

# 国立情報学研究所の オープンデータに関する取り組み

大向 一輝

@i2k

# NIIを取り巻く「オープン」

- オープンアクセス
  - 図書館コミュニティの一員として
- オープンガバメント
  - 公的機関として
- オープンデータ
  - データプロバイダーとして

# さまざまな「オープン」

- CiNii goes Open@図書館総合展2008
  - ユーザのオープン化
- コンテンツサービスのウェブAPI提供
  - CiNii Articles
  - KAKEN
  - JAIRO
  - CiNii Books
  - 機能のオープン化

# 機能のオープン化

- OpenSearchとRDF
  - 検索結果・書誌データの構造化
- JSON-LD API **NEW!**
  - JSONによるLinked Data表現
- クロスドメイン通信 **NEW!**
  - クライアントサイドでのアプリ開発が可能に

# CiNiiの書誌データ (RDF/XML)



The screenshot shows the CiNii search results page for the book 'セマンティックWebプログラミング' (Programming the semantic web). The page includes a search bar, a list of authors (Segaran, Toby; Evans, Colin; Taylor, Jamie (Edwin James); 玉川, 竜司; 大向, 一輝; 加藤, 文彦; 中尾, 光雄; 山本, 泰智), and a list of university libraries that hold the book. The book cover is also visible, featuring a cat.

<http://ci.nii.ac.jp/ncid/BB02488158>, **rdf:**

```
<rdf:Description
rdf:about="http://ci.nii.ac.jp/ncid/BB02488158#entity">
  <foaf:isPrimaryTopicOf
rdf:resource="http://ci.nii.ac.jp/ncid/
BB02488158.rdf"/>
```

```
<dc:title>セマンティックWebプログラミング</dc:title>
<dc:title xml:lang="ja-hrkt">セマンティック Web プログラミ
ング
```

```
</dc:title>
<dcterms:alternative>Programming the semantic web
</dcterms:alternative>
<dc:creator>トビー・セガラン著 ; 玉川竜司訳</dc:creator>
<dc:publisher>オライリー・ジャパン</dc:publisher>
<dc:language>jpn</dc:language>
<dc:date>2010</dc:date>
<foaf:topic
```

```
rdf:resource="http://ci.nii.ac.jp/books/search?q=
セマンティックウェブ" dc:title="セマンティックウェブ"/>
<cinii:ncid>BB02488158</cinii:ncid>
<dcterms:hasPart rdf:resource="urn:isbn:9784873114521"/>
</rdf:Description>
```

```
<rdf:Description
rdf:about="http://ci.nii.ac.jp/ncid/BB02488158#entity">
  <foaf:maker>
  <foaf:Person
rdf:about="http://ci.nii.ac.jp/author/DA15839119">
  <foaf:name>大向, 一輝</foaf:name>
  <foaf:name xml:lang="ja-hrkt">
```

# CiNiiの書誌データ (JSON-LD)



The screenshot shows the CiNii website interface. At the top, there's a search bar with the text 'セマンティックweb' and a search button. Below the search bar, there are several author names listed, including Segaran, Toby, Evans, Colin, Taylor, Jamie (Edwin James), 玉川, 竜司, 大向, 一輝, 加藤, 文彦, 中尾, 光輝, and 山本, 泰智. The main content area displays the book 'セマンティックWebプログラミング' by Toby Segaran, Colin Evans, and Jamie Taylor, translated by Ryūji Tamakawa. The book cover features a cat. The page also shows a list of university libraries that have the book, such as Aichi University of Science and Technology, Aichi University, and Keio University. The book details include the title 'Programming the semantic web', the subtitle 'Build flexible applications with graph data', and the publisher 'O'Reilly'.

```
{ "@context":  
  { "owl": "http://www.w3.org/2002/07/owl#",  
    "dc": "http://purl.org/dc/elements/1.1/",  
    "dcterms": "http://purl.org/dc/terms/"  
  },  
  "@id": "http://ci.nii.ac.jp/ncid/BB02488158.json",  
  "@graph": [ { "@id": "http://ci.nii.ac.jp/ncid/BB02488158#entity",  
    "foaf:isPrimaryTopicOf":  
      { "@id": "http://ci.nii.ac.jp/ncid/BB02488158.json",  
        "dc:title": [ { "@value": "セマンティックWebプログラミング" },  
          { "@value": "セマンティック Web プログラミング",  
            "@language": "ja-hrkt" } ],  
        "dcterms:alternative": [ "Programming  
the  
semantic web", "Build flexible applications with graph  
data" ],  
        "dc:creator": "トビー・セガラン著 ; 玉川竜司訳",  
        "dc:publisher": [ "オライリー・ジャパン", "オーム社 (発売)" ],  
        "dc:language": "jpn",  
        "dc:date": "2010",  
        "cinii:ncid": "BB02488158",  
        "cinii:ownerCount": "117",  
        "foaf:maker": [  
          { "@id": "http://ci.nii.ac.jp/author/DA15839119#entity",  
            "@type": "foaf:Person",  
            "foaf:name": [  
              { "@value": "大向, 一輝" },  
              { "@value": "オオムカイ, イツキ", "@language": "ja-hrkt" } ] ]  
        ]  
      }  
    ]  
  ]  
}
```

<http://ci.nii.ac.jp/ncid/BB02488158.json>

- 技術面の達成度は高い
  - 元データが構造化されており、機械的に変換するだけ
  - ウェブAPIの普及
- ライセンスはこれから
  - 現在は紳士協定・黙認（？）状態
  - トラブルには個別対応

# 5 ★ Open Data

Tim Berners-Lee, the inventor of the Web and Linked Data initiator, suggested a 5 star deployment scheme for Open Data. Here, we give examples for each step of the stars and explain costs and benefits that come along with it.



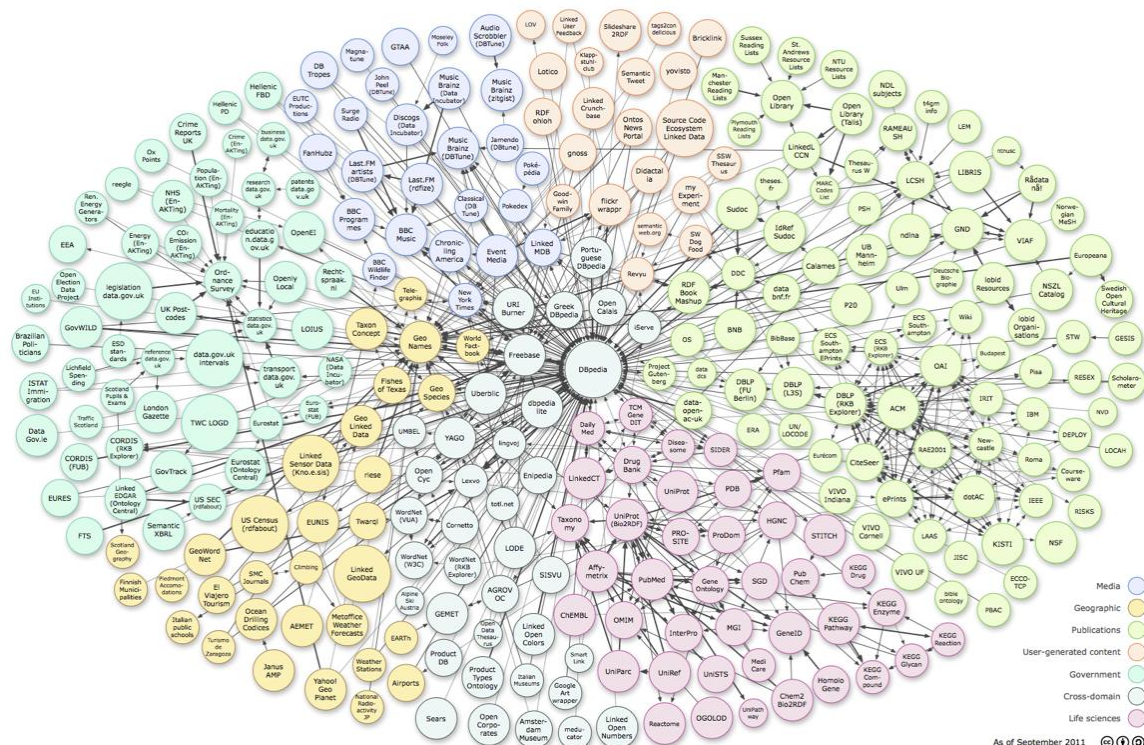


# 利用のオープン化

- オープンライセンスの採用
  - 多様なステイクホルダーとの議論が必要
    - 何を対象とするか？
    - 何の情報を対象とするか？
    - 情報に著作権はあるか？
    - どのライセンスを選択するか？
  - 意思決定のプロセス
    - NIIが単独で決定できることは「ない」

# 知識のオープン化

- 外部情報源とのリンク
  - ID・典拠：信頼性・継続性の高い「リンク先」
  - 国内連携
  - 国際連携
  - 他分野連携
- SPARQL対応



As of September 2011 © ① ②

# みなさまへのお願い

- オープンデータを知ってください！
  - 日経オープンデータ情報ポータル
  - NHK NEWS WEB
  - Open Knowledge Foundation Japan
  - Linked Open Data Initiative
- オープンデータを使ってください！
  - LODチャレンジ
  - ハッカソン・アイデアソン・街歩き

# みなさまへのお願い

- オープンデータ化を応援してください！
  - 多様なステイクホルダーとの議論
  - 意思決定者への働きかけ
    - 「うちはまだオープンデータやらないんですか？」

# データを公開しない理由 (1/5)

- 無関心
  - 誰も気にしていない
  - コミュニティからの支援が得られない (市民ハッカーがいない)
- 混乱
  - 明確な事例がない
  - データを何のために使うのかわからない
  - どんなデータが必要とされるのかわからない
  - データを公開する前に、どんな便益があるか明確にしてほしい
  - 今のシステムで十分だ
  - 人々を混乱させるだけだ

# データを公開しない理由 (2/5)

- 困難
  - とにかく複雑で難しい
  - 大量のバックログを抱えている
  - ついていけない、時間がない、変化が早すぎる
  - データの量が膨大である
  - 自分の時間の大部分が取られてしまう
- 費用
  - 予算がない
- スタッフ問題
  - それは私の仕事ではない
  - 私の上司や政治家はデータ公開を望んでいない
  - やり方を知っている人がいない

# データを公開しない理由 (3/5)

## 合法性

- データや記録を公開する法的な権限を持っていない
- このデータを公開すれば、市民は私たちを訴える
- 私たちはデータに対する権利を持っていない
- 私たちがデータの所有者なのかどうなのかわからない
- データは販売するべきである。もし売りたいになったらどうすればいいのか？すでに売っている
- 機密データではないが、公開することでよいPRになるとは思えない
- 私たちの弁護士は特別なライセンスを付与するべきだと言っている
- このデータを集めるのに、膨大な時間・人・資源を投入した
- なぜデータを必要としているのかを知りたいので、直接窓口に来てほしい
- 誰かがパッケージングし直して売り払ってしまうのではないか
- データをだれが利用して、何を行ったのかをコントロールしたい
- 連邦政府や他の政府機関がすでにデータをもっているので、そちらから入手して欲しい

# データを公開しない理由 (4/5)

- 正確さ
  - データを一括して公開すると、だれかがそれを変更してしまう
  - データを公開したりすると、私たちがハックされる
  - データを公開すると、やり方によっては人々が誤解する危険がある
  - 市民が混乱する、データの品質がそもそも高くない
- プライバシー
  - プライバシー保護の懸念があり、データを公開できない
  - もう既に公開している
  - 窓口やプロバイダーを通じてすでに公開している
  - 情報公開制度があるではないか
  - すでに十分にデータは公開している
  - APIを提供しているではないか
  - データポータルが必要なら、同じようなものはすでにある



# データを公開しない理由 (5/5)

- その他
  - データが優秀な技術者集団や高等教育を受けたコミュニティに利用されると、デジタルデバイド問題がさらに深刻化する
  - 今、それは優先事項ではない
  - それについてはペンディング状態だ、似たようなプロジェクトもすでに始まっている

出典: Reasons to Not Release Data, Part 1-Part 10, Sunlight Foundation  
(日本語訳: 「公的機関がデータを公開しない理由」 Open Knowledge Foundation Japan)

# データを公開する理由

- 一般論
  - 公共の資産
  - 透明性
  - イノベーション
- 学術情報サービスとして
  - 信頼性の高いデータをウェブに放流する
  - PRとフィードバック
  - 新規顧客・市場開拓
  - 「（まだ見ぬ）他者とのコミュニケーション」