Pragmatic data interoperability

Kirsty Kitto
Connected Intelligence Centre
@KirstyKitto • kirsty.kitto@uts.edu.au
what is UTS:CIC?

Connected Intelligence Centre
- UTS innovation lab specialising in Learning Analytics
- provides in house data science consultancy
- academics teach data science and perform research
- trains PhDs in Learning Analytics

some institutional context: UTS2027

Learning for a lifetime

- “Our commitment to bringing about a transformation for society to a lifetime of learning means we must help curate and support learning and professional development for our students, whoever they may be, for their entire lives”

Personal learning experience

- “Create a single, seamless experience learning platform. This will be a data-rich, single interface portal that dynamically adapts and supports a student’s journey.”

https://www.uts.edu.au/about/uts-2027-strategy
where does learning happen?
traditionally LA has focused upon providing analytics within the confines of specific systems built by vendors… (e.g. LMSs, eBooks, SIS)
but learning happens everywhere!
the connected learning analytics toolkit

[Diagram:
- xAPI
- Learning Record Store
- Scraping
- Analysis
- Social media
- Students
- Academics
- Learning analytics
- Admin & Developers]
I am not talking about that though!

You can read the papers if you like…


learning occurs over a lifetime!
the emerging UTS digital learning ecosystem

- LMS works as core to a decentralized system
- The core is heavily supported by other elements in the ecosystem
- Several core external tools are utilized to enrich student experience
- But many other tools are used by our academics! (how to support?)
- LA is a focus for for enabling personalization (and CIC builds tools!)
- But student data can be generated in a wide range of tools…
so data interoperability is essential to us!

but what type of interoperability?
big and comprehensive?

or loose and modular?

you could ensure that all educational technology uses one data stack…

- but how long would this remain current?
- and how quickly will it evolve as new use cases arise?
- and who is control of it anyway?
- and how comprehensive can this approach actually be?

or you could try and do something that is more modular…

- where any LRP can get up and running quickly to provide data
- but then you need to ensure that there is a way to map data between different providers
we have already broken the “comprehensive” model...
<table>
<thead>
<tr>
<th>Icons</th>
<th>Actor</th>
<th>Verb</th>
<th>Object</th>
<th>Start Date</th>
<th>End Date</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="user.png" alt="User" /> <img src="note.png" alt="Note" /></td>
<td>Sao Peina</td>
<td>created</td>
<td>Note</td>
<td>8 days ago</td>
<td>14 Nov 2019</td>
</tr>
<tr>
<td><img src="user.png" alt="User" /> <img src="note.png" alt="Note" /></td>
<td>Kirsty Kitts</td>
<td>replied</td>
<td>Note</td>
<td>8 days ago</td>
<td>14 Nov 2019</td>
</tr>
<tr>
<td><img src="user.png" alt="User" /> <img src="note.png" alt="Note" /></td>
<td>Jose Beltran Villardy</td>
<td>created</td>
<td>Note</td>
<td>8 days ago</td>
<td>14 Nov 2019</td>
</tr>
<tr>
<td><img src="user.png" alt="User" /> <img src="note.png" alt="Note" /></td>
<td>Arche Aguilar</td>
<td>replied</td>
<td>Note</td>
<td>8 days ago</td>
<td>14 Nov 2019</td>
</tr>
<tr>
<td><img src="user.png" alt="User" /> <img src="note.png" alt="Note" /></td>
<td>Arche Aguilar</td>
<td>created</td>
<td>Note</td>
<td>8 days ago</td>
<td>14 Nov 2019</td>
</tr>
<tr>
<td><img src="user.png" alt="User" /> <img src="note.png" alt="Note" /></td>
<td>Kirsty Kitts</td>
<td>created</td>
<td>Note</td>
<td>8 days ago</td>
<td>14 Nov 2019</td>
</tr>
<tr>
<td><img src="user.png" alt="User" /> <img src="note.png" alt="Note" /></td>
<td>Arche Aguilar</td>
<td>created</td>
<td>Note</td>
<td>8 days ago</td>
<td>14 Nov 2019</td>
</tr>
<tr>
<td><img src="user.png" alt="User" /> <img src="note.png" alt="Note" /></td>
<td>Arche Aguilar</td>
<td>created</td>
<td>Note</td>
<td>8 days ago</td>
<td>14 Nov 2019</td>
</tr>
<tr>
<td><img src="user.png" alt="User" /> <img src="note.png" alt="Note" /></td>
<td>Kirsty Kitts</td>
<td>created</td>
<td>Note</td>
<td>8 days ago</td>
<td>14 Nov 2019</td>
</tr>
</tbody>
</table>
Note: we have already broken the “big and comprehensive” model...
and even if we didn’t... some other new tool or data source would come along... what then?
THE IDEALIST VS. THE PRAGMATIST

DUUDE, I HAVE SOOO MANY IDEAS THAT WOULD REVOLUTIONIZE LIFE ON EARTH AND BRING AWESOMENESS TO MANKIND.

HOW `BOUT YOU JUST TAKE OUT THE GARBAGE AND DO THE DISHES FOR ONCE??

© 2011 Connie J Sun
www.conniewonnie.com

pragmatic data interoperability
a pilot ETL pipeline – the LA-API
4 related code repositories for ETL pipeline

just got permission to open source the whole data pipeline… anyone want to link up their tools or data sources?
the GraphQL schema provides flexible data access...

What if I want counts of activity?
{
  searchNotes(search: email,
  value: "kirsty.kitto@uts.edu.au"){
    parentName
    createdAt
  }
}

What if I want to do text analysis?
{
  searchNotes(search: email,
  value: "kirsty.kitto@uts.edu.au"){
    verb
    parentName
    createdAt
    author{
      name
    }
    text
  }
}
that is... the GraphQL schema delivers the pragmatic data interoperability
resolvers fetch the data from anywhere hooked up to the LA-API

```javascript
async searchNotes(parent, args) {
    let notes = [];
    if(args.search === 'subject') {
        let value = new RegExp(args.value, 'i');
        notes = await Note.find({parentName: {$regex: value}});
    }
    if(args.search === 'email') {
        let user = await User.find({email: args.value.toLowerCase()})
        if(user.length === 0) return [];
        notes = await Note.find({user: new ObjectId(user[0]._id)});
    }
    if(args.search === 'platform') {
        notes = await Note.find({platform: args.value});
    }
    if(args.search === 'verb') {
        notes = await Note.find({verb: args.value});
    }
    if(args.search === 'title') {
        let value = new RegExp(args.value, 'i');
        notes = await Note.find({title: {$regex: value}});
    }
    return notes;
}
```
what is the activity of a user?
what text did they write?

```json
{
    "date": {"searchNotes": [
    {
            "verb": "created",
            "parentName": "Date literacy series sandbox",
            "createdAt": "2017-11-16T06:31:08.000Z",
            "author": {
                "name": "Kirsty Kitto"
            },
            "text": ""
        },
        {
            "verb": "replied",
            "parentName": "36103 Spring 2018 - Statistical Thinking for Data Science",
            "createdAt": "2018-10-05T01:58:31.000Z",
            "author": {
                "name": "Kirsty Kitto"
            },
            "text": "Hi Rebecca,
Next week we will talk a lot more about AT3 tomorrow, but yes - you should be adding technical details to your report. Are you sure you are not talking about AT2C? (Which does have an appendix). Links are fine for that one, but they need to be well-curated (naming them, etc.)"
        }
    ]
```
no... actually I want to send data to OnTask...
the LA-API provides a useable solution to the ongoing problems faced by the LA community with integrating data from multiple tools
it provides a flexible and extensible way of moving data between LA tools
encourages best practice use of xAPI Profiles and Caliper, as data is more likely to already be integrated (easier if you do the right thing!)
but it still lets people integrate their other favorite (i.e. weird) tools…
Questions?