A semantic dependency-graph-based approach combining platforms hosting data and applications

Enhancing creative synergistic publishing and organizing scientific competitions on the web

Sayoko Shimoyama, Robert Sidney Cox III, David Gifford and Tetsuro Toyoda
Integrated Database Unit, Advanced Center of Computing and Communication (ACCC),
RIKEN
2-1 Hirosawa, Wako, Saitama, 351-0198, Japan.
toyoda.tetsuro@gmail.com

Linked open data increases availability of original scientific and other data. Modifiable or 'forkable' open-source programs hosted on shared platforms make applications utilizing these data ready for reuse. However, data resources and applications are often hidden from each other and external reuse by separation of publication and access to data repositories. We constructed the LinkData.org platform to automate publishing together of linked open data, applications, and introduce the 'dependency graph' illustrating relationships between data, applications and users. Dependency graphs allow transparent evaluation of data and application integration. Data and dependency graphs are accessible with open semantic APIs. Because dependency graphs combine both data and applications published on the platform, users easily create new applications for data and publish new data resources for use with applications. This yields a creative synergy cycle between data publication and application development, as shown applied to a scientific competition for design of synthetic regulatory DNA.

Keywords—semantic web; linked open data; synthetic biology; bioinformatics; data and application web publishing

I. INTRODUCTION

Data repositories and directories for open data such as The Comprehensive Knowledge Archive Network (CKAN) web-based system for the storage and distribution of data, supported by the Open Knowledge Foundation, help users register their data resources and locate related data. Resource Description Framework (RDF) graph structure data format is the standard for sharing linked open data (LOD) on the web.

The LOD model with RDF and SPARQL endpoints gives open access to the data for any external applications (Apps) worldwide, however, the act of separating the data from the applications on the web makes the synergic collaboration between data and applications invisible; resulting in the situation that contributions to opening and maintaining the data are not as appropriately evaluated as the contributions to the Apps, and thus the situation does not motivate academicians to make such contributions as to donate their own datasets.

To overcome this situation, we developed LinkData.org (http://linkdata.org) as a data publishing platform and LinkDataApp (http://app.linkdata.org) as an application publishing platform, and combined them by automatically recording dependency graphs that relate data and Apps using the data; thus, making LinkData as a repository of dependency graphs connecting Apps and datasets, as well as a repository of applications and datasets.

Here we show the cycle actually enhancing various synergistic collaborations and organizing a web-based scientific competition for synthetic biology promoter design. Further LinkData.org displays a usability analysis score calculated based on the dependency graphs to rank highly useful data and applications, so that the scores motivate users to release and update their data and applications, and to trust other's open data.

II. LINKDATA FUNCTIONS

A. LinkData as an RDF publishing platform

1) Support functions for creating table data to upload: As a support function that allows Users to easily define schema, LinkData provides a GUI by which anyone can create and download a template. When a User selects the "Input Table Data" menu and enters metadata for their data using this GUI, a table format Excel file using column names for RDF properties is generated, and this file can be downloaded. Users input their data to the template to create their own table data for uploading.

2) Conversion to RDF format and publishing: Template data tables can be uploaded, converted to RDF format, and published online at LinkData.org. When a User selects "Convert to RDF" and uploads the table data file in Excel or TSV format, anyone accessing the published data's webpage will be able to browse and download the table data, a template for table data, as well as in RDF format.

3) Reuse Data Function: Schemas of all of published Data can be reused for publishing new datasets. Users can activate the reuse table data function at the published Data webpage and download a revisable template to use with their own data.

[Paper Submission for SWWS'13 Conference]

TABLE I. ENTITIES AND LINKS OF LINKDATA CONCEPTS

Entity	**Definition**	
Data	A single data set which has been published by a User in LinkData	
Application (App)	A single application which has been published by a User in LinkData	
User	A user who had registered for a LinkData account	
Link	**Term**	**Definition**
Data(new) → Data(old)	reuse	Create new Data by reusing existing Data
Data → User	contributed	The relationship between Existing Data and the user who created the Data
App(new) → App(old)	fork	Create a new App by reusing an existing App's program code
App → Data	load	Create an App by specifying some files as input from some particular Data
App → User	contributed	The relationship between an Existing App and the user who created the App
User(A) → User(B)	follow	User A follows user B to receive updates and information of evaluated Data and Apps by user B
User → Data	vote	A user gives a rating of Useful or Un-useful for considered Data
User → App	vote	A user gives a rating of Useful or Un-useful for a considered App

4) Application development support function: For application developers who want to use Data, the LinkData platform provides APIs which allow them to access to the contents of Data. Developers will be able to get the contents by five formats: TSV, RDF/Turtle, RDF/JSON, RDF/XML and RSS in their applications.

B. LinkDataApp as an application publishing platform

1) Creating application by editing sample program: We provide two ways to create new Apps: one is to select the "Create App" menu and the other is to go to LinkData's published Data page and create a new App for the data; a sample program of JavaScript is automatically generated when a User selects a file from published Data as an input and creates a new App. The User can edit the sample program on a web browser to develop an original App. Anyone accessing to the published App page will be able to execute and download the App.

2) Forking application to publish as a new one: Users can publish new Apps by forking any App created by others. When a User selects the "Fork App" menu or goes to a published App page, click the "Fork this app as your new one" button to open the program editor. After modification, the program can be published as a new App.

3) Changing input files to create a new application: When a User forks an App, the Input Data control system allows the user to control which LinkData input is loaded. A tagging system is provided to distinguish multiple files which might be referenced by an App. By changing the input data, even a non-programmer can add new functionality to the App.

III. METHODS

A. Entities and Links of LinkData concepts

Our combined platforms use three entities: Data, App and User. Data is a single data set which has been published by one or more Users of LinkData. It must have at least one file which is uploaded by User. App is a single JavaScript application which has been published by a User in LinkDataApp. It must load at least one file from Data. User is a person who had registered for as a user of the platforms. The relationships among entities are described as eight links shown in table 1. A link relationship represents the graph association from one entity node to another by which various metrics of value can be assigned to the recipient node because of the association. The metric value of each type of link and the count of these links are used according to an algorithm to assign a usability value.

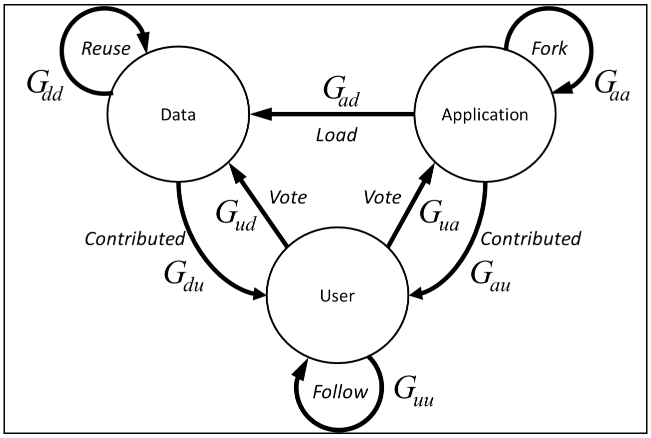


Fig. 1. **Dependency graph for usability analysis**

Three types of nodes: data, application and user, and eight types of links: Gdd, Gaa, Guu, Gad, Gud, Gua, Gdu and Gau are shown in the figure.

- Gdd is a graph connecting a data set to another data set that reuses the same data template or schema.
- Gaa is a graph connecting an application to another forked or modified application.
- Guu is a graph connecting a user who is following another user.
- Gad is a graph connecting a data set to an application loaded or used by that application.
- Gud is a graph connecting a user to a data set which is rated as "useful" or "un-useful" by the user.
- Gua is a graph connecting a user to an application which is rated by the user.
- Gdu is a graph connecting a data set to a user which is contributed by the user.
- Gad is a graph connecting an application to a user which is contributed by the user.

For the link new Data to old Data, a link termed “reuse,” will be generated when a User creates new Data by reusing an existing schema from the old Data. The link Data to User, termed “contributed”, will be generated when a User creates any Data. The link new App to old App, termed “fork”, will be generated when a User creates a new App by reusing an existing App’s program code. The link App to Data, termed “load”, will be generated when a User creates a new App by specifying some files as input from some particular Data or loads some files in his/her existing App. The link User to User, termed “follow”, will be generated when a User follows another User. For example, if User A follows User B, User A can receive updates about User B’s Data or App and information about evaluated Data and Apps by User B. The links User to Data and User to App, termed “vote”, will be generated when a user browses a Data or an App and gives rating of Useful or Un-useful for the considered Data or App. Dependency Graph for calculating Usability Scores.

B. Dependecy Graph for calculating Usability scores

In a dependency graph, the relationship between data and application make up the Utility portion of the graph measuring effectiveness of loading, data and data template reuse, and application forking (cloning and modification) in the creation of information. Three types of nodes: data, application and user, and eight types of links: Gdd, Gaa, Guu, Gad, Gud, Gua, Gdu and Gau are shown in Fig. 1. All the dependency graphs are downloadable from the LinkData.org APIs.

```
[API-TSV]
http://linkdata.org/api/1/graph/reuse_tsv.txt
http://linkdata.org/api/1/graph/fork_tsv.txt
http://linkdata.org/api/1/graph/load_tsv.txt
http://linkdata.org/api/1/graph/follow_tsv.txt
http://linkdata.org/api/1/graph/vote_data_tsv.txt
http://linkdata.org/api/1/graph/vote_app_tsv.txt

[API-JSON]
http://linkdata.org/api/1/graph/reuse_rdf.json
http://linkdata.org/api/1/graph/fork_rdf.json
http://linkdata.org/api/1/graph/load_rdf.json
http://linkdata.org/api/1/graph/follow_rdf.json
http://linkdata.org/api/1/graph/vote_data_rdf.json
http://linkdata.org/api/1/graph/vote_app_rdf.json
```

IV. Results and Discussion

A. Count of relationships among three entities indicates creative synergy cycle

LinkData hosts 557 datasets and 260 applications as of March, 2013. Datasets contain 350 public, 40 limited, and 162 private. Applications contain 160 public, 55 limited, and 45 private. There are a large number of Load (App to Data) relationships indicating Apps created by specifying some files as input from some particular Data (Table 2). There are also many Fork (App to App) relationships representing applications created by re-using program code from another application. In contrast, there are a few Reuse (Data to Data) classified relationships of new data created by using the template from another data set. It is thus clear that there is a stronger synergy cycle between data resources and applications than “in data” (between data and data). In other words, this indicates that a platform which has both capabilities of publishing data resources and creating applications has higher creativity than one having only one capability of data resource creation.

TABLE II. Count of relationships among data resources, applications and users in LinkData

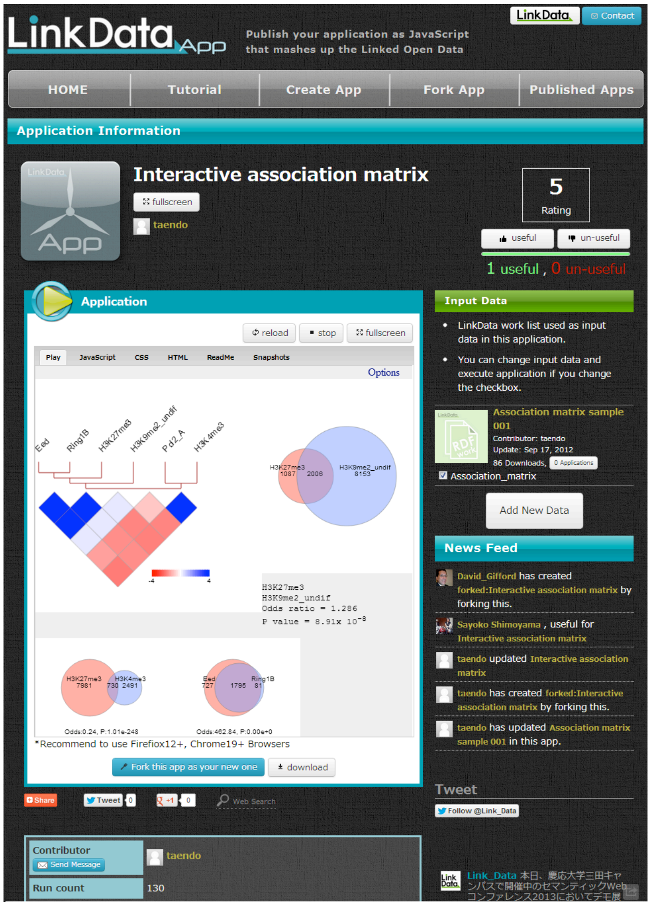
Kind of relationship	Count
Load (App to Data)	166
Fork (App to App)	137
Reuse (Data to Data)	39
Follow (User to User)	52
Vote (User to Data)	244
Vote (User to App)	89

LinkData provides a public place to publish and analyze data. Hosting both data and apps together promotes useful public RDF data exchange between fields and the creation of new interdisciplinary fields. This spreads technology for scientific data to other fields, and educates about RDF techniques for any field.

B. Visualization of a creative synergy cycle between data publication and application development

Biological data analysis is one of the most important domains of applications of LOD in science. Fig. 2 shows an app called “Interactive Gene Association Matrix” for publishing a research result visualizing research data with Linkdata.org, where association analysis of two elements using a Venn diagram indicates how these elements associate or exclude each other. Tables and diagrams of co-localization of transcription factors and conservation between different species provide unbiased views of overlap or exclusion between two conditions. However, if the number of compared elements grows it could become too complex to see which items correlate well and which ones do not, so a comprehensive and interactive visualization tool should help researchers summarize their data and provide an overall view of their dataset.

The Interactive Gene Association Matrix runs on the LinkData web platform requiring no software installation. Researchers store their own association tables in LinkData and obtain automatically clustered matrix diagrams and Venn diagrams having statistical evaluation using hypergeometric distribution. The implementation shown indicates a blue (positively correlating) or red (negatively correlating) cell for each combination of two elements. Color intensity represents logarithm of odds ratio, and statistical significance can also be incorporated into the matrix as cells are masked in gray when the displayed overlap is insignificant. The diagram responds to



Publish your application as JavaScript that mashes up the Linked Open Data

Interactive association matrix

```
H3K27me3
H3K9me2_undif
Odds ratio = 1.286
P value = 8.91x 10-8
```

Odds:0.24, P:1.01e-248

Odds:462.84, P:0.00e+0

*Recommend to use Firefiox12+, Chrome19+ Browsers

- LinkData work list used as input data in this application.
- You can change input data and execute application if you change the checkbox.

Association matrix sample 001
Contributor: taendo
Update: Sep 17, 2012
86 Downloads, 0 Applications
☑ Association_matrix

David_Gifford has created forked:Interactive association matrix by forking this.

Sayoko Shimoyama , useful for Interactive association matrix

taendo updated Interactive association matrix

taendo has created forked:Interactive association matrix by forking this.

taendo has updated Association matrix sample 001 in this app.

Link_Data 本日、慶応大学三田キャンパスで開催中のセマンティックWebコンファレンス2013においてデモ展

Fig. 2. **Interactive association matrix Application created on LinkDataApp** http://app.linkdata.org/run/app1s64i

a user’s cursor moves and Venn diagram would be shown in right panel indicating statistical significance and raw odds ratio. Preferred Venn diagrams can be saved below when a cell is clicked to compare with overlap of other elements.

The matrix in Fig. 2 is made using epigenetic marks of transcription start sites of genes in mouse embryonic stem cells. This tool is not limited to gene by gene analysis, and can also be applied to any type of datasets if 2x2 contingency tables are available. The application can interpret RDF of data uploaded on LinkData. LinkDataApp allows all users to fork the application so that modified versions can have differences such as color representation and clustering algorithms. For example, it may be possible to compare various x and y elements for other species as well. The LinkData platform is very flexible for recombining different datasets as well as modifying the programming in LinkDataApp according to researchers’ needs; and thus reuse of data and applications are observed as shown in Fig. 3.

C. Rating data and apps based on the dependency graphs

The dependency graph allows users to dynamically contribute to and benefit from an automated rating of both data and applications. The usability analysis score LinkData.org displays (Fig. 2 upper right) is calculated based on the dependency graphs to rank highly useful data and applications, so that the scores motivate users to release and update their data and applications, and trust other’s open data. The core version of this rating system combines various works rating parameters as follows:

Rating published data resources + applications

In this type of rating, a user “votes” by using their judgment of the usability of the application. (Fig. 5) Users click a “useful” or “un-useful” button for positive or negative rating.

Rating for a LinkData work（LinkData）

Score = Useful count - Un-useful count + App count

This rating metric also integrates an indication of the App count measuring the number of apps using a data resource.

Rating for an Application （LinkDataApp）

Score = Useful count - Un-useful count + Fork count

An application’s ranking also benefits from how many other applications have been generated as modified “forked” versions created by using the program code of another app.

For example in Fig. 2 and Fig. 3 application app1s69i is a fork of app1s64i. App1s69i loaded 2 data sources, contributed 1 time for a rank of just 3. App1s64i loads 1 data source, was contributed 1 time, and as well was forked 2 times and voted for 1 time to give a total rating of 5. In this fashion each app, dataset and user can be compared for total activity and usefulness in turn, as shown in Fig. 4.

D. Application to a Scientific Competition showing creative synergy cycle on a massive scale

For the synthetic biology competition GenoCon2 (http://genocon.org) [1], we challenged participants to design novel regulatory DNA for controlling gene expression in the thale cress plant *Arabidopsis thaliana*. Participant DNA designs will be synthesized and tested for tissue and time specificity in a real plant. To allow non-experts an opportunity for DNA design we built a computer aided design tool on the LinkData platform, called PromoterCAD (Fig. 5).

Using PromoterCAD function modules, genes with the desired properties can be found and mined for regulatory motifs. These are introduced into the synthetic promoter by user choice of regulatory position. Repeating this process can create complex regulation at the promoter. Finally, the DNA design is exported for error and safety checking, DNA synthesis, and experimental characterization.

Using PromoterCAD function modules, genes with the desired properties can be found and mined for regulatory motifs. These are introduced into the synthetic promoter by user choice of regulatory position. Repeating this process can create complex regulation at the promoter. Finally, the DNA design is exported for error and safety checking, DNA synthesis, and experimental characterization.

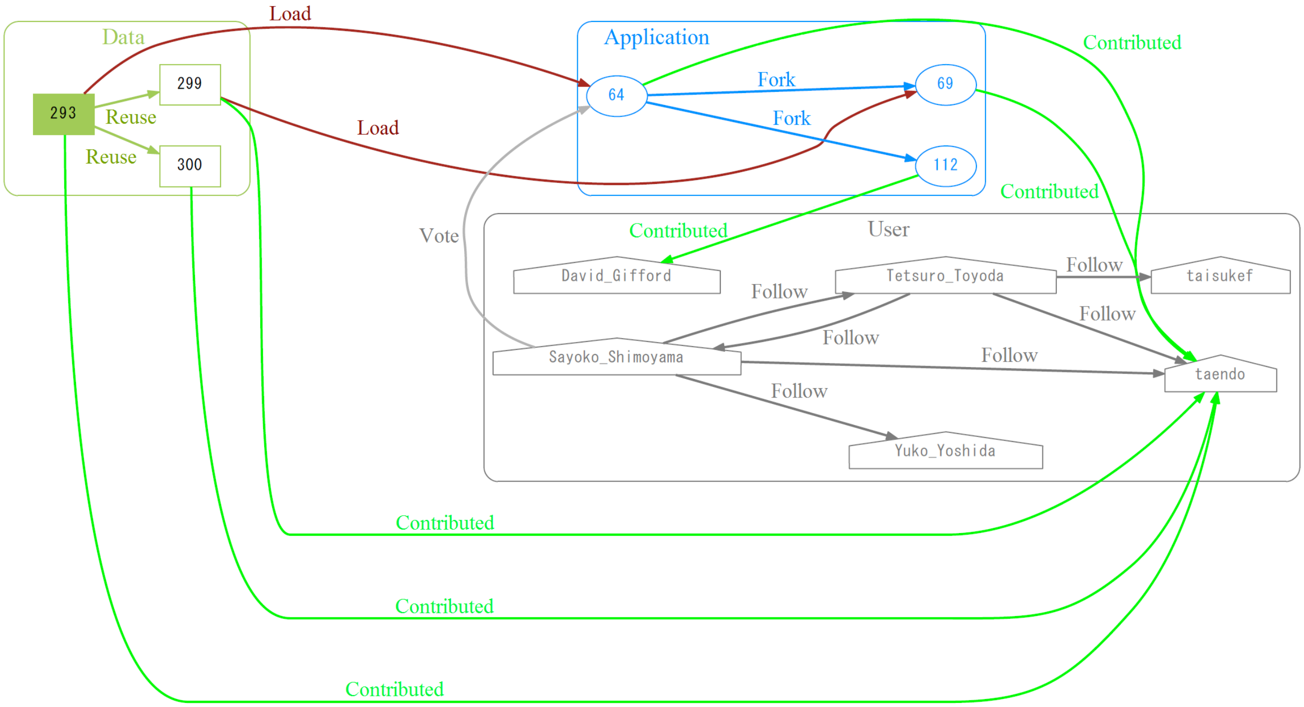


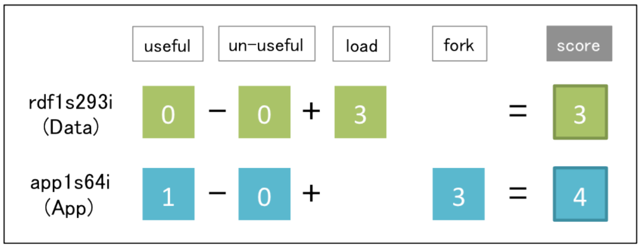
Fig. 3. **An example of dependency graph among Data, Apps and Users.** Dark Green edges indicate Data to Data **reuse** where a new data resource is published using a template of data published on LinkData, Red edges indicate Data to App **loading** in which a new application is created for data published on LinkData, Blue edges indicate App to App **forking** where new application is created by using program code of application created on LinkData. Bright green indicates User ownership **contribution** of a Data-set or an App. Grey edges indicate **votes** to rate various applications by users, and **following** of other users to receive updates of their activity and evaluations of their works.

PromoterCAD rests on a rich set of high throughput microarray and DNA sequence data containing over one million measurements and annotations of 20,000 genes. These were uploaded to LinkData as a series of data mashup tables and data rank lists (Fig. 6). Where other DNA design tools act as sequence editors with DNA specific functions, PromoterCAD is able to pull sequence data directly from the data sources in the LinkData system, guided by the menu-driven interface. PromoterCAD allows users to quickly perform advanced data queries, retrieve useful sequences, and organize them into their promoter sequence designs.

PromoterCAD also allows users to add their own knowledge of regulatory sequence data. Users may have literature knowledge of useful DNA sequences, so PromoterCAD allows these to be typed in and manipulated in the same manner as sequences retrieved from the LinkData sources. For example, one team of GenoCon2 participants introduced a DNA sequence that had been experimentally confirmed to confer dark inducibility to a plant gene. This sequence was combined with the LinkData to generate a DNA design they predicted to allow gene expression only in the flowering tissue of the plant, and only at night. In this way, PromoterCAD and LinkData allow expert users to combine their biological knowledge along with data mining operations from the LinkData sources.

The LinkData system provides code extensibility to PromoterCAD. With the forking function, users can write their own JavaScript data mining modules to PromoterCAD, and draw upon the rich linked data in new ways. For example, one participant in GenoCon2 modified a PromoterCAD function to display the top 10 expressing genes in a specific plant tissue. Other GenoCon2 participants used this module, and the forked utility has since been merged back into the main PromoterCAD functionality.

The architecture of PromoterCAD allows new LinkData sources to be added without any direct code modification. The LinkData forking system includes a flexible “Input Data” loading system. This allows users to control the LinkData that gets used for the PromoterCAD data mining tools. In a series of tutorials (http://promotercad.org), we clearly document how users can make their own LinkData tables and register them into forked versions of PromoterCAD. Examples are provided which explain how to add different types of experimental and



	useful		un-useful		load	fork		score
rdf1s293i (Data)	0	−	0	+	3		=	3
app1s64i (App)	1	−	0	+		3	=	4

Fig. 4. **Example of Calculating a score** that integrates several Data and Application activity and User rating ranking score.

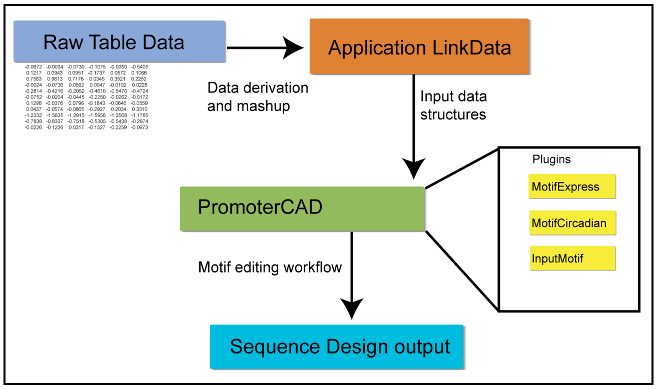


Fig. 5. **PromoterCAD LinkData system for DNA design incorporates database information with user knowledge** Overview of the PromoterCAD architecture. The source data is linked and then processed into a system of data suited to promoter design. PromoterCAD accesses this data, along with data that may be directly added by the user (user knowledge). The design workflow is similar to the revision history of a text, with each step recorded in the output. This allows for easy checking of the design and for collaboration.

sequence data. The data is structured with LinkData (Excel) templates so that users only need to copy and paste gene expression values or regulatory sequence lists. The uploaded data is then converted to RDF on the LinkData.org server. This function is intended to allow scientists who are not programmers to add their own databases to PromoterCAD. By replacing all of the data tables, a user could adapt PromoterCAD to design regulatory DNA in other organisms such as mouse, human, or bacteria.

PromoterCAD is also a learning tool. A special LinkData file contains pointers to external links, including the original data sources. This links appear directly in PromoterCAD web interface, so that users can quickly obtain more information about a particular gene or sequence. These include links to gene expression visualization websites such as the "electronic fluorescent pictogram" browser for *Arabidopsis* [2] and HanaDB [3]. This allows users to see illustrations of the gene expression patterns, which are presented in PromoterCAD as Highchart plots (http://www.highcharts.com).

Powerful tooltip functionality allows all LinkData sources to be annotated in a separate tooltip table. This provides guidance for the users who might not have familiarity with gene expression microarray data or promoter analysis. Furthermore, the tooltip files can be easily modified to create interactive tutorials for guiding users in promoter design.

This GenoCon system is used to empower Open Genomic Design, coupled with closed construction and safe experimental verification of the designed DNA sequences. The system of Linked Data driving Computer Aided Design, with evaluation by experiment, will foster a rapid biological knowledge cycle where programmers, researchers, and amateurs can all contribute.

Dependency Graph for GenoCon PromoterCAD: Here we show the cycle enhancing synergistic collaboration in this web-based scientific competition for synthetic biology promoter design. (Fig. 7) For example in Fig. 7 and Fig. 8 highly voted for and followed application app1s137i "A Promoter Design to Maintain the Fertility of Transgenic Plant by new Plugin MotifRanking" is a fork of app1s94i GenoCon PromoterCAD. App1s94i was forked by 8 apps and was voted for by 1 user for a total of 9 rank score. App1s137i was forked 0 times and was voted for by 5 users for a rank score of 5. In this fashion each app can be compared for total activity and usefulness in turn.

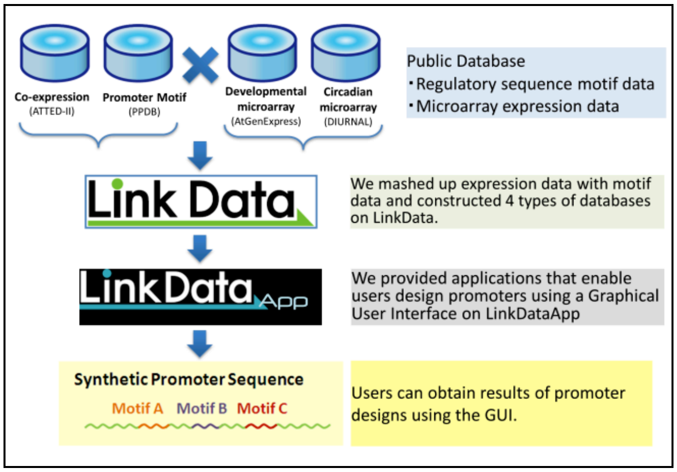

Co-expression (ATTED-II)
Promoter Motif (PPDB)
Developmental microarray (AtGenExpress)
Circadian microarray (DIURNAL)
Public Database
・Regulatory sequence motif data
・Microarray expression data
Link Data
We mashed up expression data with motif data and constructed 4 types of databases on LinkData.
LinkData App
We provided applications that enable users design promoters using a Graphical User Interface on LinkDataApp
Synthetic Promoter Sequence
Motif A Motif B Motif C
Users can obtain results of promoter designs using the GUI.

Fig. 6. **PromoterCAD database integration.** PromoterCAD uses several data sources for Tissue / Time specific promoter design. 4 types of sample data combine two promoter motif data, and two expression microarray experiments.

Contest Activity: The GenoCon2 promoter design contest generated active user groups and over 40 international submissions including from the USA, Egypt and Japan. Key users cooperated to create original designs that were modified and possibly improved by other users. Team collaboration was aided by the open nature of the design platform, and 13 promoter designs are being considered for final construction in transgenic plants. Application to further design challenge projects for other organisms is also planned.

V. Conclusions

Ease of generation of new applications on top of existing data is a practical benefit for scientists, with faster development making it potentially easy for other scientists to jump in at any step of a research process and test pre-existing data analysis, and more easily recreate to check what the original researcher has done. A major benefit of the LinkData platform for biological research is that unique analysis modules and database structures could possibly be reused for future and different organism related research.

Because dependency graphs combine both data and applications published on the platform, users easily create new applications for data and publish new data resources for use with applications. This yields a creative synergy cycle between data publication and application development. As a future plan we propose the use of the LinkData.org integrated database/application concept including dependency graphs to be applied for CKAN and other major repositories.

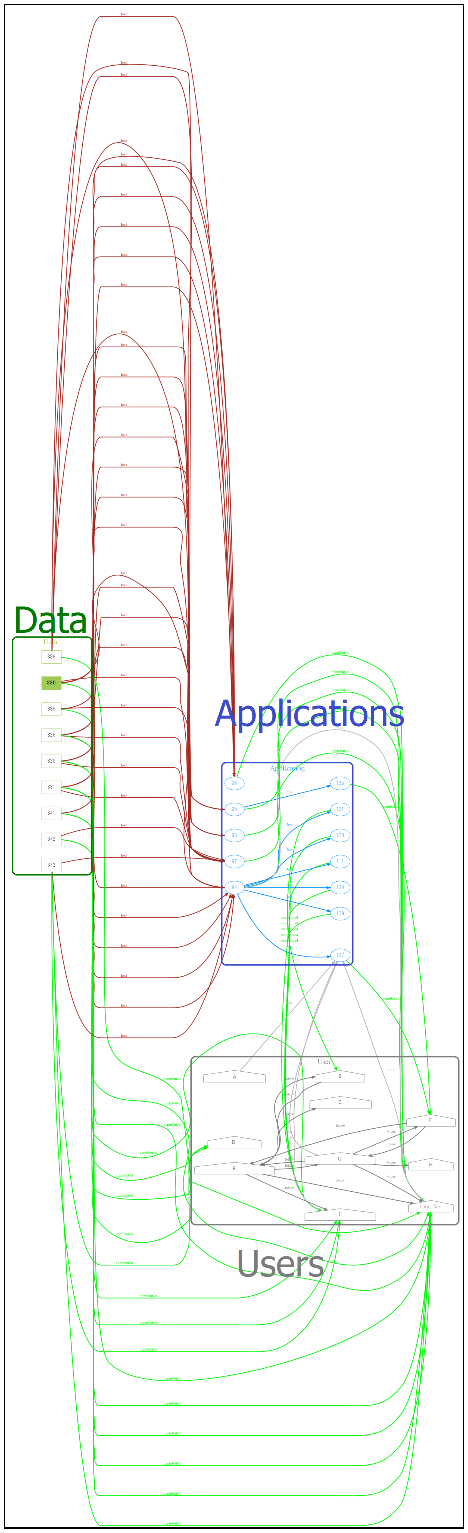
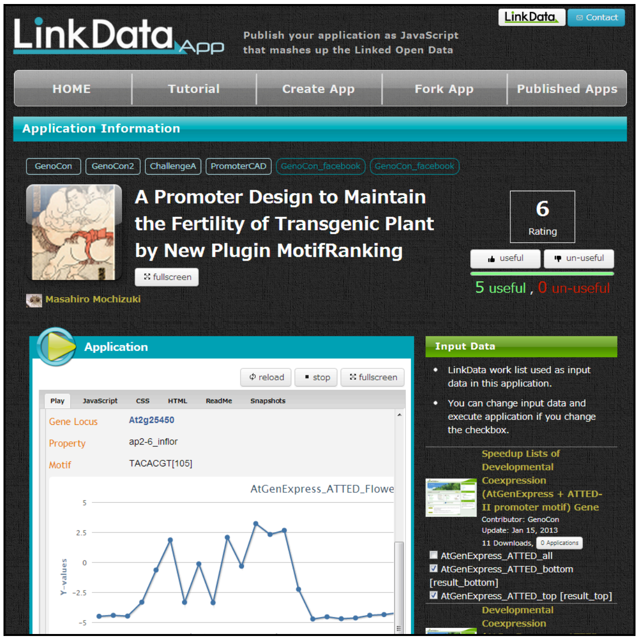


Fig. 7. **Dependency graph of the PromoterCAD application on LinkData.** This graph shows the interaction between the LinkData (Green color box), the Apps (Blue color box), and the Users (Grey color box). Red lines show the loading, creation of apps by specifying some particular Data . Blue lines indicate the forking of an App into a new App. Green lines show which users have created each Data or App. The Grey lines indicate the interest expressed in each Data, App or User by the Users. In this graph highly rated and followed application app1s137i is a fork of app1s94i which is ranked higher.



A Promoter Design to Maintain the Fertility of Transgenic Plant by New Plugin MotifRanking

6 Rating

5 useful , 0 un-useful

Masahiro Mochizuki

Gene Locus At2g25450
Property ap2-6_inflor
Motif TACACGT[105]

- LinkData work list used as input data in this application.
- You can change input data and execute application if you change the checkbox.

Speedup Lists of Developmental Coexpression (AtGenExpress + ATTED-II promoter motif) Gene
Contributor: GenoCon
Update: Jan 15, 2013
11 Downloads,

☐ AtGenExpress_ATTED_all
☑ AtGenExpress_ATTED_bottom [result_bottom]
☑ AtGenExpress_ATTED_top [result_top]

Developmental Coexpression

Fig. 8. **LinkData Application app1s137i showing usability ranking and user voting buttons on top right.** http://app.linkdata.org/app/app1s137i

ACKNOWLEDGEMENT

We would like to acknowledge: Dr. Takaho Endo for creating biological visualization tool on LinkDataApp. Ms. Yuko Yoshida for development of converter and valuable discussion. Dr. Shuji Kawaguchi for giving advice on score calculation. Dr. Koro Nishikata for testing LinkData. Dr. Masahiro Mochizuki for testing and adding the MotifRanking tool. Mr. Chanaka Perera, Mr. Uditha Punchihewa, Mr. Gayan Hewathanthri, Mr. Hiroaki Osada, Mr. Kazuro Fukuhara and Mr. Kiyoshi Mizumoto (Axiohelix Co., Ltd.) for web application and LinkData development. The committee of Linked Open Data Challenge Japan for continuing interest and encouragement.

This work was supported by the National Bioscience Database Center (NBDC) of the Japan Science and Technology Agency (JST).

REFERENCES

[1] T. Toyoda, et al.: “Methods for Open Innovation on a Genome – Design Platform Associating Scientific, Commercial, and Educational Communities in Synthetic Biology,” Methods in Enzymology., Vol. 498, 189-203, (2011)

[2] D. Winter, Ben Vinegar, H. Nahal, R. Ammar, G. V. Wilson, and N. J. Provart, “An ‘Electronic Fluorescent Pictograph’ Browser for Exploring and Analyzing Large-Scale Biological Data Sets,” PLoS ONE, vol. 2, no. 8, p. e718, Aug. 2007.

[3] K. Hanada, M. Higuchi-Takeuchi, M. Okamoto, T. Yoshizumi, M. Shimizu, K. Nakaminami, R. Nishi, C. Ohashi, K. Iida, M. Tanaka, Y. Horii, M. Kawashima, K. Matsui, T. Toyoda, K. Shinozaki, M. Seki, and M. Matsui, “Small open reading frames associated with morphogenesis are hidden in plant genomes.,” Proc Natl Acad Sci USA, Jan. 2013.