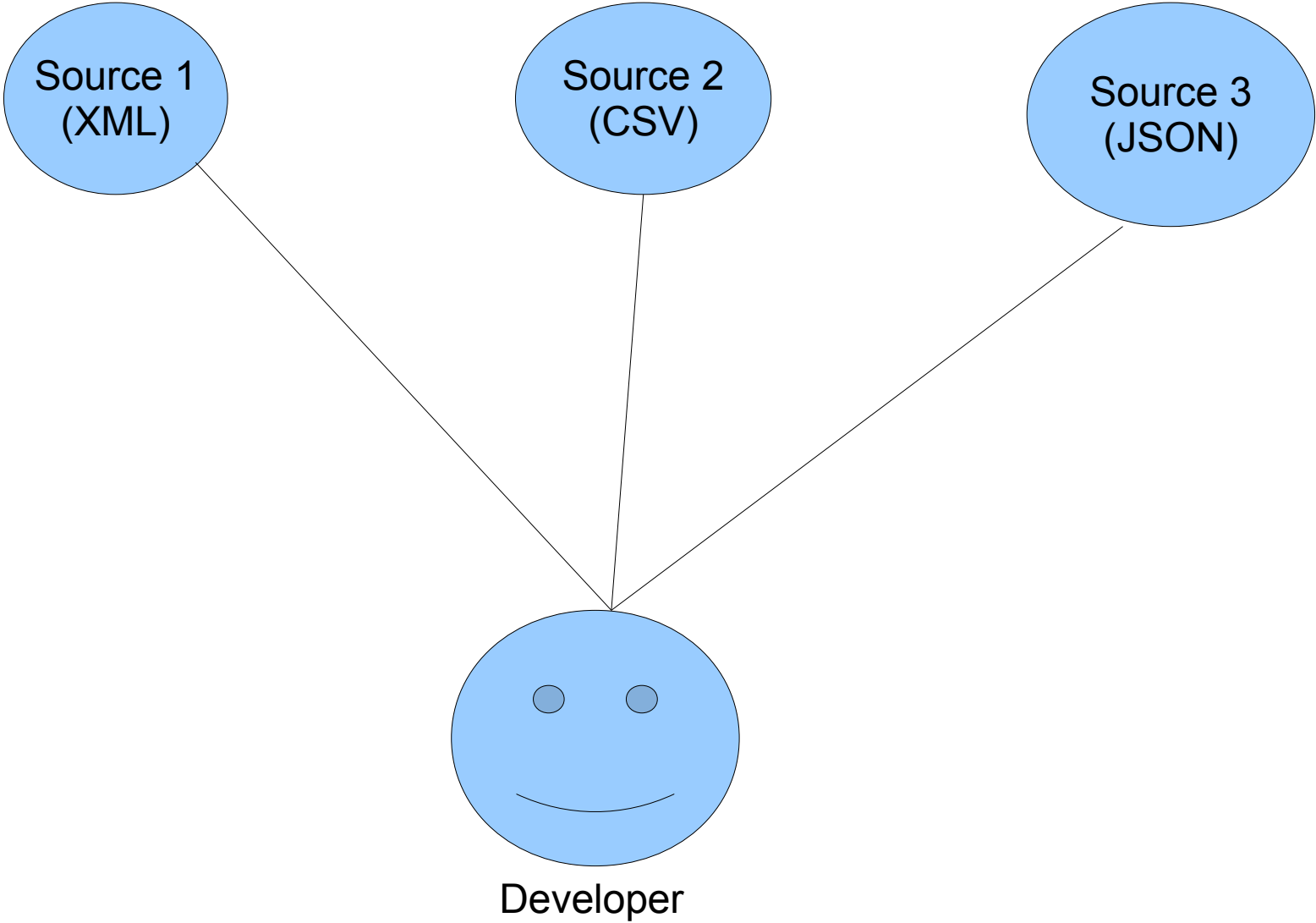


Data sharing platform

Kai Kousa

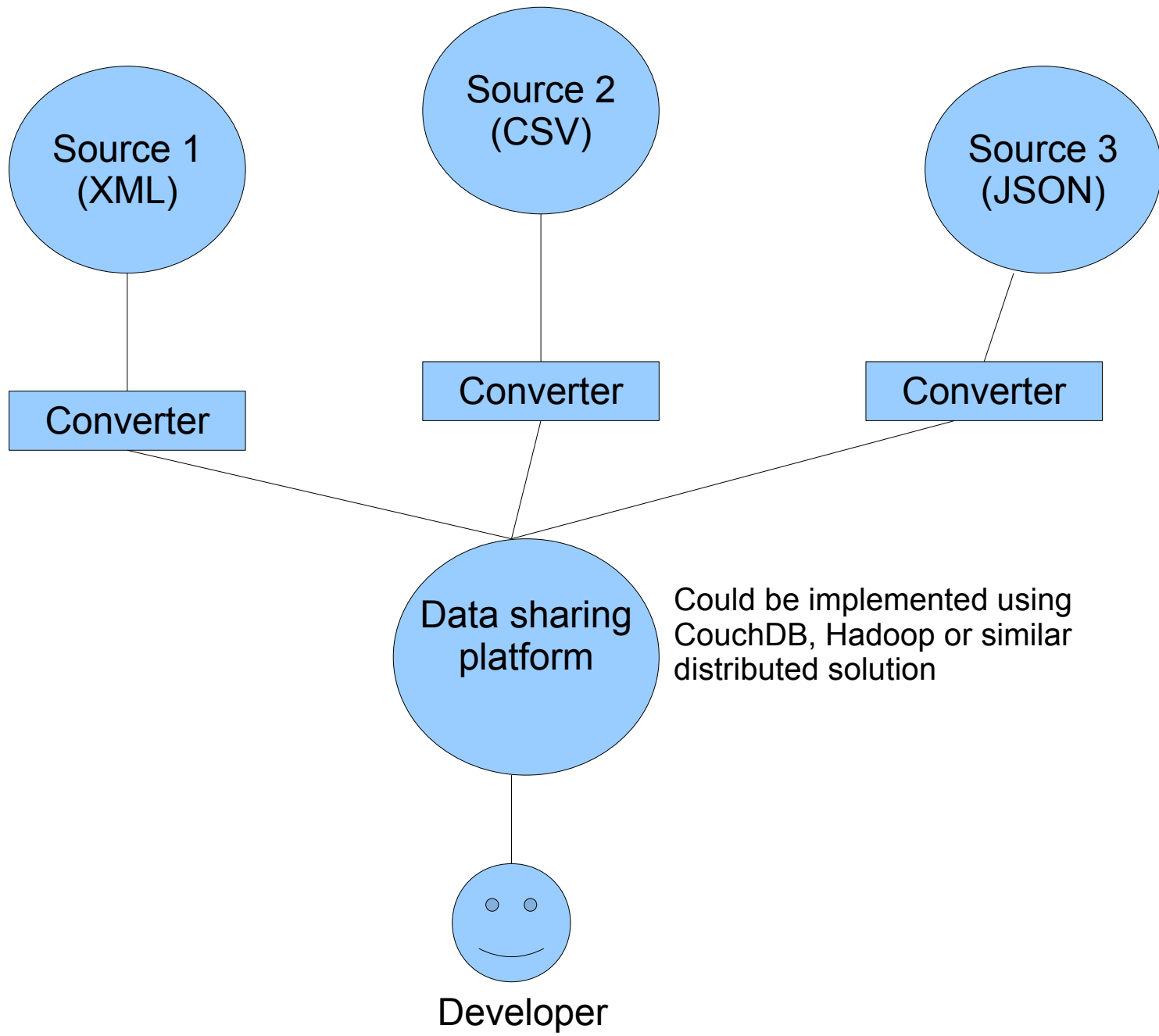
Creative Commons Attribution-Noncommercial-Share Alike 1.0 Finland -licensed



Scattered data

- It's up to the developer to gather and process all relevant data
 - This usually means one or all of the following
 - Parsing the data
 - Converting the data
 - Pushing data to a relational database or similar for querying purposes
- Lots of boilerplate code before the data can be used for the end product or anything useful

The solution?



Data sharing platform

- The platform takes the burden of processing the source data to a usable and query-ready format
- The developer chooses the data sets he/she wants to work with and then starts querying and playing with the data

Converters

- Provided by third party developers or the administrators of the original data source

Back-end

- The data storage would be best to implement using a schemaless database solution such as CouchDB
 - The traditional solution of a schema based relational database would be too burdensome to support
 - The data would still be easy to query as CouchDB uses map-reduce algorithms to query the data

Developer front-end (or API)

- Developers could write simple applications using a web-based UI
 - Choose the data sets, write the map-reduce functions in Javascript, choose the graph to plot with the data and the submit
 - Applications could also be saved for other users to expand on
- For more complex applications there would also be libraries for accessing the service

Summary

- Uniform programming interface for many different data sets
- Access to data sets is made available by converters / plugins
 - Provided by third party developers or administrators of the original data source
- Developers can get straight to the data without extra work

References

- <http://couchdb.apache.org/docs/intro.html>