Proposal for a new basic information carrier on the Internet: URL plus number sequence

Wolfgang Orthuber
UKSH, Kiel University
Arnold-Heller-Straße 3, D-24105 Kiel, Germany

Wilhelm Hasselbring
Software Engineering Group, Kiel University
Christian-Albrechts-Platz 4, D-24118 Kiel, Germany

Foundation of information

What is digital information?

Foundation of information

What is digital information?

(sequence of bits which is a)

number sequence

Foundation of information

What is digital information?

(sequence of bits which is a)

number sequence

which describes a

selection within a set of possibilities or "**domain**"

→ a **common domain** of sender and receiver

is precondition for precise information transfer

Digital information

2016 the basic carrier of digital information is a bit sequence or

number sequence

which is defined by context and **often not well defined**
(e.g. free text) and **not comparable** (because not identified).

The here proposed combiation

URL (of definition) plus number sequence

is called "**Domain Vector**" (**DV**): In it the **domain** of selection is **identified and globally uniform,** therefore the numbers are **comparable** to others with identical URL
and (in case of standardized and complete online definition) **well defined**

Efficiency of (BIG) Data

There are a lot of proposals for improvement of the data structure on the internet.

If we want maximal efficiency,
there is not so much freedom

We have to optimize the original (**numeric**) data.
These should contain maximal **density** of "useful" information.

"Useful" must be defined by users.

→ **URL (of online definition) plus number sequence**

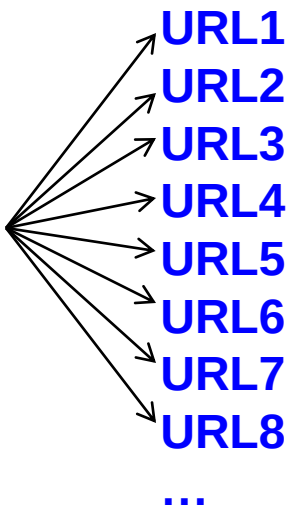
Abbreviation due to efficiency

In the DV

URL (of online definition) plus number sequence

the URL can be abbreviated to a number which is a pointer into a local table of external URLs:

URL of online definition as index **number**



Features of the new information carrier

The structure of the DV

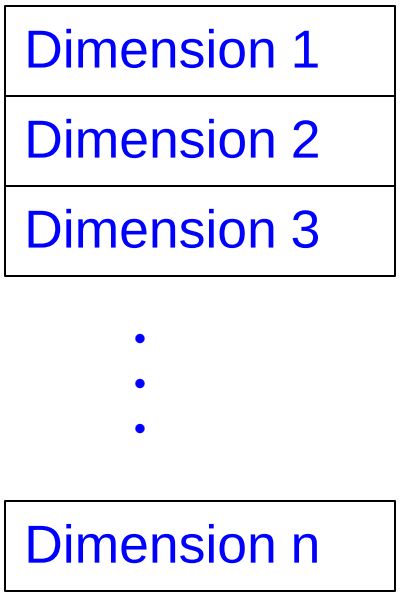
URL (of the online definition) plus number sequence

enables the combination of
maximal competence (definition by ALL internet users)
with
maximal efficiency (number sequence)

- The "online definition" defines in standardized (machine readable) way a "**Domain Space**" (**DS**). This is a metric space (which allows user defined similarity search and) whose elements are the DVs.
- A DV can precisely represent every definable information, from a simple word to complex multidimensional information e.g. in science, medicine, industry.

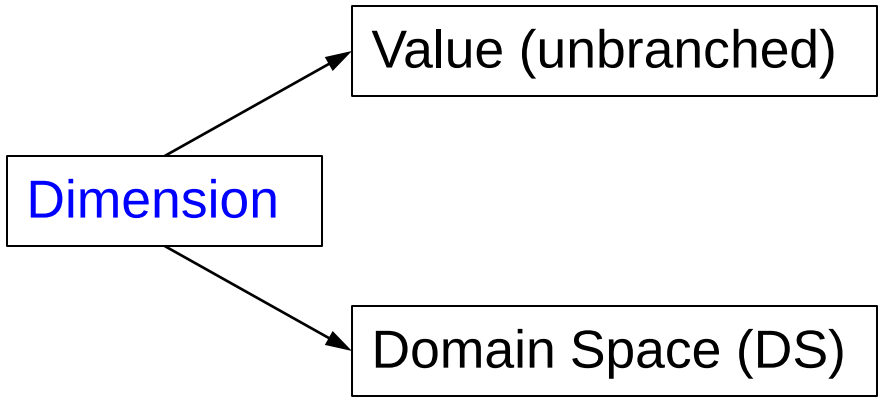
Domain Space Structure

Domain Space (DS)



The DS and every of its dimensions have a unique name (URL).

Dimension of a DS



Every dimension of a DS can represent an unbranched value (as number) or again a DS. So external DSs can be integrated and nested (like directories).

BW example, current approaches

RDF is a well known approach for machine readable data on the internet. We show an RDF example for representation of

"date and bodyweight"

```
<?xml version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns: cd="http://www.example.ns/cd#">

<rdf:Description
rdf:about="http://www.example.ns/cd/Weight measured">
   <cd:date>2014-01-30</cd:date>
   <cd:weight>83.914</cd:weight>
</rdf:Description>
</rdf:RDF>
```

Modification from http://www.w3schools.com/xml/xml_rdf.asp
to code the observation of body weight.

There is no Link (URL) to a complete standardized definition of the data (domain).

BW example, current approaches

FHIR is a next generation standards framework created by HL7 for medical data. At this data are coded in resources. The data **"date and bodyweight"** are e.g.:

```
<Observation xmlns="http://hl7.org/fhir">
  <text>
    <status value="generated"/>
    <div xmlns="http://www.w3.org/1999/xhtml">
     Jan 30 2014: Body Weight = 185 lbs</div>
  </text>
  <name>
    <coding>
      <system value="http://loinc.org"/>
      <code value="3141-9"/>
      <display value="Weight Measured"/>
    </coding>
  </name>
  <valueQuantity>
    <value value="185"/>
    <units value="lbs"/>
    <system value="http://unitsofmeasure.org"/>
    <code value="[lb_av]"/>
  </valueQuantity>
</Observation>
```

Excerpt of http://www.hl7.org/fhir/observation-examples.html which codes the observation of body weight.
There is no Link (URL) to a complete standardized definition of the data (domain).

BW example, the proposed approach

Example of the **DS Definition** for **"date and bodyweight"**

```
<DS>
  <kw>BodyWeight</kw>
  <dim>
    <kw>Date</kw>
    <unit>yyyy-mm-dd</unit>
    <format>yyyy-mm-dd</format>
  </dim>
  <dim>
    <kw>Weight-Morning</kw>
    <co>Weight at morning directly after stand up</co>
    <unit>kg</unit>
    <format>float</format>
  </dim>
</DS>
```

Exemplary (long) text form of a DV

```
<v http://numericsearch.com/bw.xml; 2014-01-30; 83.914>bodyweight</v>
```

After click its content can be displayed e.g.:

BodyWeight	
2014-01-30	Date | yyyy-mm-dd
83.914	Weight-Morning | kg

Comparison of size

RDF:

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns: cd="http://www.example.ns/cd#">

<rdf:Description
rdf:about="http://www.example.ns/cd/Weight measured">
   <cd:date>2014-01-30</cd:date>
   <cd:weight>83.914</cd:weight>
</rdf:Description>
</rdf:RDF>
```

DV:

```
<v http://numericsearch.com/bw.xml; 2014-01-30; 83.914>bodyweight</v>
```

The (by URL adressed) **online definition** of the DV
can be very complex and detailed and **can be reused worldwide**.

Implementation: http://numericsearch.com/

Dimension

keycomment of dimension owner

Keyword: Link:

Weight-Morning A

Unit: Link:

kg A

Comment:

Min: Max: Weight: 1 SetOneDivSd

representation: list tux

integer money floating point: medium length floating point: max. length

date in: yyyy-mm-dd hh:mm:ss yyyy-mm-dd hh:mm yyyy-mm-dd hh

yyyy-mm-dd yyyy-mm yyyy hh:mm:ss hh:mm

Implementation: http://numericsearch.com/

DS (Domain Space)

Definition of DS 1029 (BodyWeight) owner

|< << < > >> >| 0..0

0 BodyWeight
0 Date | yyyy-mm-dd
1 Weight-Morning | kg

Keyword: Link:

BodyWeight A

Comment:

This is: draft ok deprecated

Nested metric: Manhattan Euclidean Maximum

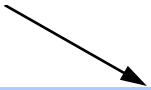
Implementation: http://numericsearch.com/

NumericSearch in DS 1029 (BodyWeight)

|< << < > >> >| 0..0 search spar

	sim	min	max	g
0				BodyWeight
0				Date | yyyy-mm-dd
1	80			Weight-Morning | kg

d	a		
0	17	o Bodyweight-1	2014-02-05, 80,
3.914	55	o Bodyweight	2014-01-30, 83.914,
5	26	o Bodyweight-2	2014-03-10, 75,
10	19	o Bodyweight-3	2014-04-20, 70,



BodyWeight	
2014-04-20	Date | yyyy-mm-dd
70	Weight-Morning | kg

Examples (from http://numericsearch.com)

Select i7 (index of Domain Space) kw0 add X copy nx

|< <<< << < > >> >>> >| 1000..1024..4017

i7	s	r	
1000	99	75	space-of-spaces || v627138
1001	30	9	o ride
1002	31	4	o my-location
1003	28	2	o Real-Estate
1004	7	0	o car
1005	530	10001	o test-space || try search 0..10: sub-DS1, sub-DS2 filled with pseudo random numbers 0..
1006	93	25	o Cupboard | Schrank
1007	109	11	o Diode || (for rectification)
1008	32	15001	o 260dim-demo || try search 0..10: sub-DS1, sub-DS2 filled with pseudo random numbers
1009	203	57	o text-as-dimension-example || dimensions (not used for similarity comparison) can also re
1011	11	3	o cardiovascular-disease
1012	24	2	o tamiflu-test || Is Tamiflu indicated? For answer of this question we could fill this space wi
1013	49	85	o NOx-Pollution-in-1000tons || exemplary data from Australia, Austria, Belgium, Germany
1014	14	8	o Screw | Schraube
1015	135	24	o datacube-example-as-TS || data like "The RDF Data Cube vocabulary" example chapter 5
1016	4	0	o opinion-about-xx
1017	1	0	o climate-fluctuations
1018	6	0	o Meeting | Treffen
1019	50	11	o ball-bearing | Kugellager-Edelstahl
1020	0	0	o Help || Search help (kind of help, time, location, duration etc.)
1021	0	0	o SleepDay || Documentation of one day sleep with result, result optionally after daycount
1022	5	0	o MRT-usage-year || yearly usage data about one magnetic resonance tomograph
1023	556	100001	o test-150dim || try search 0..10 in subv1(Euclidean metric) and subv2(Manhattan metric)
1024	10	1	o traffic-accident || DVs can become increasingly part of legislative vocabulary, existing juc

Example "Real-Estate"

NumericSearch in DS 1003 (Real-Estate) search-stat DS-stat

|< << < > >> >| 0..5 search spar

	sim		min	max	g	
0						gps-coordinates
0					☐	latitude | degree
1					☐	longitude | degree
1						financial
0					☐	price | euro || (if for sale)
1					☐	price-per-square-meter-living-area | euro / square-meter
2					☐	monthly-rent | euro || (if renting)
3					☐	monthly-rent-per-square-meter-living-area | euro / square-meter
4					☐	maintenance-costs-per-month-average | euro
5					☐	this-per-square-meter-living-area | euro / square-meter
2						energy-efficiency
0					☐	energy-costs-per-year | euro
1					☐	this-per-square-meter-living-area | euro / square-meter
3						age
0					☐	build | year
1					☐	last-renovation | year
4						size
0					☐	count-of-living-rooms
1					☐	living-area | square-meter
2					☐	area-of-corridors | percent-of-living-area
3					☐	area-of-windows | percent-of-living-area
4					☐	basement | square-meter
5					☐	lot-size | square-meter
5						equipment
0					☐	toilets
1					☐	showers
2					☐	baths
3					☐	garages

Every dimension (number) can be defined freely.

On the internet definitions can be reused and combined to build new definitions.

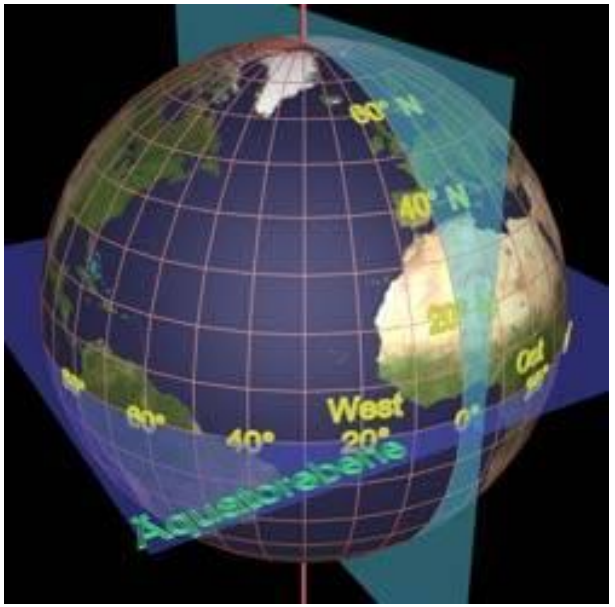
Examples

GPS-Coordinates

Feature Vector:

a_1 = Latitude

a_2 = Longitude



Examples

Industrial products, e.g. electric motors

Feature Vector:

a_1 = power (in Watt)

a_2 = rpm (revolutions per minute)

a_3 = energy efficiency (in percent)

a_4 = axial dimeter in mm

a_5 = length in mm

a_6 = height in mm

a_7 = weight in kg

Examples

Customized clothes

Feature Vector:

a_1 = collar size (in cm)

a_2 = abdominal girth (in cm)

a_3 = chest measurement (in cm)

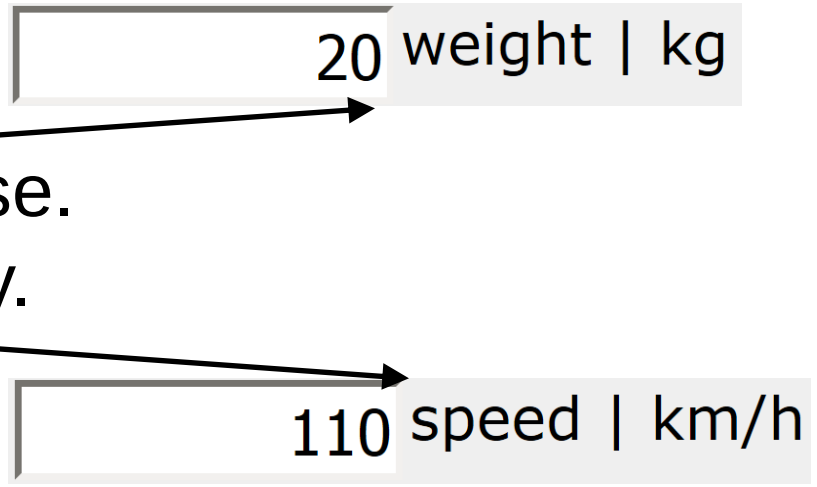
...

This DV be also used for ordering clothes.

Applications

Quantified text

Words of language can be made more precise by additional quantitative features shown after click, e.g.:



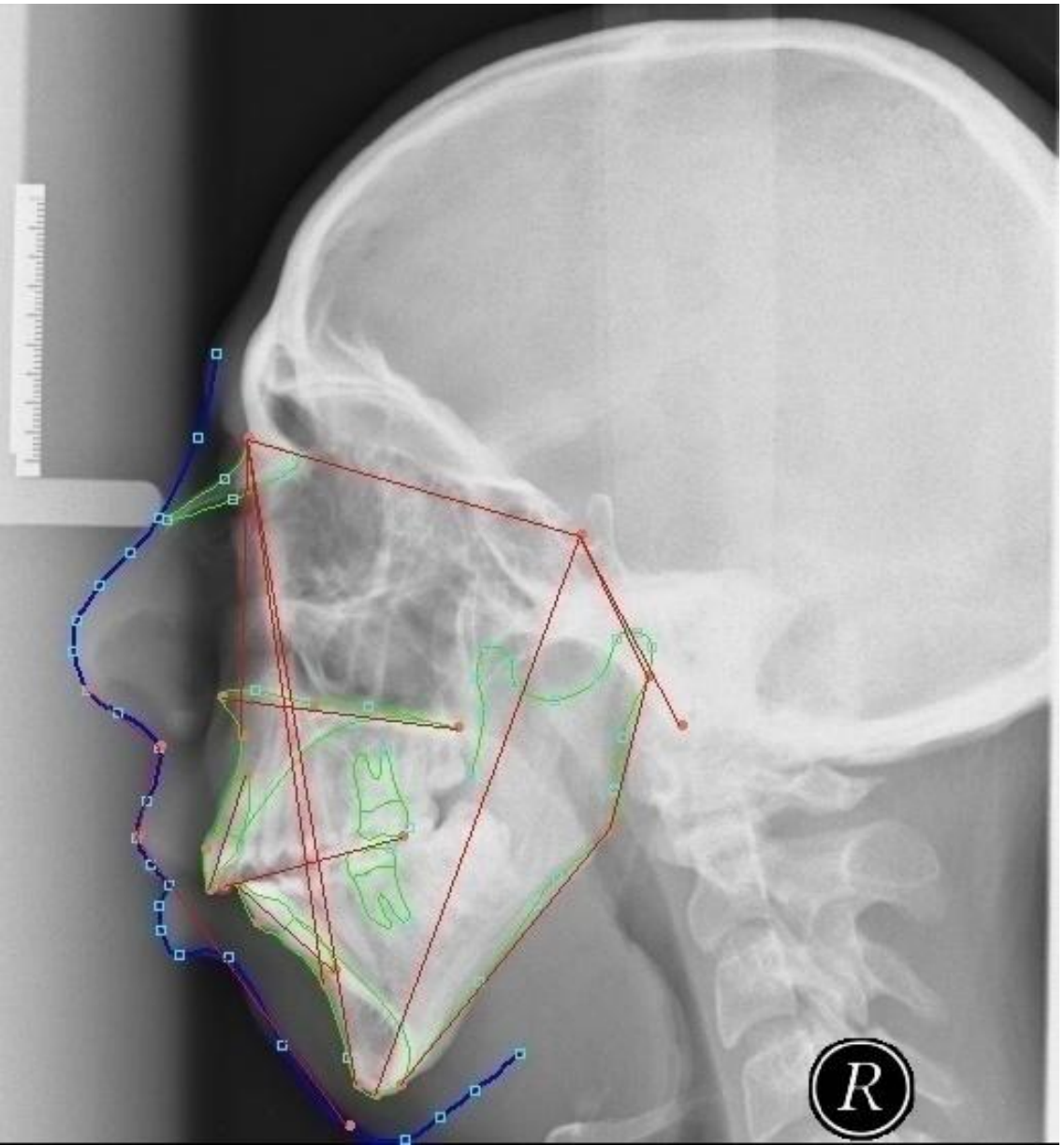
If the DS definition is done in multiple languages, automatically the definition in (by user selected) system language can be shown.

Applications

Quantified legislative text

DVs can make legislative text more precise. Similarly like for description of medical decisions DVs can be also used for description of judgments and (internationally) large searchable internet collections of judgments can be built. So it would be possible for judges to compare existing cases to past cases in the collections more precisely and to check past judgments. This could help jurisdiction towards better reproducibility and precision.

Examples (Medicine)



History of idea:
Medical applications

e.g. Cephalometry:

- A scientific study of the measurements of the head with relation to specific reference points
- utilizing a fixed, reproducible position for lateral radiographic exposure of skull
- used for orthodontic treatment planning, for evaluation of facial growth and development, including soft tissue profile.

SureSmile 5.6 -

File Edit Go Case-Mgmt Tools Help

TS

Upper Occlus

Reference Model: Diagnostic Model 1 (07/16/07 19:21:16)
Reference Arch: Occlusal Plane: Upper, Arch Form: Lower

Bolton Ratio (6..6): 86.3 (3..3): 65.3
Maxilla Sum (6..6): 92.1 mm (3..3): 44.7 mm
Mandible Sum (6..6): 79.5 mm (3..3): 29.2 mm
Surplus (6..6): Maxilla 5.1 mm
Surplus (3..3): Maxilla 5.9 mm
Measured Overjet: -0.5 mm
Arch Length Discrepancy:
UR: 0.0
UL: 0.0
LR: 0.1
LL: 0.0

8.3
8.1
7.0
6.9
7.5
6.9
7.7
6.8
6.4
6.8
10.4
10.2
10.2
9.6

Camera Navigation

[20.7.2007] Intra-oral Right Buccal

[20.7.2007] X-ray Lateral Ceph

Global Registration | References | Occlusal Plane | Arch Form | Upper Disp. | Lower Disp. | Fixed Teeth | Notes

U

Midline: r. 0.2 / r. 0.2

Arch Width

	(27.0)	(27.4)
Molar	27.0	27.4
R / L		
Canine	16.9	17.6
	(16.9)	(17.6)

AP Position: R none | L none

L

Midline: r. 2.2 / r. 2.2

Arch Width

	(11.5)	(12.2)
Canine	11.5	12.2
R / L		
Molar	24.3	25.0
	(24.3)	(25.0)

AP Position: Front none; R / L Molar none none

Extract / Space (+) / IPR (-)

Apply | Cancel

0.1

Molar Class: Right / Left — Maint., I, II, III

Overjet: cur. 3.3 — 0.0

Arch Selection: Currently: Natural — Natural; Symmetric / Asymmetric

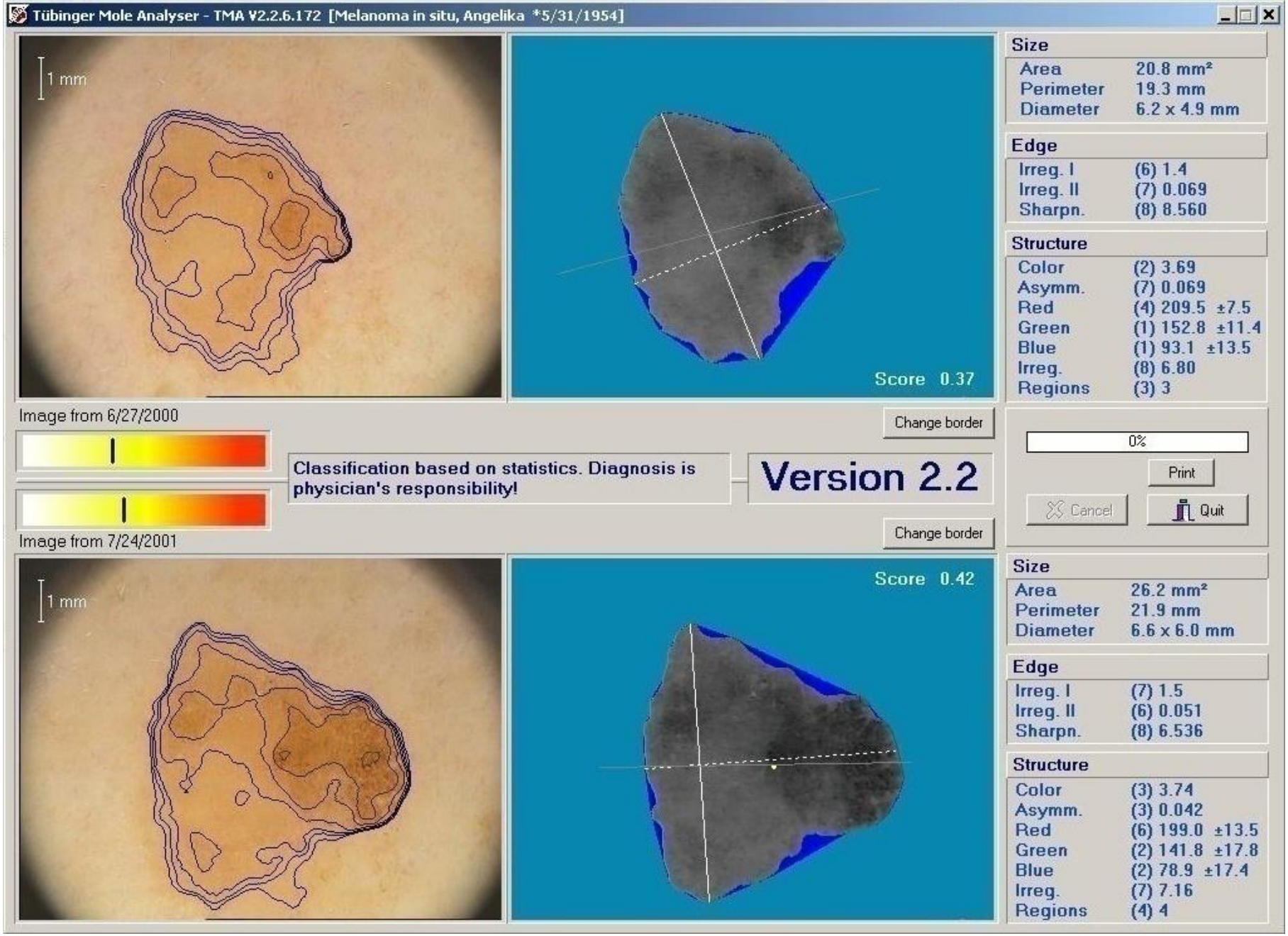
Setup: Arch: Upper, Lower; Vertical, Vert. Contact, Horizontal, Horiz. Contact; Go; Reset

Max. IPR [mm]: 0.0 — 3 - 3 / 4 - 4 — Apply

Align: Front — Align

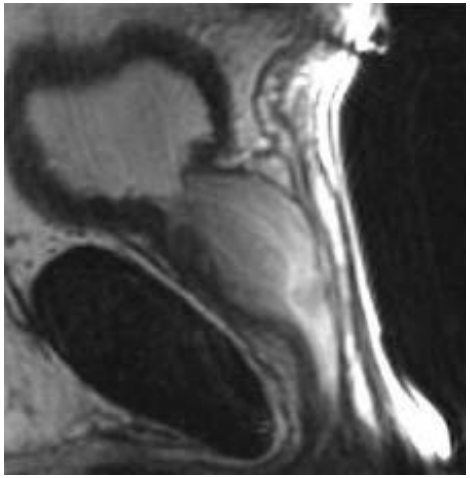
For Help, press F1

Elapsed Treatment Time: 7 Months, 3 Weeks | Admin | Active | Real



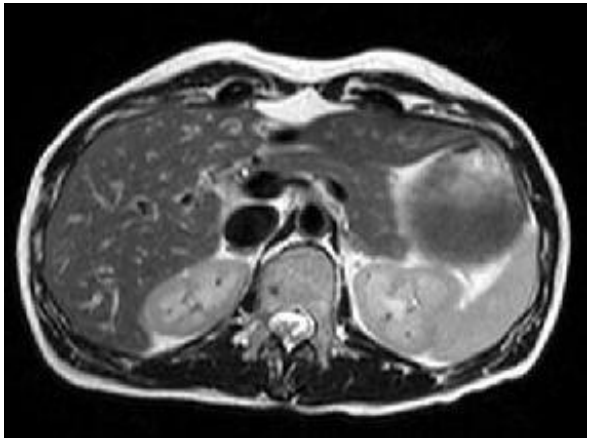
Source: Tübinger Moleanalyzer

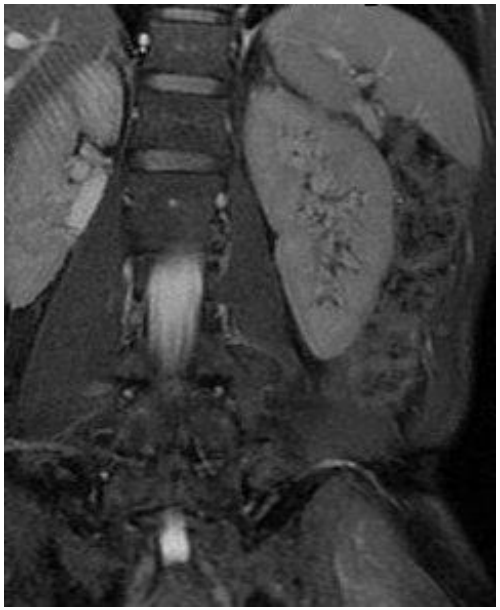
Examples (Medicine)

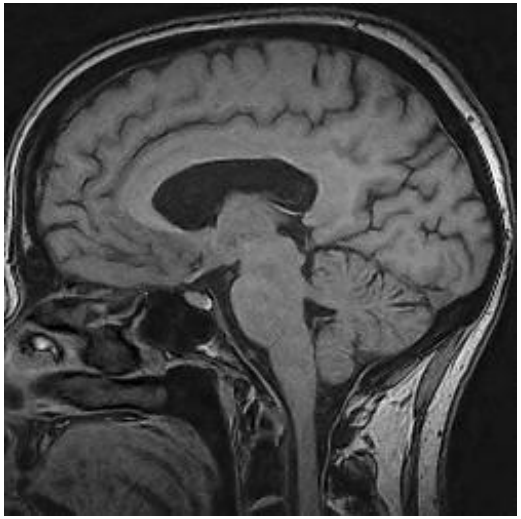


MRI prophylaxis

- Selection of frequent and serious diseases which are best detectable by MRI
- Description and quantification of decision relevant features (initially 2D, later 3D)
- Comparison with previous findings and cases







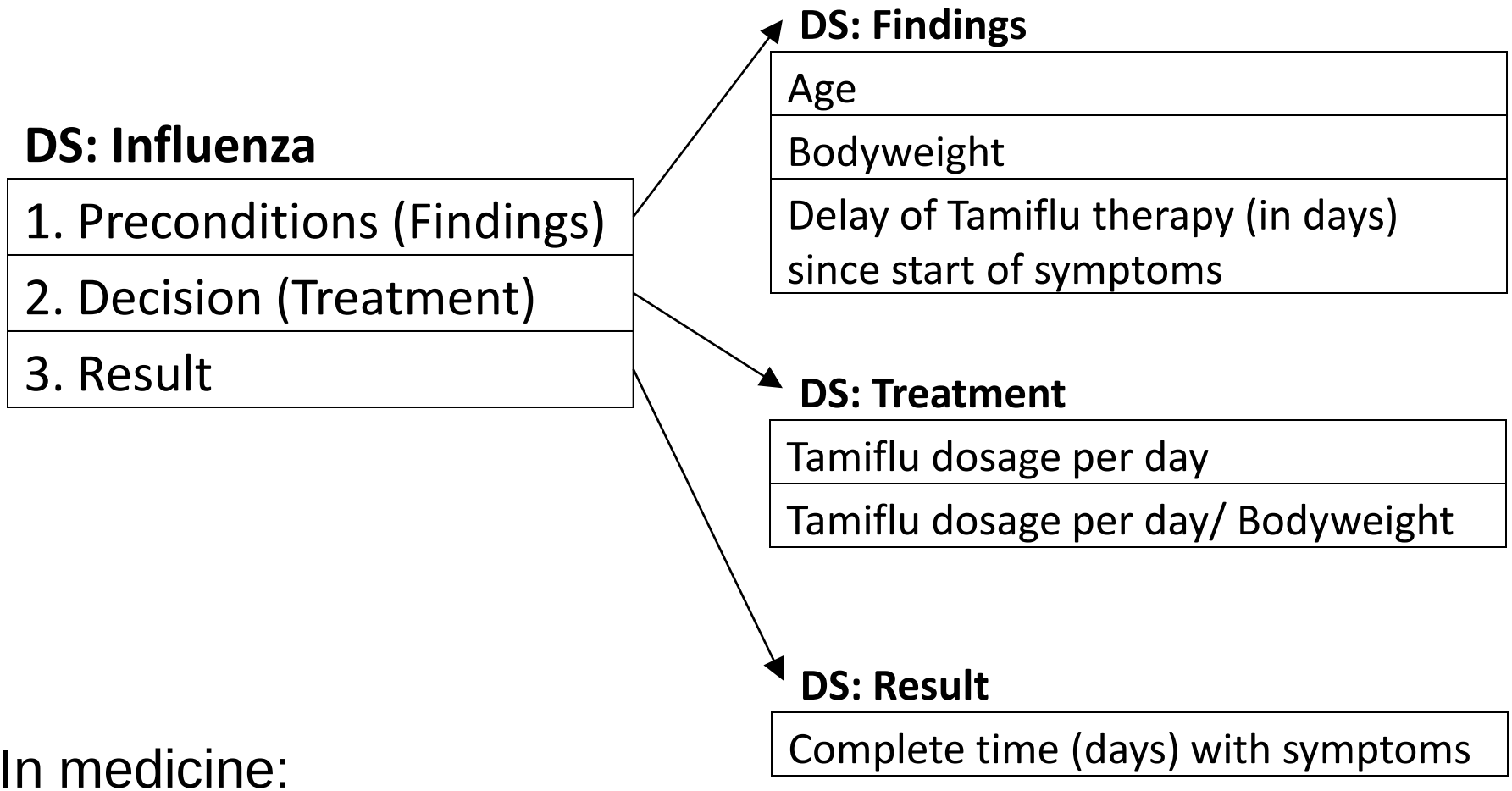
Searchable Feature Extraction

There are uncountable many possibilities for Feature Extraction. Representation of important features of a resource as dimensions of a DS is an important application. It could be used to make complex resources identifiable, comparable and searchable.

Application: Decision support

- A decision means a **new** selection within a domain (i.e. generation of new information).
- So precondition of well defined decision support is that all speak and think about the same domain.
- So a common standardized Domain Space definition (and with this the definition of the domain) on the internet is natural also for decision support.

Application: Decision support



In medicine:

- search patients with similar findings
- at this vary possible treatment decisions
- look for decision with best result

Applications

Searchable original scientific data

- Scientific original data are usually detailed quantitative data.

- As DVs on the internet these would be searchable and interoperable.
- Quantitative data could be defined that automatic combination is possible.

Synchronized Index for combined DVs

DVs can be grouped together, so that one group describes the same resource. The index can combine multiple DSs.

So data providers can select dimensions which they group together.

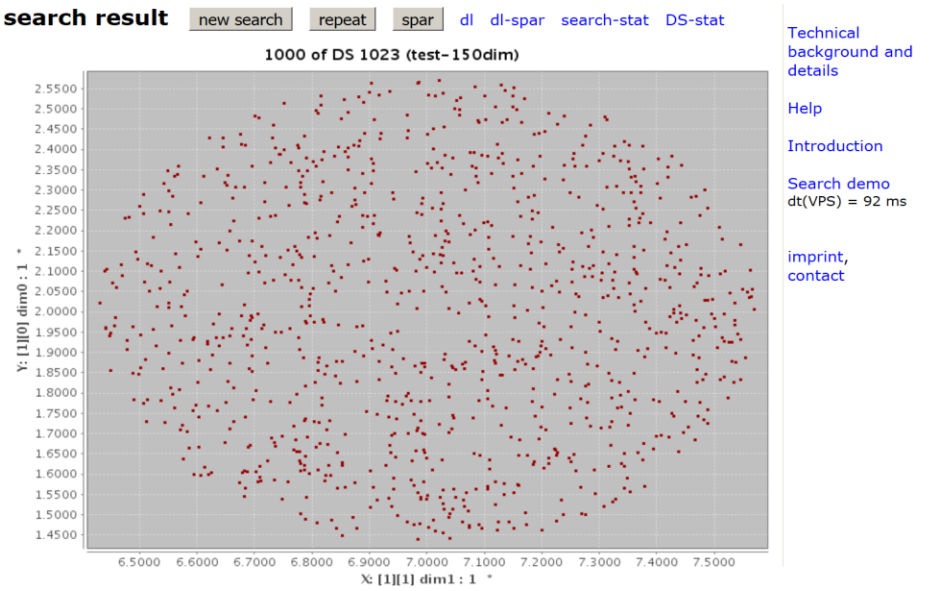
Later it is possible to combine dimensions of different DSs also for search.

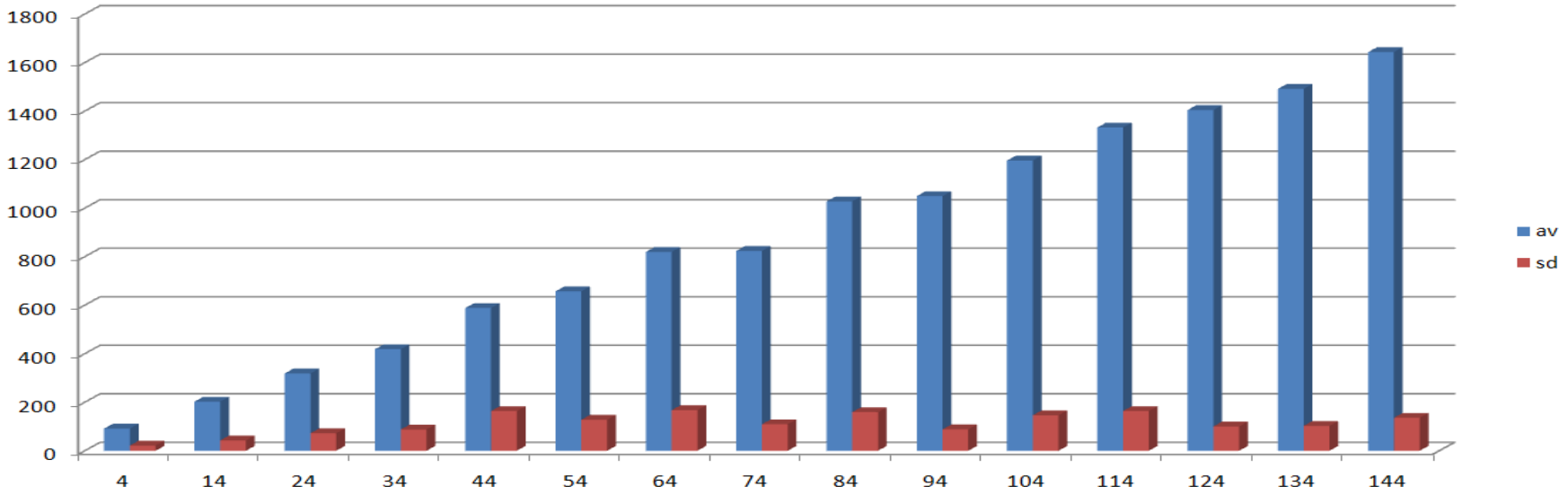
Synchronized index for combined DVs

NumericSearch in DS 1023 (test-150dim)

|< << < > >> >| 0..2..2 search spar

	sim |	min	max |	g
0				subv0
0				☐ dim0
1				☐ dim1
1				subv1
0	2			☑ dim0
1	7			☑ dim1





The search time within 100001 DVs in ms (vertical) in dependence of searched dimensionality (x 64 bit).

Discussion

Only after considering existing well known concepts (e.g. LOD, RDF, FHIR) for machine readable data on the internet we decided for this approach - to achieve the features a) to k) purposeful and efficiently. We list some possible counterarguments and answer to them:

- Without (online) DS definition the data of the DS are no more readable.

Answer: Therefore periodical backup of published DS definitions is included.

- Data are not self-contained.

Answer: To get self contained data, online definitions can be automatically downloaded together with DVs. (Remark: currently many data are not self contained due to not standardized or not locatable or missing exact definition. This approach provides a systematic solution.)

- Definitions by user may be not stable.

Answer: internet services can allow (identified) users to create definitions and expand these, but not to delete these.

Discussion

- There will be many DS definitions for the same thing.

Answer: Up to now digital information is defined by context which (e.g. free text) is even for the same thing VERY variable. By the combination "URL (of definition) plus numbers"
we get a systematic approach to solve resulting problems: Search engines can be used to search within existing DS definitions and provide in the results (besides the URL and the responsible creator) also the date of creation, the frequency of usage and the count of contained DVs. So users can prefer e.g. the most popular definition.

Searching within existing DSs definitions is obligatory before creation of a new DS definition. New definition can be connected with existing definitions e.g. by a "sameAs" directive. Search engines can provide the option to treat all by "sameAs" connected definitions as one definition.

Discussion: redundant definitions

So every user can generate DSs and searchable spaces with quantitative data. (Purposeful) Redundant definitions of dimensions are to be expected.

All redundant definitions can be connected. For this the (e.g. in http://www.w3.org/TR/owl-ref/#sameAs-def described) **sameAs** directive can be extended to the form:

this Dimension is sameAs (algebraic) expression of other DS Dimensions

Usually definers of DSs are interested to connect their definitions with other definitions, so that searches there can also include the own space.

Conclusion

- Digital data are represented as bits or (more precisely) as
  number sequences
  which are defined by context.

Much more is possible:

- All users can **define digital data globally** on the Internet. The definition can be optimized globally
  in dependence of the application.

- Using the DV data structure
  URL of the online definition plus number sequence
  data are **interoperable, comparable** and **searchable**,
  because every kind of information is identified uniformly
  (by the URL of definition).

Uniform definition of information by users

The DV data structure
URL of the online definition plus number sequence
enables the combination of
maximal competence (definition by all internet users) with
maximal efficiency (number sequence)

Creation of a common standard for online definitions and data is the next step.

Interested to contribute?

Contact: Wolfgang Orthuber, Kiel University, Germany
orthuber@kfo-zmk.uni-kiel.de

Further Information: **http://numericsearch.com**

Repetition

Well defined information means selection from a well defined set or domain.
So preconditions for precise transfer of information are:

(1) Well defined common domain (for all participants of conversation)
(2) Ordered domain (so that its elements are selectable by numbers)
(3) Transfer of the numbers which show the selection in the domain

In this approach the domain is defined online.
So it is defined uniformly in the internet - for uniform definition of information.
It is called "domain space" (DS) and it is ordered: a DS is a n-dimensional metric space. Its elements are called "Domain vectors" (DVs).

Every DV has the form

URL (of the online definition) plus number sequence

where the (online definition at the URL) defines the domain and the number sequence describes the selection. The URL can be abbreviated.

HTML like syntax example (later abbreviated syntax and binary representation is recommended):

```
<v http://numericsearch.com/bw.xml; 2014-01-30; 83.914>clickable</v>
```