity / article linking or article interoperability

3. SciVerse Applications enable data sets to be further exposed, and be put into the context of the article, further increasing value of article and data sets

4. Elsevier invites data repositories to collaborate!
Time’s Up!

About your speaker:

Name: IJsbrand Jan Aalbersberg
Company: Elsevier
Tel: +31 20 485 3021
Email: ij.j.aalbersberg@elsevier.com