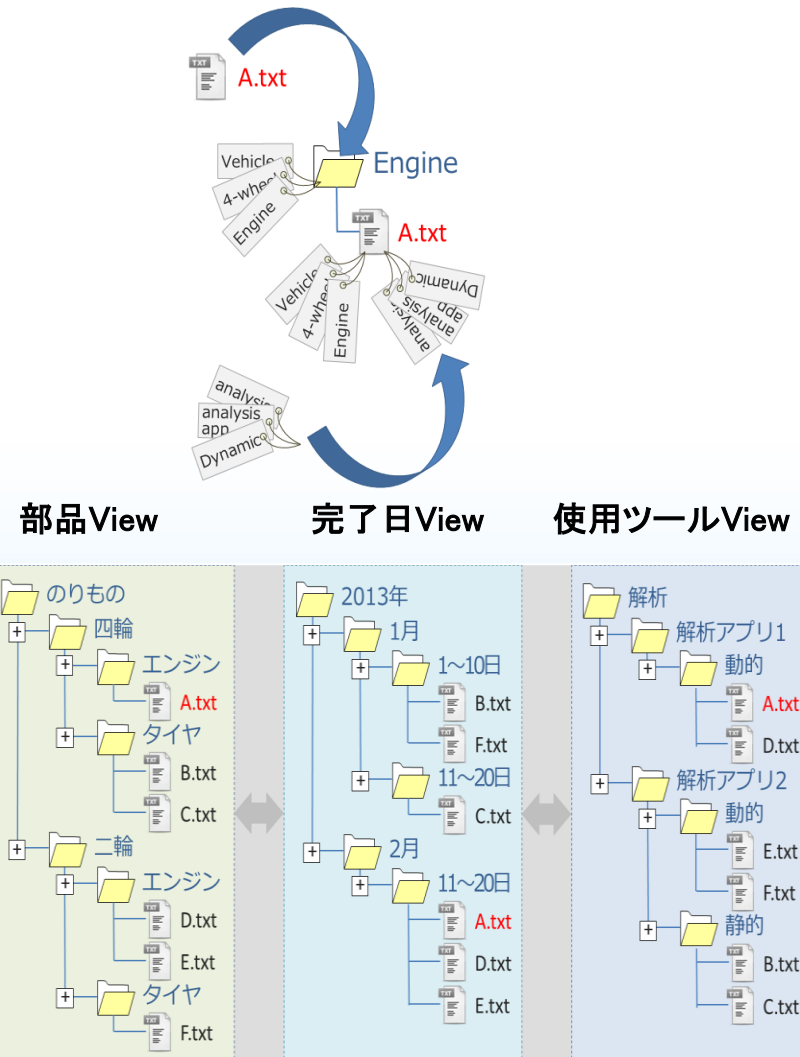
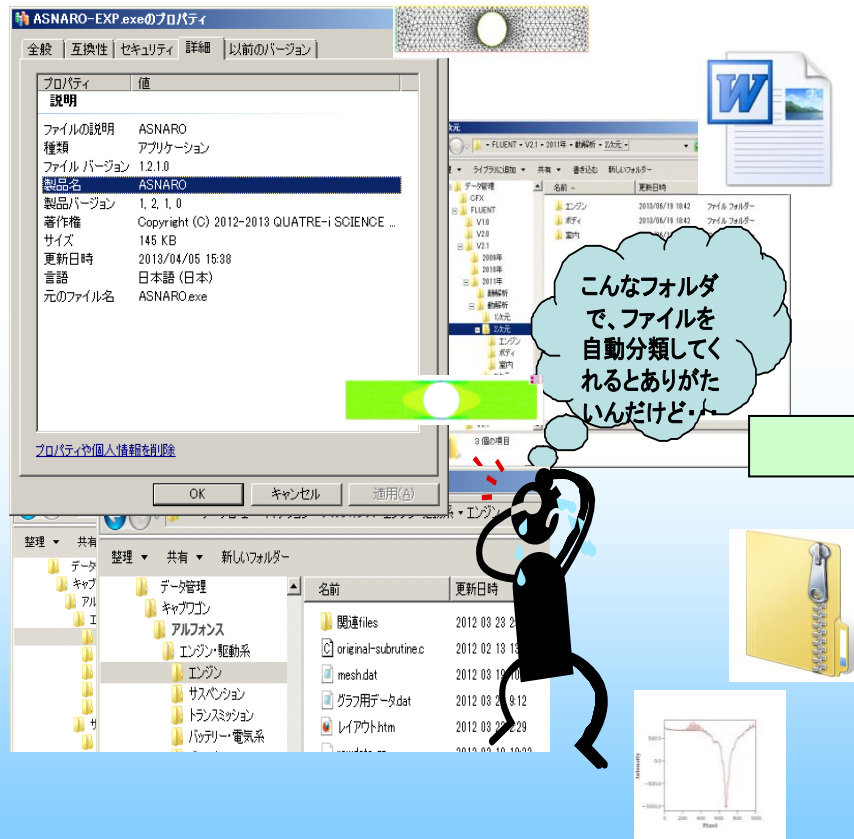


データへのメタ情報付与と 仮想フォルダツリーを使った情報分類の勧め



目次

- 1) 会社紹介
- 2) 階層的カテゴリ分けによる分類、整理
- 3) 「超整理法」、「超超整理法」を実践してみました！
- 4) 新しい分類、整理方法の提案：ASNAROを使ったデータ管理
- 5) ASNAROを使ったデータ管理の事例
- 6) まとめ

研究所で培ったR&D業務のシステム化ノウハウを結集し、 R&D統合プラットフォームを製品化しました。

社名の由来

感性と知性を

情報技術で

革新します



・会社名	株式会社 キャトルアイ・サイエンス
・代表取締役	上島 豊
・本社	京都府京田辺市山手南2丁目1-3 ハチセンビル 3号館 3階
・設立	2006年1月
・資本金	22,500,000円

会社設立経緯

上島が研究員だった当時の課題(日本原子力研究開発機構でHPCによるシミュレーションを行う)

- 1) HPC性能向上によるデータ数・容量増大とそれらの処理・管理作業の増大
- 2) 属人的作業のミス発生や効率化の限界
- 3) 刻々と進展する結果及びコード、ツール共有の煩雑さ

これの問題を解決する仕組みを研究員の時代に考案・開発

2000年 大規模データマネジメント研究会を立ち上げ

2003年 製薬会社や各種研究機関よりシステム相談を受け始める

2006年 キャトルアイ・サイエンスを設立

『研究開発を革新する』を原動力とし、R&D統合プラットフォームを製品化



販売開始から9年を経ており、R&Dプラットフォームとして、
お客様から厚い信頼を得ております。

主要取引先(順不同、敬称略)

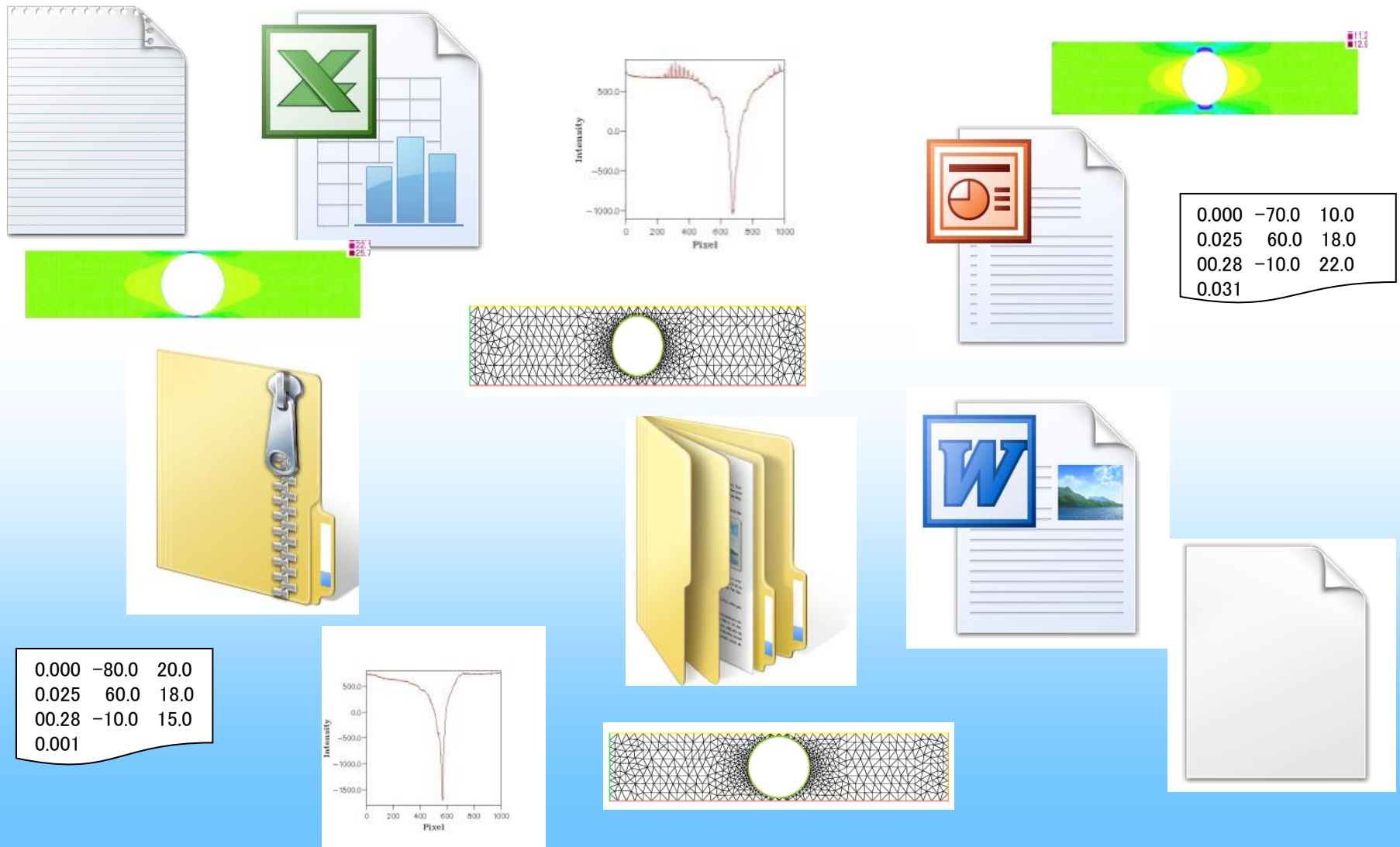
東京大学
大阪大学
北海道大学
東京工業大学
宇宙航空研究開発機構
日本原子力研究開発機構
防災科学技術研究所
高エネルギー加速器研究機構
情報・システム研究機構
情報通信研究機構
国立天文台
理化学研究所

計算科学振興財団
鉄道総合技術研究所
パナソニック株式会社
シャープ株式会社
株式会社神戸製鋼所
川崎重工業株式会社
新日鐵住金株式会社
中部地区自動車関連会社 等



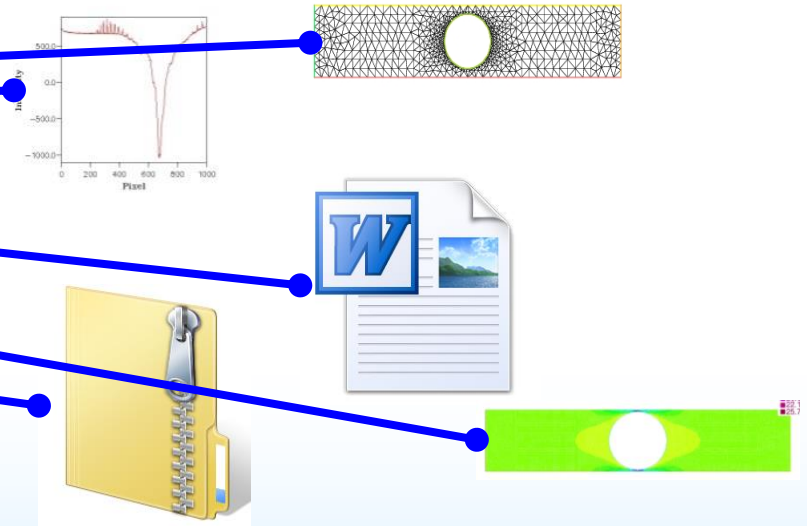
2) 階層的カテゴリ分けによる分類、整理

ファイル整理、ファイル探査は、どうしていますか？



[illegible]

- コード名
- - コードVersion
- - - 条件設定値
- - - - 観測量種別
- - - - - 可視化手法
- - - - - 可視化パラメータ



2001年 上島の研究
 固体への高強度レーザー照射
 P-cube
 20011102
 $D=1.0E+24$, $I=1.0E+23$, $N=100\cdots$
 電場、磁場、プラズマ密度、速度分布
 断面カンタ、等値面、ベクトルアロー
 $Z=0$ 面, $D=1.0E+22$, 密度勾配 $\cdots$

階層的分類カテゴリ分けは、図書館、生物分類、商品カタログなど様々な分野で使われています。

図書分類法

流体動力学の分類

key	value
類	4:自然科学
- 綱	2:物理学
- - 目	3:力学
- - - 分目	8:流体力学
- - - - 厘目	3:流体動力学
- - - - - 毛目	

生物分類

現代人の分類

key	value
ドメイン	真核生物
- 界	動物
- - 門	脊索動物
- - - 綱	哺乳
- - - - 目	サル
- - - - - 科	ヒト
- - - - - 属	ヒト
- - - - - 種	ホモサピエンス

- 各階層自体が内容的に包含関係にあり、末端が決まれば上位階層は自動的に決まる形になっており、分類法としては、強力だが、柔軟性に乏しい。
- valueが異なるkeyのものを含めても、ユニークになっていることが多い。

階層的分類カテゴリ分けには、 以下のような問題があります。

○こうもり問題

「複数の分類項目のどちらに分類すべきか？」を明確に決められないものがあること

○その他問題

どの分類項目にも分類されないものがあること

○誤分類問題

人が分類を間違ってしまうこと

○分類追加時の再分類問題

分類項目を細分化したときに細分化前のものを再分類する必要があること

○分類項目名の忘却問題

どんな分類項目名があり、それはどんな分類意図であったかを忘却してしまうこと



階層的分類カテゴリ分けは、研究のデータ管理においてさらに困った事態を引き起こします。

各階層自体が内容的に包含関係がなく、厳密な意味では分類分けではない。

研究におけるデータ分類

key

分析対象

- コード名

-- コードVersion

--- 条件設定値

---- 観測量種別

----- 可視化手法

----- 可視化パラメータ

2001年 上島の研究

value

固体への高強度レーザ照射

P-cube

20011102

D=1.0E+24, I=1.0E+23, N=100...

電場、磁場、プラズマ密度、速度分布

断面カンタ、等値面、ベクトルアロー

Z=0面, D=1.0E+22, 密度勾配...

全ての条件、パラメータをフォルダ名にできないし、階層を分けるのも非現実的

- ・組織内/共同
- ・内部資金/外部資金
- ・実験/シミュレーション/理論
- ・既発表/未発表
- ・未追試/追試済
- ・既論文化/未論文化

他観点の分類情報も入りたいが、フォルダだけでは、対応できない。

3)「超整理法」、「超超整理法」を実践してみました！



「超整理法」って知っていますか？

「超整理法」は、一橋大学名誉教授の野口悠紀雄先生が1993年に執筆した本で紹介している書類の整理方法です。

以下で、簡単にその方法を書いておきます。

- 1)資料を封筒に入れ、本棚に日付ごとに並べる。
- 2)使った資料は必ず一番左に戻す。
- 3)そうすると、一番左によく使う資料がくる。
- 4)逆に使わない資料は一番右に集まる。



実際に、「超整理法」を実践してみました。

実践してみたのは、今から15年前で、私がまだ日本原子力研究開発機構（当時は、まだ、日本原子力研究所）に在籍していた時です。

当時、自身の研究の他、数十件の共同研究、スーパーコンピュータの調達、運用、研究所の広報、渉外担当の業務を行っており、様々な資料が溜り、困っていました。

全ての紙資料を「超整理法」で整理し直すことを実践しました。

a) 資料量は、段ボール10数箱分

⇒ 背丈のキャビネット1個分をほぼ使い切りました。

b) 本家、「超整理法」では、「廃棄」と「神様ファイル化」という特別処理
がありましたが、手間と不安があり、行いませんでした。

c) 結果、このレベルでも1週間程度で平均探索速度は向上し、
見つからない事態はなくなりました。

○ 分類されているものの分類を崩す不安感と、手間のかかる封筒入れのハードル
○ 見つかるまで、探す忍耐力（全資料確認も厭わない決意：とはいえ最大数十分程度）



「超超整理法」って知っていますか？

「超超整理法」は、野口悠紀雄先生が2008年に執筆した本で紹介している書類の整理方法ですが、1993年の「超整理法」にその片鱗が書かれていました。

以下で、1993年時点では、以下のように書かれています。

1)ディレクトリを月単位で作成し、そこにその月に作成したすべてのファイルを置く

実は、事務用、日誌用、DBの3つは、例外で分類ディレクトリを作ると書いています。

○探査方法は、時間軸(ディレクトリ)、拡張子、ファイル名、
全文検索(当時は特殊アプリで)絞り込んでいくと説明されている

全文検索において、「あいまい検索」の充実が大きなカギと指摘されており、1993年時点(Windows95がまだない)で、そこまで見通せている野口先生の卓越した先見性には、脱帽です。

「超整理法」から「超超整理法」に移行してみました。

「超整理法」を実践して、1年後、富士通(PFU)のScanSnapというスキャンデバイスに出会い、「超超整理法:上島勝手に想像版」への移行を実行に移しました。

注)当然、この時点で「超超整理法」は、出版されていませんし、
実は、いまだに「超超整理法」は読んだことがないので、本家の方法は全く違うかもしれません。

なぜ、このタイミングで「超超整理法(仮)」を実践しようかと思ったのかというと、

a)PC内のファイルデータと紙面データを統一的に扱いたかった。

b)安価で高速(15枚/分)、高機能(両面同時スキャン、OCR機能付き)であるScanSnapが発売された。

上記2つが最大理由です。

実際には、「超整理法」管理外にしていた本型の冊子も裁断し、スキャンした(今でいう自炊ですね)ので全データ移行までに約2ヶ月を要しました。

単純計算で2ヶ月*20日*8時間*60分*15枚= 417,600枚程度ということになります。

実際に、「超超整理法」を実践してみました。

「超超整理法(仮)」を実践して、最初に感じたのが、「本棚が空っぽ！」の不安でした。しっかりした書籍(自炊をためらった)以外は、本棚には何もない状態で、同室の他の人に場所を貸してしまいました。

ただ、本当にうまくいくかも不安だったので、スキャンした紙を半年間は、段ボールに詰めておいて置き、半年間、全く使わなければ捨てようと思えました。

「超超整理法(仮)」の効果は、どうだったか？=>効果は絶大でした。

a)私のアレンジでディレクトリは、月単位も例外も作らず、1ディレクトリとしました。

b)日時+全文検索で殆ど事が足り、ファイル名、ファイル種別をたまに使う程度でした。

c)スキャン資料はOCRで文字列化を失敗したものがあるのもわかり、
後日、スキャン資料だけ別ディレクトリ管理にし、2ディレクトリ運用としました。

- ScanSnap Organizerは、PDFのプレビューがサムネールレベルでページ送りも含めてでき、強力で便利あるが故に、スキャンしていないファイルの方にも使いたかった。
- 単純探査はいいが、関係ファイルを集めたり、比較の為に分類するのは難しい。

研究では、比較の為、分類したいことが多々あるが、 「超超整理法(仮)」も階層型分類フォルダも無力！

上記は、「超超整理法(仮)」だから生じた問題ではなく、「超整理法」でも起こっていた問題です。格段に便利になったので、クローズアップされてきたということです。

「比較の為の分類」は、「階層型の分類フォルダなら・・・」と思うかもしれませんが、研究者が行いたい「比較の為の分類」は、あらかじめ決まっているのではなく、研究過程で変わってくるのです。例えば、電場をコードVersion間で比較をしたければ、電場,コードVersion以外の階層名は同じものを集め、電場とコードVersionの比較表にまとめたいはずです。

＝＞解析ケース数～フォルダ量が増えると如何に大変かは、皆様はご存じのはず！

そして、それは研究の宿命で、本質的なものだと思います。

分析対象	固体への高強度レーザ照射
- コード名	P-cube
-- コードVersion	20011102
--- 条件設定値	D=1.0E+24, I=1.0E+23, N=100・・・
---- 観測量種別	電場、磁場、プラズマ密度、速度分布
----- 可視化手法	断面カンタ、等値面、ベクトルアロー
----- 可視化パラメータ	Z=0面, D=1.0E+22, 密度勾配・・・

Windowsのタグ追加、 検索条件フォルダ機能を使ってみました！

Windowsには、「超超整理法(仮)」をもう少し便利にする機能があります。

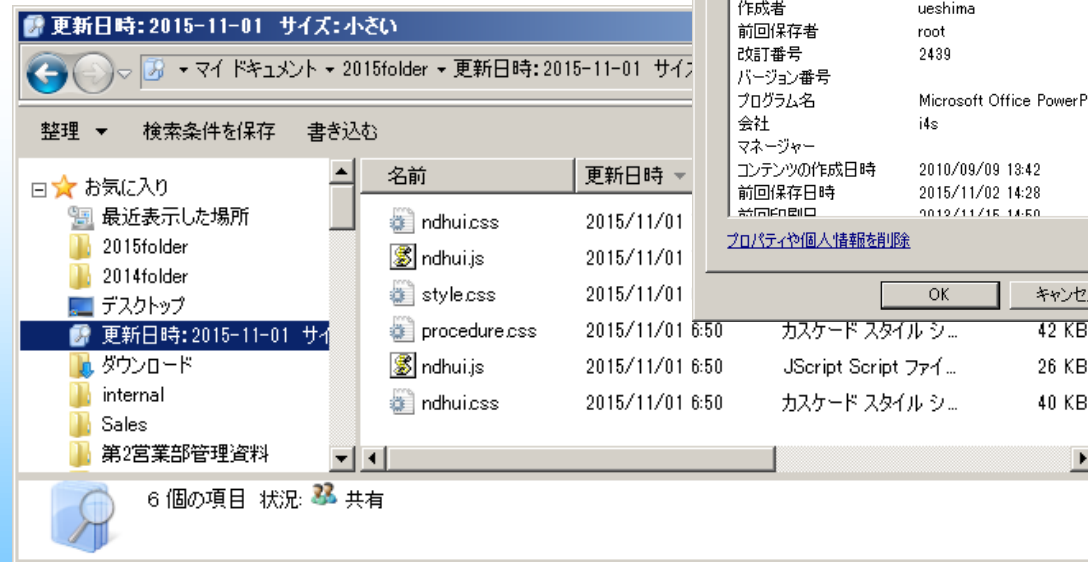
1) タグ追加機能

ファイルに任意の文字列(keyword)を追加できる機能で、そのkeywordでファイルを検索することができます。
「,」区切りで、複数keyword登録することもできます。



2) 検索条件フォルダ機能

複雑な検索条件を指定したものをフォルダとして保存でき、フォルダを開くだけで、リアルタイムに再検索されます。



- ファイルに多くのタグを追加しておけば、複雑な条件でも絞り込み、抽出できます。
- 検索条件フォルダを作って、どんどん蓄積すれば、検索自体も簡単です。

それでも、シミュレーションデータ、実験データに関しては、 全文検索、追加タグ、検索条件フォルダは、ほぼ無力！

1)全文検索機能の無力

シミュレーションデータ、実験データは、ほぼ同じフォーマットで、数値だけが変わっていることが多いです。つまり、全文検索では、同じ種のファイルを集めることはできても同種ファイルを絞り込むことはできません。

2)追加タグ機能の無力

膨大なファイルにどのようにタグを追加するのでしょうか？また、そもそもPC(Windows)に移動させられないサイズのものもあります。Windowsのタグは、key,valueのkey=タグということで固定で、value=keywordを複数登録できるといっても、それぞれが速度、圧力というように区別できません。区別できるよう(速度マッハ1.0等)にすると数値大小比較ができない。

例)速度ピークがマッハ0.9～1.2のものを抽出したい。

3)検索条件フォルダの無力

研究は、宿命的に定型的でなく、役に立ったと思われる検索も次の時に役に立つ確率が高いわけでもなく、また、その逆も真です。つまり、検索条件フォルダが非常に沢山になります。これ自体も検索はできますが、タグが追加できません。また、比較では、表とともに階層構造で理解を促進させるので単発フォルダでは不十分です。



問題点のまとめ

分析対象	固体への高強度レーザー照射
- コード名	P-cube
- - コードVersion	20011102
- - - 条件設定値	D=1.0E+24, I=1.0E+23, N=100...
- - - - 観測量種別	電場、磁場、プラズマ密度、速度分布
- - - - - 可視化手法	断面カンタ、等値面、ベクトルアロー
- - - - - 可視化パラメータ	Z=0面, D=1.0E+22, 密度勾配...

- 組織内/共同
- 内部資金/外部資金
- 実験/シミュレーション/理論
- 既発表/未発表
- 未追試/追試済
- 既論文化/未論文化

0)階層的分類カテゴリ分けの問題点

全ての条件、パラメータをフォルダ名にできないし、階層を分けるのも非現実的です。

1)全文検索機能の無力

実験、解析データなどでは、同種ファイルを絞り込めません。

2)追加タグ機能の無力

タグ追加が手間。数値大小比較ができません。

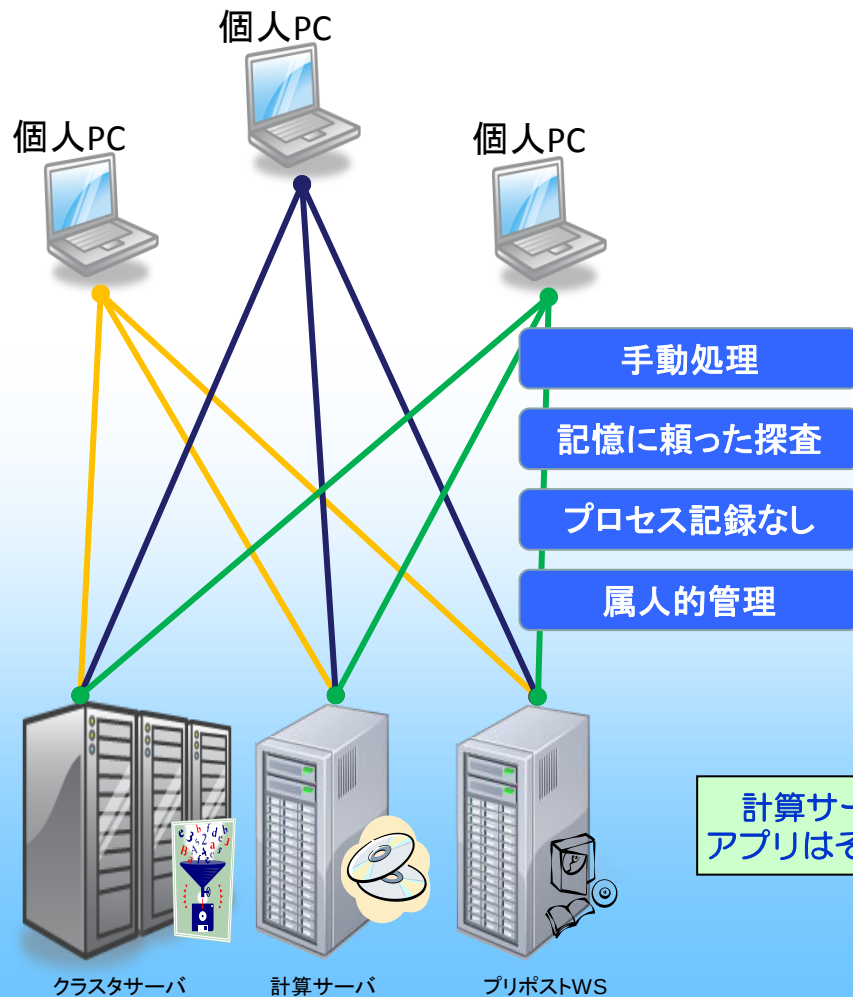
3)検索条件フォルダの無力

検索条件フォルダは、沢山になり煩雑。比較では、単発フォルダでは不十分です。

4) 新しい分類、整理方法の提案: ASNAROを使ったデータ管理

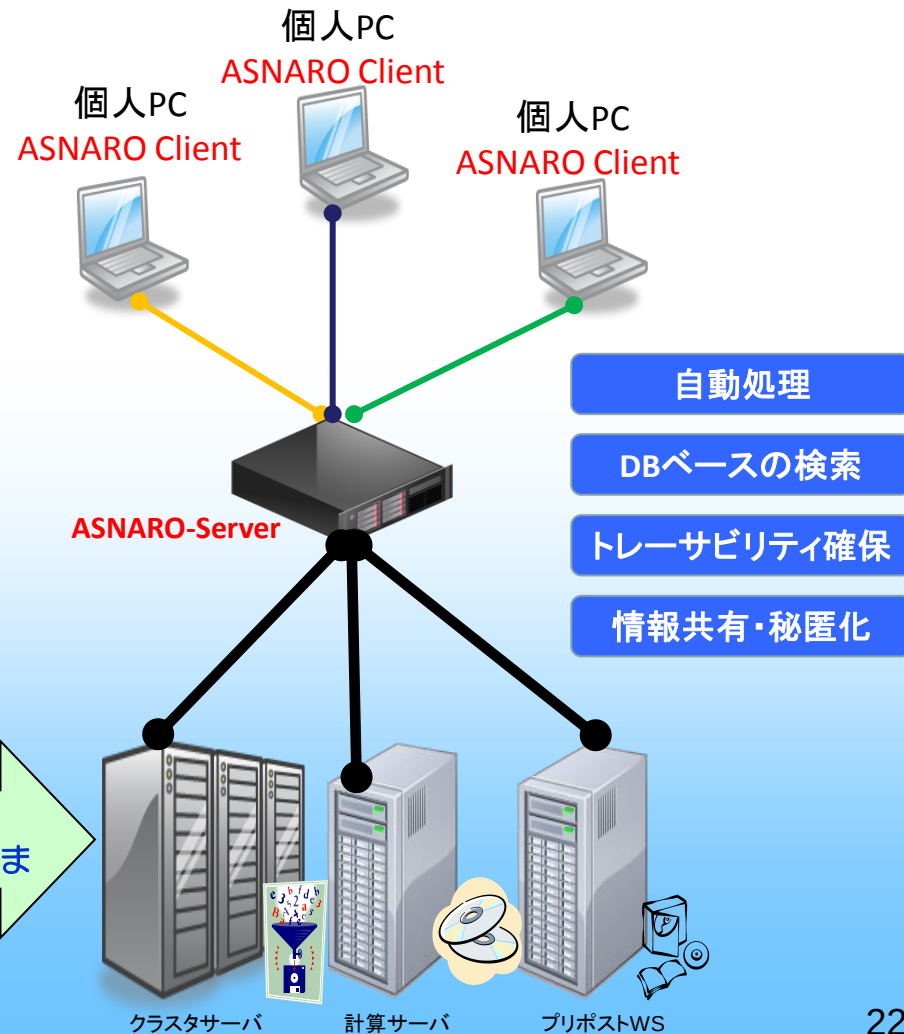
ASNAROは、自動化、プロセス/ 結果のDB化を支援し、 解析/データ分析業務を革新するソフトウェア！

従来の業務方式



計算サーバ、
アプリはそのまま

ASNAROによる業務改善



仮想フォルダにより開かれる新しいデータ管理、探査

ASNAROのフォルダは、Windowsの検索条件フォルダですが、階層構造にできる仮想的なフォルダになっており、比較しやすい構想構造、表構造を取りながら、ファイルを簡単に任意の分類(横串)で表示できる！

従来データ管理

ASNARO

フォルダを横串で、纏め直したいんだけど...

仮想フォルダで、横串探査が容易

あのVersionのアプリに不具合があったから、2次元動解析の結果に関して、商品化したもののみ遡って計算し直させて言われても、対象ファイルを探すなんて無理だよ！

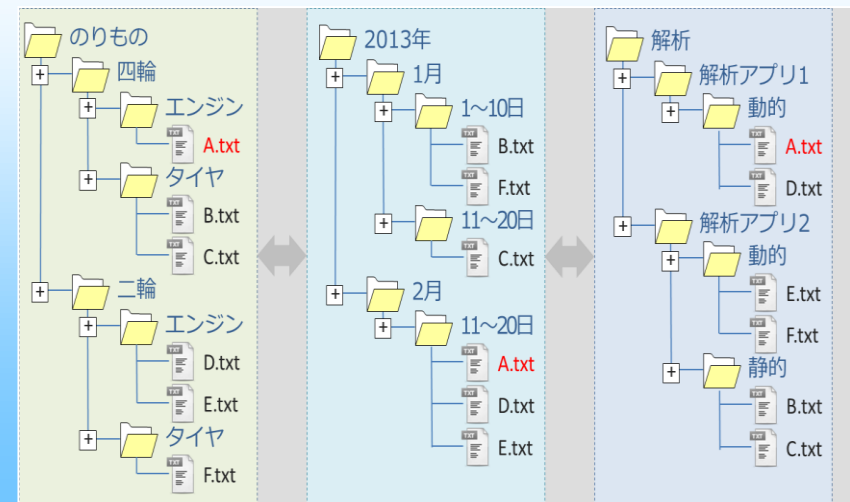
こんなフォルダで、ファイルを自動分類してくれるとありがたいんだけど...



部品View

完了日View

使用ツールView



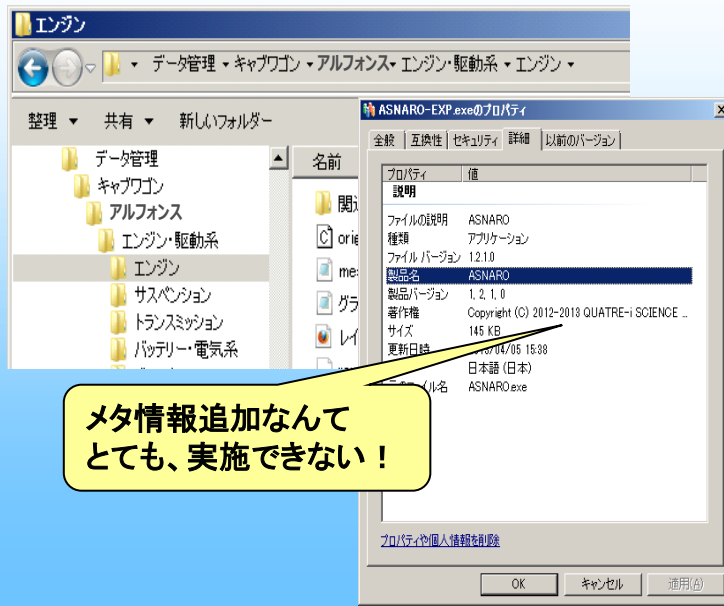
Key-value型のメタ情報を簡単に追加できる メタ情報Quick-ADD機能

ASNAROは、フォルダヘドラック&ドロップするだけで、ファイルにメタ情報を簡単追加！

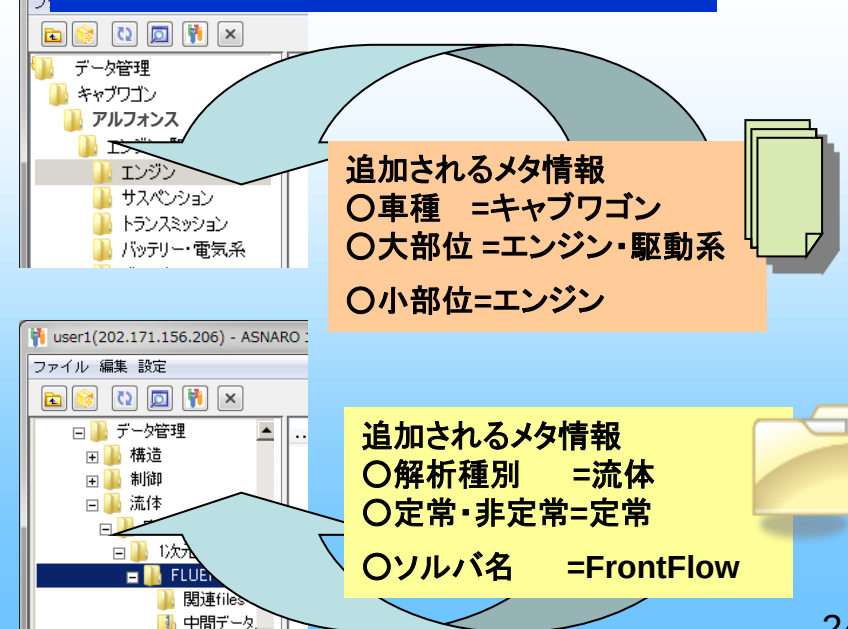
従来データ管理

ASNARO

1ファイル毎にプロパティを開いて・・・！



D&Dで複数ファイルに複数メタ情報を一括追加



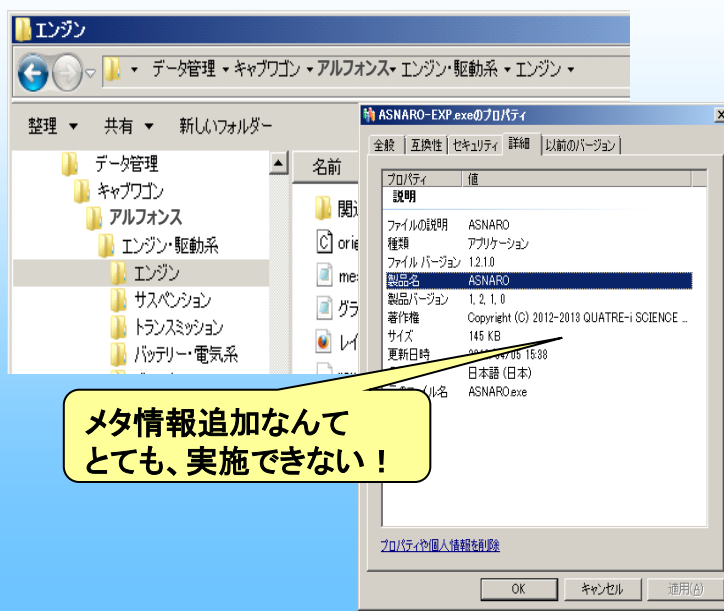


メタ情報を自動で追加できるメタ情報Auto-ADD機能

ASNAROで、データ処理実行すれば、I/Oファイルに自動的にメタ情報付し、I/Oの紐付も自動で行われる！

従来データ管理

1ファイル毎にプロパティを開いて・・・！

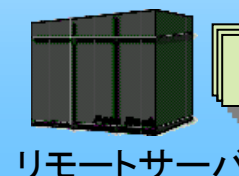


ASNARO



解析自動実行

メタ情報
自動登録



リモートサーバ



ファイル間の
紐付情報登録



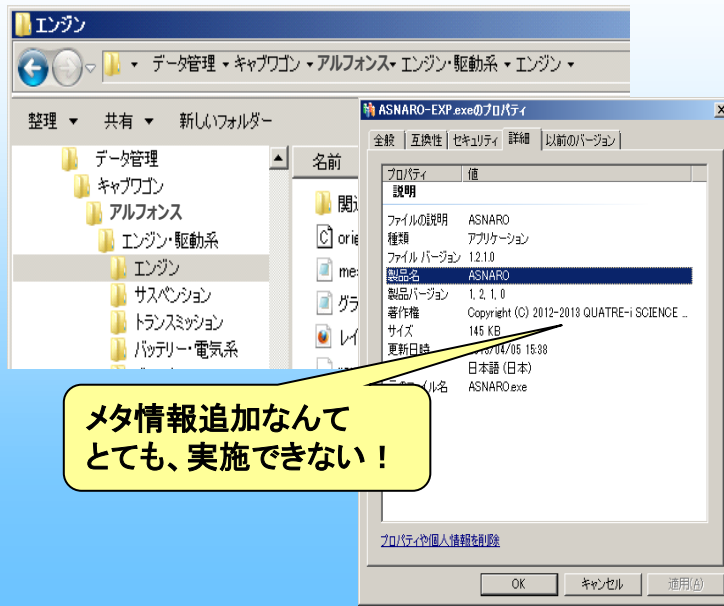
データベース

テキストファイルからKey-value型のメタ情報を自動でpickupし、自動DB登録するメタ情報pick-set機能

ASNAROで、データ処理実行すれば、ファイル内のデータパターンも分析し、パターンマッチ値をメタ情報としてDB登録！

従来データ管理

1ファイル毎にプロパティを開いて・・・！



ASNARO

入出力ファイル



パターン
マッチング

```
1 // input data
2 title = "Test data"
3 pressure = 1.8
4 temperature = 96
5 moisture = 25
6 angles = 10, 20
7
```

メタ情報として
自動登録

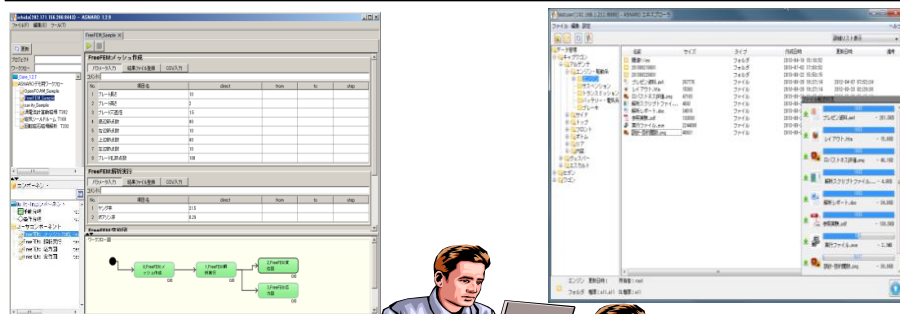


データベース

5) ASNAROを使ったデータ管理の事例



民間研究所で実験データの手動で属人的なデータ管理を 結果ファイルから特性情報を抽出し、グラフ化とともに自動化しました！



効果

•実験データの自動分割

⇒ 手動処理のヒューマンエラー多かった。

•特性情報の自動pickup/メタ情報付与

⇒ 手間がかかりすぎて行えていなかった

•実験後、すぐに結果を共有/閲覧

⇒ 共有できるまで1週間以上かかっていた。

実験装置



閲覧

ASNAO

NQS投入

特性情報の
自動pickup
DB化

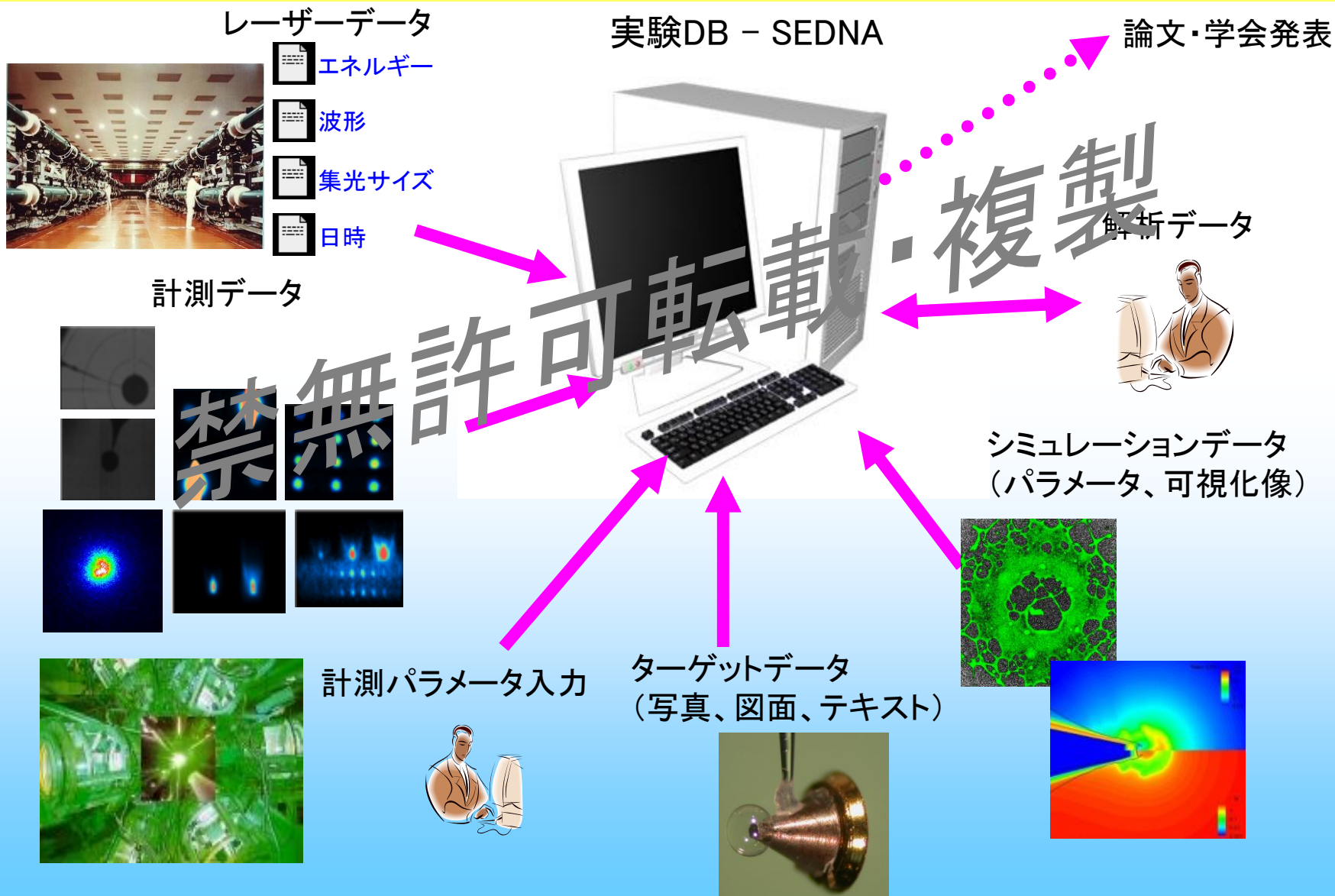
実験データの自動分割
グラフ化

ファイルサーバ

解析、グラフ化クラスターサーバ

SEDNAの目標

実験データベースSEDNAをハブとして
複数の研究者が参加したバーチャルな研究環境を提供



ASNAROで多様なレベルでのデータを共有、探査を実現！

Experimentsフォルダ

- 実験データを格納するためのフォルダ
- Experiment/#SERIES_ID/#SHOT_ID/#DIAGNOSTICSの構造
- #DIAGNOSTICSまでは管理者が自動作成する。

Diagnosticsフォルダ

- 計測器単位でデータを一括検索
- Diagnostics/#DIAGNOSTICS/#SERIES_ID/#SHOT_ID

Documentsフォルダ

- 研究関係のドキュメントを共有するためのフォルダ
- 解析データ, 学会発表資料, 論文等を入れて, グループ内で共有
- Document/#YEAR/#GROUP_ID
- #GROUP_IDまでは管理者が作成する

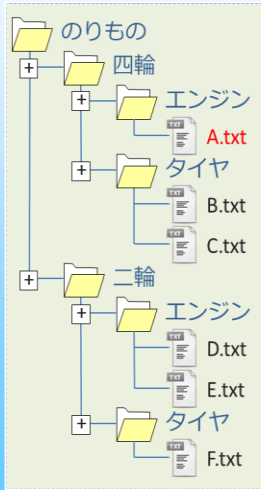
Informationフォルダ

- ユーザーへの連絡用のフォルダ
- アクセスは全員可能だが, ファイル登録は特定の人
- 各種マニュアル, 共同研究関係の書類を登録

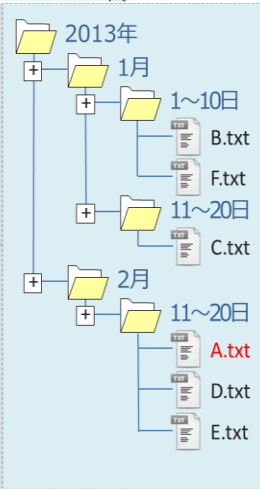
利用する人の立場、シーンにより、その時々適切な仮想フォルダからファイルを簡単に探査できます。

禁無許可転載・複製

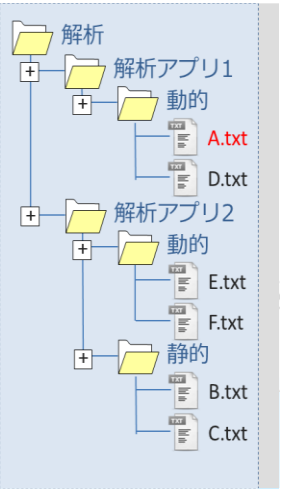
Experiments



Documents



Diagnostics



000567(sedna.ile.osaka-u.ac.jp) - ASNARO エクスプローラ

検索フォルダ
未分類フォルダ
Experiments
FK-02
FK-01
Diagnostics
Documents
2015
I4Sテストデータ
2015FIREX
150324JPS...at meet ng
150407FIREX_meeting
Information

名前	サイズ	状況	タイプ
150324物理学会春_Fuji...	949452	OK	ファイル
150322物理学会_Zhang2...	998332	OK	ファイル
物理学会2011春_F小...	725000	OK	ファイル
20150324物理学会_早...	21659462	OK	ファイル
150324JPS...pdf	32366364	OK	ファイル
150324JPS...docx	1703391	OK	ファイル

仮想ツリー機能による任意視点による分類

共通A1議事.txtのプロパティ

プロパティ

文書種類:議事録

登録日時:2014-02-10 18:04:58

更新日時:2014-02-10 18:04:58

bytes:11

relativePath:true

latest:false

delete:true

メタ情報:議事録

プロジェクトカテゴリ:実験プロジェクト

名前:実験A

解析番号:A001

管理番号:A001

文書種類:議事録

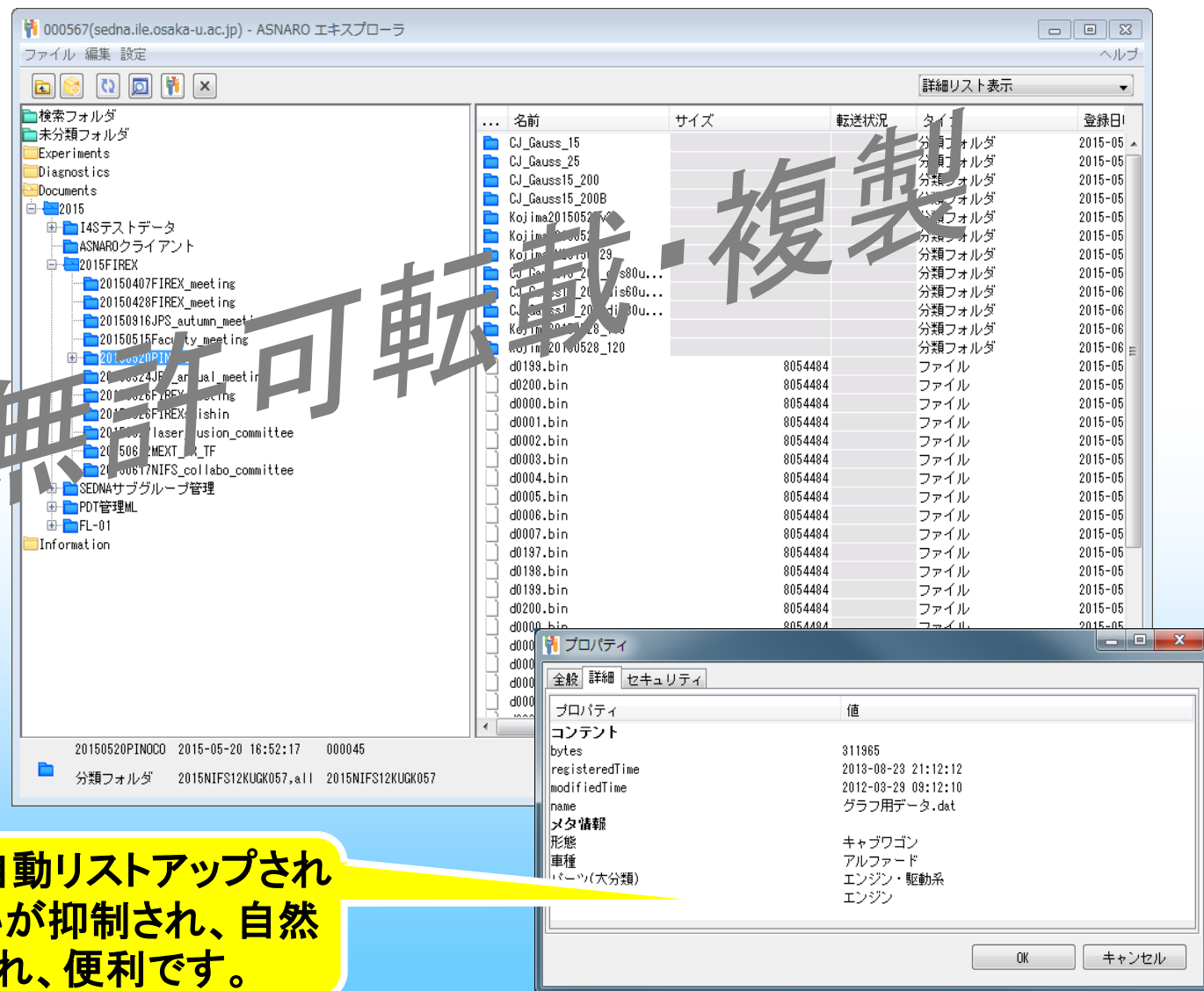
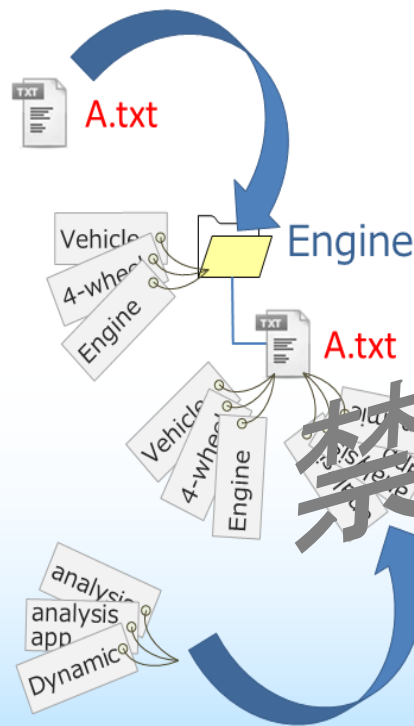
登録日時

更新日時

更新日時と議事録

共通プロジェクトの実験AのA001の

ILE, Osaka



既登録メタ情報が自動リストアップされるので、付け間違いが抑制され、自然とメタ情報が統一され、便利です。

6) まとめ



データへのメタ情報付与と仮想フォルダツリーを使った情報分類の勧め

研究におけるデータ管理は、階層的カテゴリ分類も超整理法も決め手を欠く。

○階層的カテゴリ分類は、

予め何が重要か不明な研究領域では使えない。

○「超整理法、超超整理法」にも

メタ情報を付与する手間の壁が立ちはだかる。

○タグ追加/検索条件フォルダ機能は、

ファイルに含まれるメタ情報の多種、数量性の為、有効でない。



ASNAROを使ったデータ管理で、これらの問題を一気に解決！

○検索と階層的カテゴリ分類を融合した**仮想フォルダ機能**

○メタ情報の付与手間の壁を消し去る**メタ情報Quick,Auto-Add機能**

○ファイル内のメタ情報の多種、数量性問題を解決する**メタ情報pick-set機能**



「ASNARO」は、R&Dプロセスを革新し、 技術者の創造性を支援します！

キャトルアイ・サイエンスは、
感性と知性を
情報技術で
革新します！

ASNARO

