

データアーカイブ化に対する データ提供側の悩み ～確実なデータ提供体制を築く為に～

篠原 育 (ISAS/JAXA)

今日の論点

- “Data Provider”と“Data Archiver”は対立するものなのか？
 - 真の対立相手はEnd User…？
- 科学データの品質・信頼性向上は誰が担うのか？
 - 科学者の参加が必要
- データシステムは便利に使えさえすればそれでよいのか？
 - データに付加価値をつける
- 科学データにとって Data Provider – Archiver – End User のあるべき姿は…

牧歌的な時代のデータ利用

- Data Provider から End User は直接顔が見えていた。
- End User はデータ利用の敷居が高くても利用した。
- 相互にrespectがあった。

喜んで！でも、使い方にはいろいろ注意点があるので勉強してもらう必要があるよ。

確かにそうだね。データ評価も手伝ってもらえるかな？

ありがとう！！
おかげで科学成果もあがったし、データの品質もあがったよ！

データを使わせてください！

もちろん！沢山質問するからね。

このデータは何かがおかしいよ？
もちろん、手伝うよ！

おかげでいい論文が書けたよ！
共著者にはいってね。

データも使いやすい形で整備したから、よかったら使ってみてね。

ネット公開時代のデータ利用

- Data Provider から End User の顔は見えなくなった。
- End User はデータ利用の敷居が高いと利用しない。
- 相互respectが失われた。

使い方にはいろいろ注意
点があるので勉強して
もらう必要があるよ。
マニュアルはないんだ。

全部は見切れていないんだ。
データ評価も手伝って
もらえるかな？

みんな論文書くことばかりで
データ評価を手伝ってくれないよ。
馬鹿馬鹿しくてやってられない…

公開データに質問
あるのだけれど。

えー？面倒くさいなあ。
マニュアル整備してよ？

公開データなんだから、怪しい
データなんてないよね？

いやだよ。 そんな時間があつたら
論文を書かなくっちゃ。
観測屋の義務でしょう？

使いにくいデータなん
か使わなくてもいいか。

Data Provider と Archiver の対立？

わかっているけれど、対応には限界があるし、そもそも悪い品質のデータを公開したくないよ。

それもわかっているけれど、公開してしまえばどんな使い方をするかわからないよ…

ユーザ対応に追われて自分の研究がまったくできないよ。どういう使われ方をするかもわからないし、データ公開はよいことばかりでないよ…**データ屋は催促するばかりで気楽なものだな。**

公開データは義務だから早くデータを納品してね。

品質はユーザの目に触れればだんだんよくなるよ！

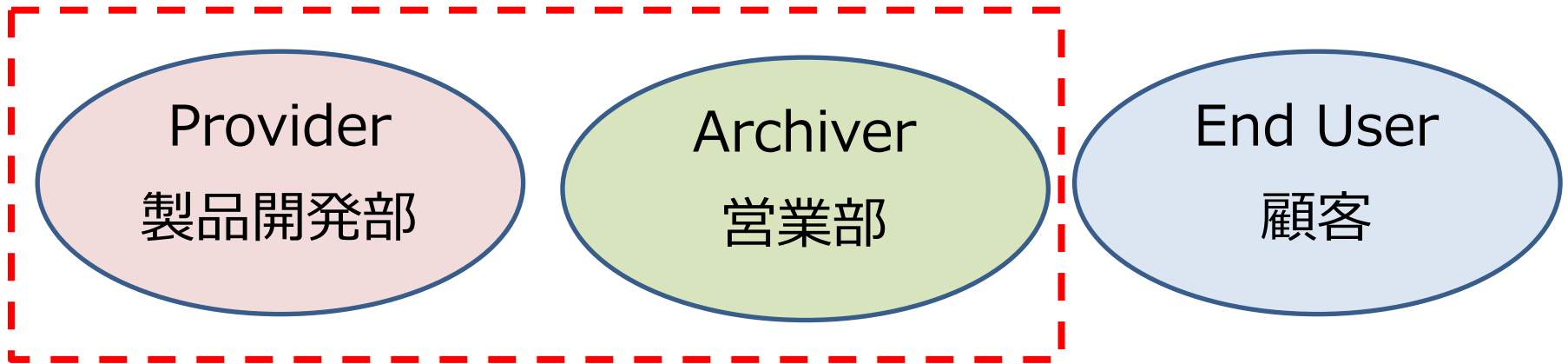
それから、ユーザーからの問い合わせにも答えてね。利用に必要な書類も整備してね。

データがなかなか出てこないし、書類も整備されていないから、なかなかユーザがつかないよ…**観測屋がしっかりしないからデータ・サイエンスがすずまないよ。**

リソースのあるアメリカはいいなあ！！

Data Provider と Archiver の対立？

- Data Archiver の言い分は、実は End User の要望を代弁しているだけ。
- ユーザ・サービスという事業を抱えている分だけ、End User よりも Data Provider への要求は厳しめ。



ネット公開時代のデータユーザは一般消費者と同じアレ？ 科学者が科学データの単なる消費者でいいの？

製品開発部と営業部って同じ会社だよね？

Data Provider 側の意識

- データを秘匿しておきたいわけではまったくなく、むしろ、多くの人にデータを利用して欲しい。
- できるだけ多くの人々の目に晒されることで、データのクオリティーが向上する。
- データを公開することによって、科学成果は飛躍的に増大する。
- しかし、データ処理に必要なリソース（特にマンパワー）は不足しているので、すべてのリクエストには応えられない。
- 怪しいデータを使って、ヘンな解析をして欲しくない。
- 較正や評価は手伝って欲しいが、必ずしも誰にでもできる訳ではない。
- 自分たちの研究者なのだから、自分の観測機器のデータを使って研究がしたい。
- データ取得・処理・公開に払った努力についてはrespect (authorship や acknowledgement) して欲しい。

このような主張に膠のない反論はできるけれども、問題はなにも解決しない。

科学データの評価は誰が担う？

- 観測機器のことをよくわかっている装置開発者がすべてを行うことができれば理想的。しかし、できることには限りがある。
- 較正などに必要な情報が公開されていれば、ユーザも較正作業に参加できる。多くのユーザが較正や評価に参加できれば、装置開発者が想定していなかった新しい手法によってデータ品質を向上させられるかもしれない。
- X線天文では較正作業自体が研究成果になり得る。
 - 機器チームがノイズを除いたデータを公開する（利便性を重視）のか、ノイズを引かないデータ+ノイズ除去アルゴリズムを公表する（処理の発展性を重視）のか、過去のデータ公開時に論争があったのだとか。結局、後者が選ばれた。

ユーザの利便性のみを追求すると、ダウンロード・サイエンティストを増やしてしまう。データ解析の専門家ならばデータの目利きぐらいは行えなくてははいけないのでは？（牧歌的な時代にはユーザにその訓練がなされた。）

科学データの評価は誰が担う？

- 2種類の End User のタイプ
 - 専門家としてデータの目利きができるコア・ユーザ
 - 周辺分野のデータ利用の裾野を広げることを担うエンド・ユーザ
- データにも2種類ある.
 - 専門的な知識がないと使えないデータ（より生に近いデータ）
 - 誰でも（ほとんど背景知識がなくとも）使えるデータ
 - すべての取得データを後者にまでプロセスできる…？
- 専門家なのにエンド・ユーザ？
 - データを消費するだけのユーザが増え過ぎると、データ提供側が疲弊し、巡り巡ってその研究分野のアクティビティが弱くなってしまふ。
 - 消費するばかり（ダウンロード・サイエンスのみ）ではダメ…

STP分野では、機器校正にデータ解析ユーザが参画するケースはそれほど多くなかった。（機器開発者以外に）**データ評価や校正に関わるコア・ユーザを増やす努力をすることが急務ではないか。**

データシステムの方向性

- 現状のデータシステムの整備の方向性は対End Userサービスに偏っている？
 - 「データ利用の裾野を広げる」という目的の為には必要不可欠なサービスだが、そのサービスを支えるデータ提供体制が不足.
 - データ提供体制を強化するには、コア・ユーザを増やす必要.
 - コア・ユーザは、データ品質向上の観点から、フィードバックをData Providerに戻したり、エンド・ユーザサポートを担う.
 - コア・ユーザと Data Provider の連携を促進する仕組みもデータシステムのスコープ.
 - ユーザからのフィードバックとして、高次データ処理の成果を収集することもデータシステムの大切な機能.
 - Data Provider と Archiver は敵同士ではなく、同志！
- データを消費するだけでない、データの品質を向上し、さらに付加価値をつける仕組み

Provider-Archiver-Userの理想の関係

- ネット公開時代では、ユーザからデータを消費されるだけではProviderもArchiverも疲弊してしまう。
- 利用促進として裾野を拡大するだけでなく、根本的なデータ提供体制を強化しなければならない。
- コア・ユーザ層を開拓し、データ提供側にもデータ利用側がデータ評価に深くコミットできる雰囲気作り。
- Provider -> Archiver -> User の一方通行の流れではない、双方向の情報の流れを作ることによって豊かなデータ提供体制を築く。

ERGプロジェクトでは

- End User向けサービスは、ERG-SCが中心となって、データ公開、ユーザサポートにこれまでにない体制を構築しつつあります。
- ERGでは9つもの科学観測機器を搭載しており、データ提供体制は必ずしも十分とはいえません。
- ERG-SCの準備したユーザ利用態勢に見合う以上のデータ提供体制を作らなければなりません。
- ERGのサイエンスWGメンバーには、様々な形でデータ品質向上やデータ付加価値付与などへの貢献をお願いしていきたいと考えています。
- 機器チームには、データ評価・校正への参加に対する敷居を下げる雰囲気作りをお願いしていきます。
- ERG衛星のみならず、将来のSTP衛星を実現し、成果をより確実にあげていくために必要な取り組みなので、コミュニティの皆さんのご協力をお願いいたします。