

CGIクエリでコントロールする ウェブシステム開発の試み

今井 弘二(NICT), 村山 泰啓(NICT), 海老沢 研(ISAS)
池田 大輔(九州大学), 瀬口 大介(九州大学)

1. オープンサイエンスの流れと分野横断型研究

- 観測機器や計算機の小型化・高精度化・低コスト化により、多量のデータが蓄積されており、それらのデータ資源のオープン化、あるいは共有化への取り組みの一環として、**オープンサイエンス(オープンデータ)**が叫ばれている。
- 例えば、地球惑星科学の分野では、宇宙・地球環境を総合的に理解することが最大課題の一つであるが、多分野のデータが容易に扱えるようになれば、大きな前進となる。

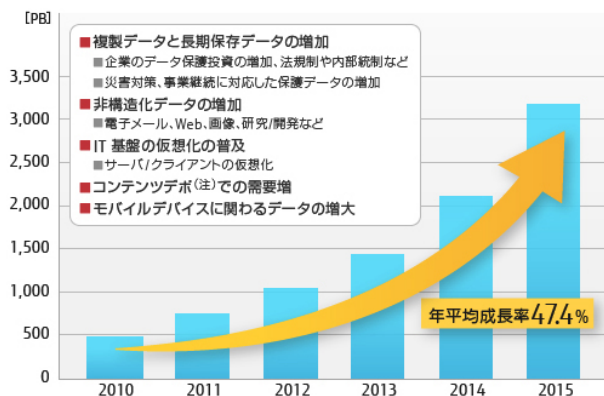
⇓ しかし

- 各学術分野ではデータ形式が異なるため、**多分野のデータ**を取り扱うことは難しく、**分野横断型研究**も容易ではない。

⇓ そこで

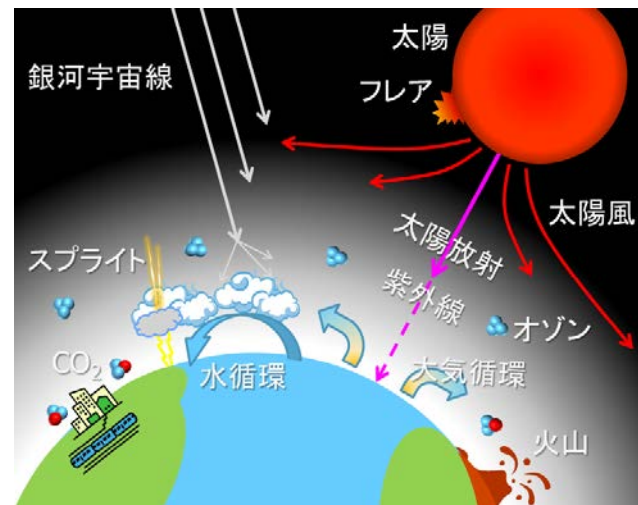
- データの利活用を促進するために、国際組織(RDA; Research Data Alliance、ICSU-WDS、OECD、G8/G7等)が課題解決を目指して取り組んでいる。

2010年を基準とした場合、5年後の2015年の外付けディスクストレージのデータ量は約7倍



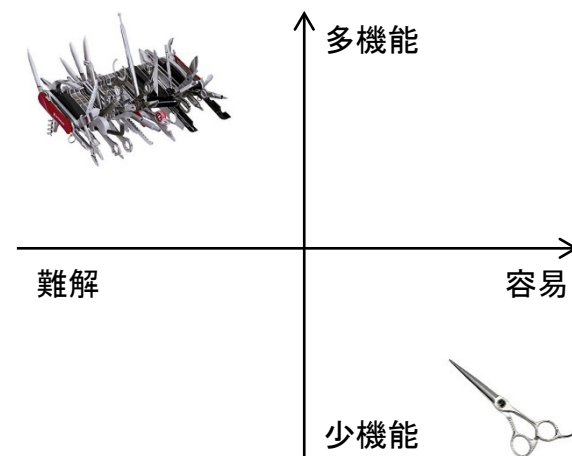
(注) コンテンツデポとは、多様なデータやコンテンツを収集し、組織化した巨大なコンテンツの貯蔵庫へのアクセスを提供するサービスプロバイダーの総称のことです。

出典：IDC Japan, 12/2011 国内ディスクストレージシステム市場 2011年上半期の分析と2011年～2015年の予測 (J11420110)



2. 多分野を扱うサービスを提供するための課題

1. 各学術分野で点在するデータをどう扱うか？
2. 多種多様なデータを処理するシステムを構築するか？
3. 多種多様なデータを表示するビューワを開発するか？
4. 豊富な機能と利便性のどちらを優先とするか？
5. データやチャートの共有・引用はどうするか？



3. 識別子を用いた引用について

DOI(Digital Object Identifier; デジタルオブジェクト識別子)とは

- 解決可能、持続可能、相互運用可能なリンクを提供するための国際的な識別子である。
- DOI登録機関を通じて登録される。
- 構文は **Prefix** と **Suffix** で構成される。

dx.doi.org/**10.1002**/**2015GL063323**

- DOIの発行件数は**1億**件（2015年時点）

CrossRef(1999年～)	7,300 万件
------------------	----------

Japan Link Center(2012年～)	300 万件
---------------------------	--------

⇒ ファイルの増殖とカオス化を回避する。

4. データ引用についてのシステム要求

Data Citation of Evolving Data



Recommendations of the Working Group on Data Citation (WGDC)

Andreas Rauber, Ari Asmi, Dieter van Uytvanck and Stefan Pröll

R1 – Data Versioning

R3 – Query Store Facilities

R5 – Stable Sorting

R7 – Query Timestamping

R9 – Store Query

R11 – Landing Page

R13 – Technology Migration

R2 – Timestamping

R4 – Query Uniqueness

R6 – Result Set Verification

R8 – Query PID

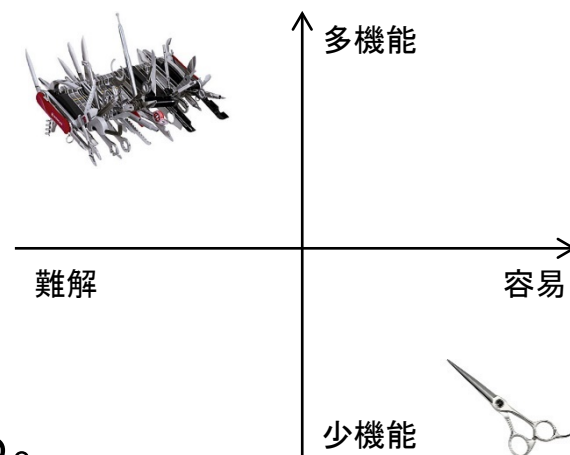
R10 – Automated Citation Texts

R12 – Machine Actionability

R14 – Migration Verification

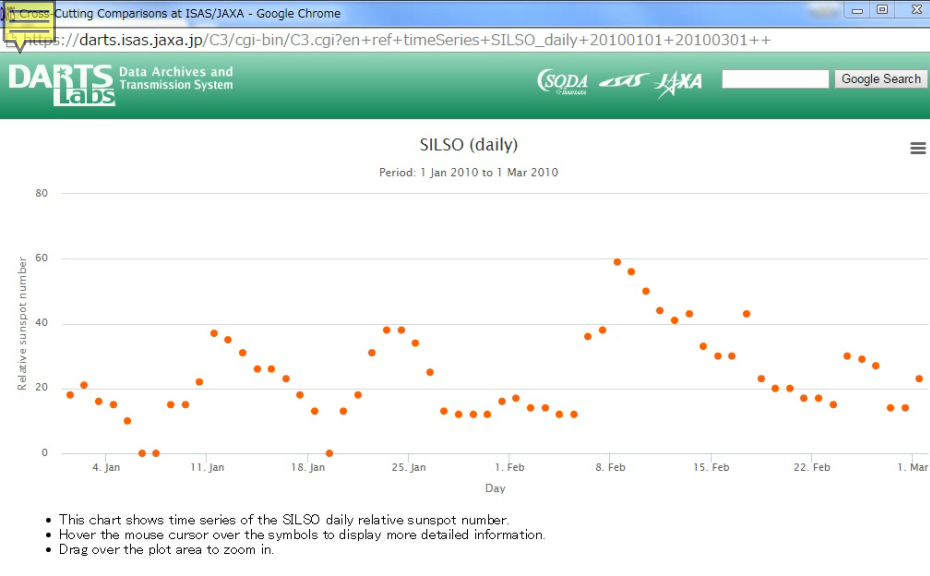
2. 多分野を扱うサービスを提供するための課題

1. 各学術分野で点在するデータをどう扱うか？
⇒ データを軽くして保有し、その他はネット経由で提供。
2. 多種多様なデータを処理するシステムを構築するか？
⇒ 生データではなく、高次データのみを取り扱う。
3. 多種多様なデータを表示するビューを開発するか？
⇒ 車輪の再開発はせず、オープンソースを利用する。
4. 豊富な機能と利便性のどちらを優先とするか？
⇒ 可能な限り豊富な機能と利便性を両立させる。
5. データやチャートの共有・引用はどうするか？
⇒ CGIクエリでコントロールする。
6. 保守・運用はどうする？
⇒ ウェブベースのアプリとして、DARTSにただ乗りする。

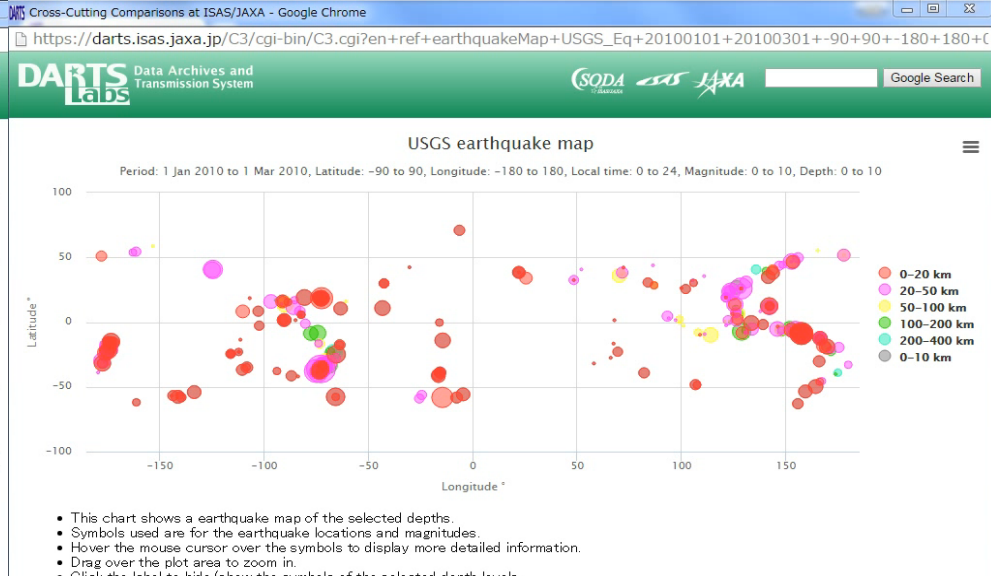


などなど…を予算0円、開発者1名でどう実現する？

⇒ 多種多様なデータ形式が混在する地球惑星科学を
テストケースとしてサービスの構築を試みた。



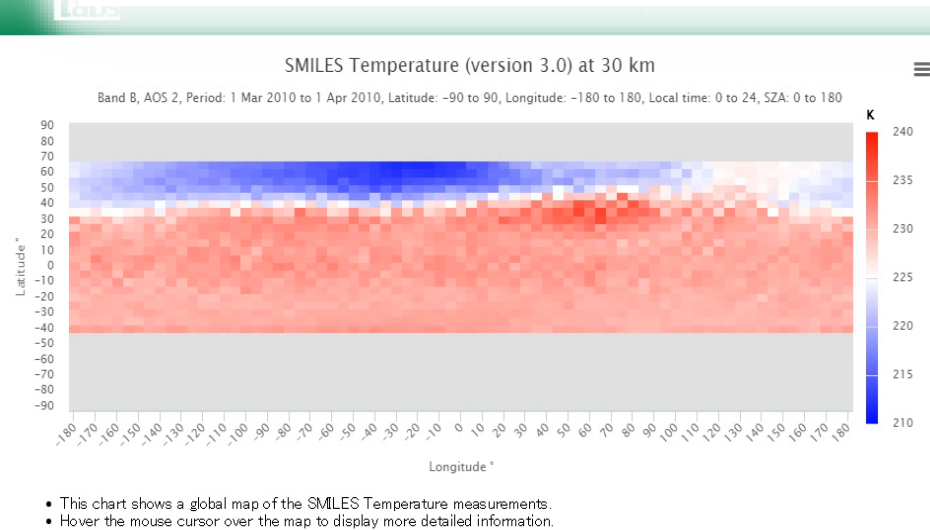
Please send questions or comments to "darts-admin AT ML.isas.jaxa.jp".
DARTS is maintained by C-SODA at ISAS/JAXA
Created date and time: 05/17/2016 18:43:37



Please send questions or comments to "darts-admin AT ML.isas.jaxa.jp".
DARTS is maintained by C-SODA at ISAS/JAXA
Created date and time: 05/17/2016 18:52:21

(URL CGI プログラム)?(ヘッダ)+(データ選択部)+(データ抽出部)

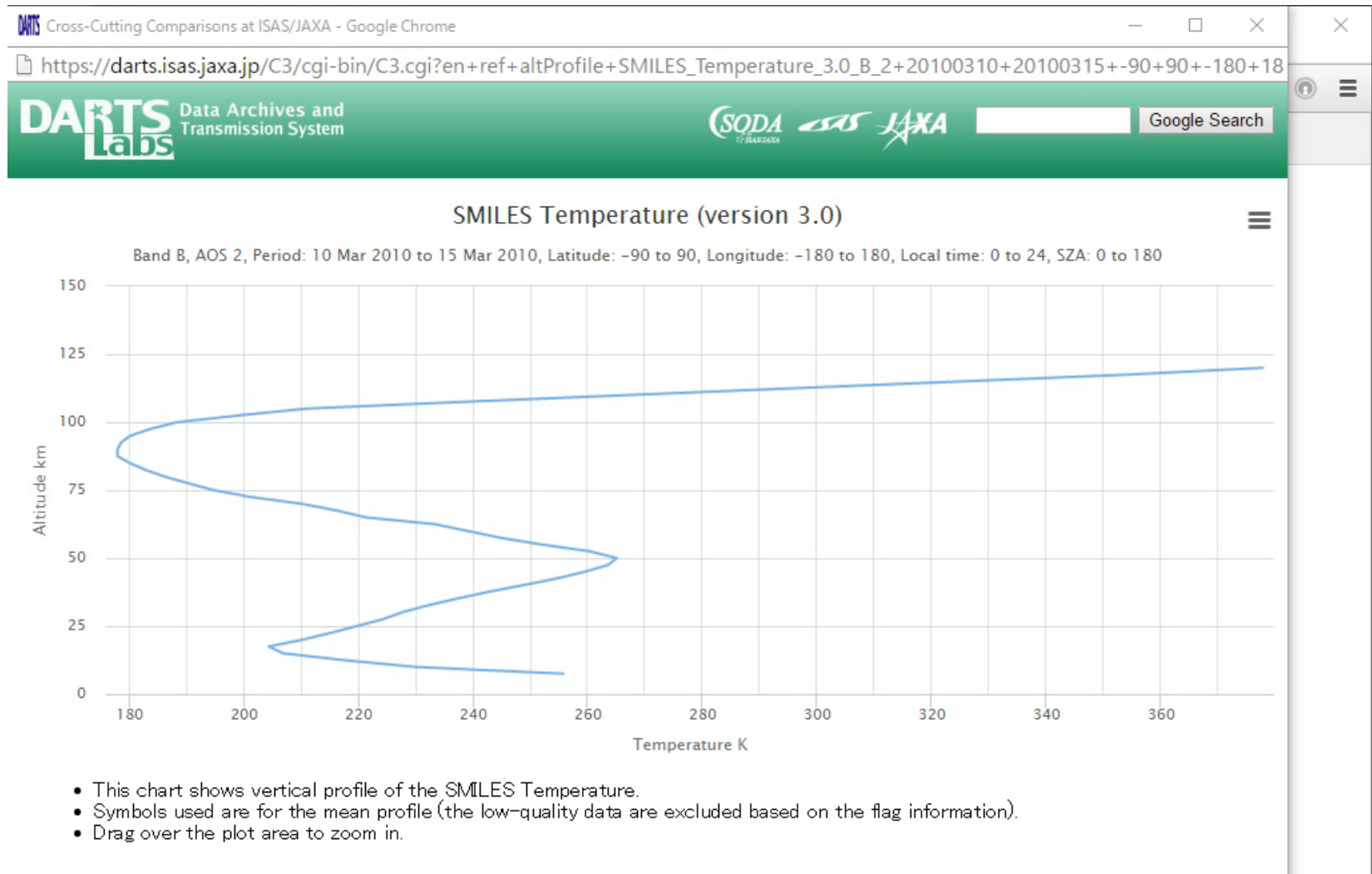
darts.isas.jaxa.jp/C3/cgi-bin/C3.cgi?en+SORCE_TSI_17_timeSeries+20100101_20110101



Please send questions or comments to "darts-admin AT ML.isas.jaxa.jp".
DARTS is maintained by C-SODA at ISAS/JAXA
Created date and time: 05/18/2016 22:09:27

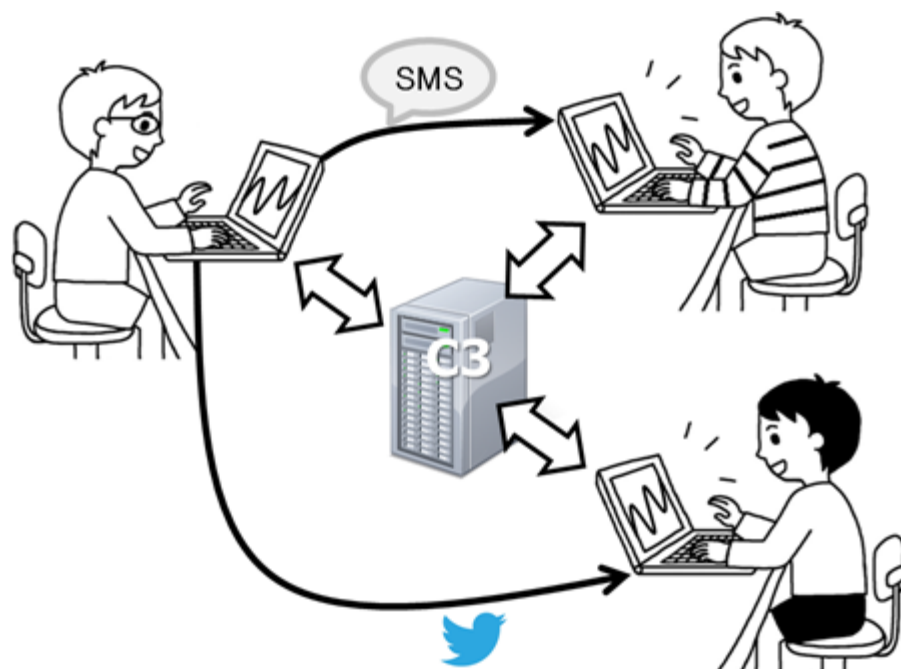


5. CGIクエリを用いたチャートの再現



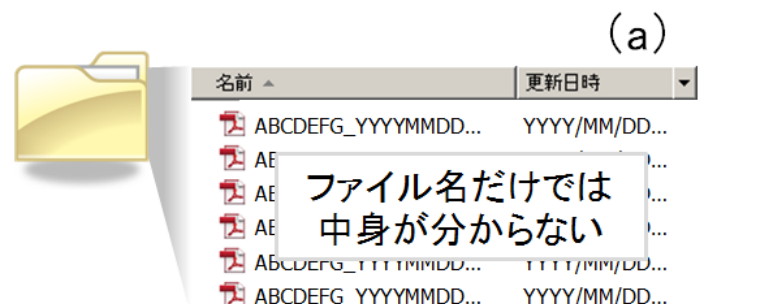
Please send questions or comments to "darts-admin AT ML.isas.jaxa.jp".
DARTS is maintained by C-SODA at ISAS/ JAXA.
Created date and time: 05/20/2016 07:49:59

6. CGIクエリを用いたチャートの共有と引用

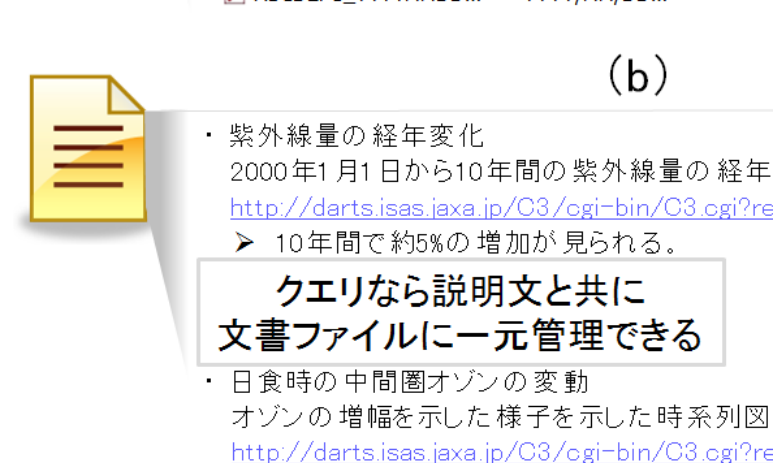


チャートを共有することができる。

(a)



(b)



- ・ 紫外線量の経年変化
2000年1月1日から10年間の紫外線量の経年
<http://darts.isas.jaxa.jp/C3/cgi-bin/C3.cgi?re>
➤ 10年間で約5%の増加が見られる。
- ・ 日食時の中間圏オゾンの変動
オゾンの増幅を示した様子を示した時系列図
<http://darts.isas.jaxa.jp/C3/cgi-bin/C3.cgi?re>

文書に引用することができる。



7. データ引用についての要求とその対応策

- R1 – Data Versioning (適用外, OK) R2 – Timestamping (適用外)
- R3 – Query Store Facilities (OK) R4 – Query Uniqueness
- R5 – Stable Sorting R6 – Result Set Verification
- R7 – Query Timestamping R8 – Query PID
- R9 – Store Query (OK) R10 – Automated Citation Texts (OK)
- R11 – Landing Page (OK) R12 – Machine Actionability (OK)
- R13 – Technology Migration R14 – Migration Verification

RX: データに関する要望で、C3は対象外である。

RX: C3はクエリをデータベースで管理せず、リアルタイム処理をしている。

RX: C3はクエリを自動生成している。

RX: C3のページそのものが“Landing Page”であり、クエリ構文に可読性を持たせているので必要ない。

RX: クエリにタイムスタンプを付けて解決する。

RX: タイムスタンプとチェックサムで解決する。

(URL CGI プログラム)?(ヘッダ)+(データ選択部)+(データ抽出部)

darts.jaxa.jp/C3/cgi-bin/C3.cgi?(タイムスタンプ, チェックサム)+SORCE_TSI_17_timeSeries+20100101_20110101



システムはタイムスタンプで管理し、
クエリ自体はチェックサムで確認する。

多種多様なデータ形式が混在する地球惑星科学をテストケースとして、サービス(C3)の構築を試みた。

システムをCGIクエリでコントロールすることで、

1. チャートの引用
2. チャートの編集(識別子の作成)
3. チャートの共有

また、クエリにタイムスタンプを加えることで、

識別子の Uniqueness、Sorting、PID、Technology Migration の情報を持ち、チェックサムを合わせて、アプリのVerificationを保証する。

オープンサイエンス時代に、多分野のデータを取り扱うサービスの一つの形態として、CGIでコントロールするウェブシステムが考えられる。

<謝 辞>

本プロジェクトは、宇宙科学研究所 科学衛星運用・データ利用センター(C-SODA)の助成を受けて実施しております。ここに感謝の意を表します。