Focus group 1

Managing the supplier lock in risk

Facilitator: Kieren Lovell

Summary

1. Definite agreement:
   a. Export *will* happen.
   b. PDF is not an export format for importing.
2. Think about it ahead of time.
3. Vendors always tell the truth like Estate agents always tell the truth.
4. Accept reality of poor import/export fidelity but specify minimal acceptable level at start.
5. Consider user's specific requirements. (E.g. InChi metadata absolute requirement for a chemist to reproduce.)
6. If you can't import into new system, allow referencing to old system.
7. Exporting to open standards is import but to the *right* format. PDF is great for viewing by humans but useless moving forward. Make a distinction between
   a. Human thought process (maybe PDF summary of human-written docs)
   b. Data which was put in should come out in the format it was put in
   c. Metadata which was put in should be understandable.
8. Simply don't buy things you aren't 100% sure will allow useful export. Define what useful export is at purchase/selection time.

Points which don't fit below

1. Lock in is helped by having a culture of making import/export not the unusual case. A case study was given of a system which routinely exports to four other systems.
2. Metadata export/import is basically impossible with 100% fidelity. Example was given for DSpace @ Cambridge which is open but still only had an anecdotal 9%5 fidelity.
3. When migrating, we care about *referencing* old data when migrating but not necessarily that it becomes *live* data in a new system.
4. Would be interesting to know statistics on what sort of per-user storage is actually needed.

Problems that the group wants to solve

1. Address concern about investing, e.g., 1 year of work importing information into system and then determining that the system is not useful.

2. Effort to get data back out of PDF is incredibly large.
3. Open but semantically poor formats (docx, PDF) are not useful for data export.

Possible long-term solutions to these problems

1. Check that export files are both *readable* and *useful.*
2. Educate users that PDFs are *view* formats but not an archive format.
3. Open standards for export. But be careful about what data those standards try to achieve.
4. Open standards for metadata export too. (It was suggested that [DataCite](#) schema would be useful.)
5. At start of ELV procurement/selection, grasp the realities and up-front specify a minimal level of fidelity of output rather than asserting the impossible that export with 100% fidelity is possible.
6. Certain specific domains will need strong guarantees on specific metadata. E.g. Chemists would like to mandate [InChi](#) export. General point: get users to think carefully about *vital* metadata at system design/selection stage.

Quick wins

1. Only buy ELVs will well specified metadata export. Prioritise non-proprietary standards.
2. Make sure the "at rest" format for data or metadata is in a format which more than one piece of software can read and which is ideally an open interoperable standard.