ABCD comments commented – March 8, 2005

While working on the update of ABCD, I have collected my notes on the comments provided by other people and included them in this document. Please note that some of the comments refer to earlier versions of ABCD, so that they may have been implemented already e.g. in versions 1.10 or 1.20 of the schema. See the related document “Changes from Version 1.20” to see only the changes made since that version (which is the current reference installation providing data to BioCASE and GBIF under the BioCASe protocol).

Some issues still under discussion are highlighted in red.

Comments from: Adrian Rissone (ARI), Agnes Kirchhoff (AKI), Alexander Kroupa (AKR), Andrea Hahn (AHA), Anthony Isenor (AIS), Anton Güntsch (AGU), Armando Urquiola Cabrera (AUC), Ben Richardson (BRI), Bob Morris (BMO), Charles Copp (CCO), Dave Watts (DWA), Denis Lepage (DLE), Donald Hobern (DHO), Dorothea Gleim (DGL), Edward Vanden Berghe (EVB), Gregor Hagedorn (GHA), Hannu Saarenmaa (HSA), Javier de la Torre (JTO), Jessie Kennedy (JKE), Jörg Holetschek (JHO), John Wieczorek (JWI), Markus Döring (MDO), Mark Jackson (MJA), Michael Malicky (MMA), Michael Wiedemann (MWI), Neil Thomson (NTH), Patricia Mergen reporting for the Belgium Coordinated Collection of Microorganisms (PME), Peter Davyndt (PDA), Peter Strobl (PST), Sabine Roscher (SRO), Steve Ginzbarg (SGI), BioCASE Technical Committee (TC), and Yde de Jong (YDJ)

March 8, 2005. Walter Berendsohn (WGB)

Contents

1. Dataset – general and metadata issues

We intend to completely revise dataset-level metadata to provide a common structure with the other TDWG standards (SDD and Names/Concepts as well as the common protocol). However, our present priority is to produce a new version which incorporates all comments received over the past 2 years, but we do not propose to implement the UBIF structure suggested by SDD, which offers many exiting possibilities for future development, by which also needs further and wider discussion. Consequently, we are here using an evolutionary approach, taking in some of the datatypes suggested by other standards, but keeping the schema compatible with the existing BioCASe protocol software.

1.01. NTH: A number of the new elements that have been put in have taken the default of "mandatory". An example is <ContinentOrOcean> so a review of optionality would be good.
WGB: Done.

1.02. HSA: Shouldn't "OriginalSource" be under Datasets, not Dataset? One can derive many Datasets from one source. It is unlikely that one XML file would contain datasets from many sources.
WGB: I do not agree. For example, the SysTax database in Germany will soon be accessible and it will supply a single ABCD document as a response when asked, e.g., for Apis. The data, however, will come from several entomological datasets hosted by SysTax, and those are from several institutions, each of which is an OriginalSource.

1.03. HSA: GatheringCode is good. However, ObservationUnitIdentifier which serves similar purpose is named differently. We need to be consistent how we call them (ID, Code, GUID, ...) We also need similar IDs for other elements such as Site, Person, ... This is necessary so that we can go back to original data if needed, or locate it elsewhere.
WGB: I think this is an important point for further discussion and standard development. I agree that clearly defined key elements should be named accordingly. Currently, this refers to SourceInstitutionID (changed from SourceInstitutionCode), SourceID (changed from SourceName) and UnitID (unchanged) in ABCD, which are roughly equivalent to the three IDs in the Darwin Core. In renaming, I followed HSA's suggestion for these elements. However, it has to be pointed out that these have provider-determined contents - the only rules are that (i) the SourceInstitutionID should be unique globally (it may be a code or a name or a combination of code-source and code [e.g. IH-K]; (ii) the SourceID must be unique within the institution, and (iii) the UnitID must be unique within the source. The naming of other ID's, Codes and Identifiers used in the schema now refers to content rather than to their role in the processing of the information. I posit that we should keep it at that as long as no generally agreed scheme for ID's in the biodiversity information emerges. ObservationUnitIdentifier is a good example: it may or may not be the unit level ID - this is left to the data provider and their domain-specific needs. However, the UnitID must be given, if decided so, it can be the same as the OberservationUnitIdentifier.

1.04. NTH: <DatasetDerivation> - is there a need for this to be repeatable? Our feeling was that it shouldn't be if they have come from one supplier.

WGB: Yes, it needs to be. For example: NHM in Karlsruhe (original source) provides Butterfly data to SysTax in the German GBIF project; SysTax provides data to the Global Butterfly Information System (first derivation), which may be accessed by a GBIF portal, which in turn provides a secondary derivation. I think we discussed this extensively in Singapore.

1.05. NTH: I do not remember why repeatable elements were now to be wrapped in a container element. Was this some kind of good practice technique - or does it just bloat the element count? Either way, there is inconsistency, with some elements wrapped and others not (yet) wrapped.
WGB: Done (I hope). The wrapper elements make all the repeated elements in a document based on the schema addressable under a single path.

1.06. NTH: .. would like a flattened version with an index number against each element ....
WGB: We discussed the question of indexing of elements (and attributes), especially with regard to the simplification of version changes (automatic mapping and re-writing of configuration files and interfaces, etc.). However, we concluded that in any case a mapping old-to-new has to be provided, and that unique identification of elements in types would be not be helpful for version changes anyway. Thus, we should use the individual xpath as the index.

1.07. NTH: As BioCASE is no longer tied to DiGIR, it follows that it is no longer tied to Darwin Core - although this would remain a useful option. Should we determine a smaller subset of access points? Maybe based on the Who / What / Where / When questions?
WGB: Yes, this would indeed be very useful. It should be part of the forthcoming documentation of ABCD, for which all of us lack the time. In Oeiras we already defined on of the name elements (the fully concatenated one) as an access point, same as the InformalName.

1.08. MDO: All elements should have a type.
WGB: Done; MDO wrote a program to test for untyped “leaf” elements.

1.09. MDO: Think about using an optional ID attribute for every repeatable element to allow for better imports (you can detect real duplicates with this - so you know Müller ID=1 is really not the same record as Müller ID=2 within the same dataset). This would be particularly useful for data exchange purposes (import of dataset).
WGB: A dbid attribute added to every repeatable element would be possible. This should be discussed in the context of the relational normalisation envisioned for version 3 of the schema. It also relates to the ongoing discussion about globally unique identifiers (GUIDs).

1.10. SBL: .. narrow the definitions of all these string elements to something more precise than "xs:string". Most are probably xs:Name or xs:NameToken or something like that. The O'Reilly XML Schema book has a pretty good discussion of simple XML data types.
WGB: We think that we should see how things develop first. One of the principles of ABCD is not to be too prescriptive. This puts more of a burden on the developers, of course, but a bunch of invalid documents because of to tight typing would not help us either.

1.11. PST: Since the file is rather large, would not it be better to split it up into senseful pieces, as XML schema offers several options for modularization? I also guess this would improve readability and ease of maintainance. Or vice versa, are there any technical problems hindering this?
WGB: The version for download on the web represented a file concatenate from several subunits (as depicted on the SourceForge site). For this the Includer software provided by Markus Döring (also on the SourceForge site) was used. While the old subdivision into modules largely followed editorial puposes, a more meaningful distribution is provided in versions >1.20. Also, a zip-file with all individual files will be provided besides the concatenated version. The latter is needed for the Schema Viewer to work.

1.12. WGB: Remove OpenIDType and replace occurrances with xs:string.
CCO: No objections, was introduced earlier when the thinking was still more in modelling terms.

1.13. DWA: I note there is not a lot of controlled vocabularies with some of the ABCD elements. I presume that will increase with time.

WGB: See discussion about that subject during the Oeiras meeting (http://www.bgbm.org/TDWG/CODATA/OeirasWorkshop.htm). In some cases controlled vocabularies were introduced (especially in the domain specific parts of the schema, where standards are used). In other cases it will be the work of the providers acting together with specific user communities to ensure that data adheres to certain (sometimes community-specific) term lists.

1.14. PST: Formatting the code using indents would be rather nice for people like me who look at the code directly.
WGB: This has been corrected, the code needs to be imported into a tool like XML-Spy once after it has been concatenated with the Includer-Tool.

1.15. PST: It seems the file was not created using a standard text editor since there are a lot of tags only with whitespace content. In this place may I warn you not to believe any of these tools creates valid XML schema, but rather use an external validation tool.

WGB: Whitespace content was removed and the schema should be valid.

1.16. SRO: Thinking about the PlantGeneticResources elements I miss an indication whether a DataSet will also be provided by another network, e.g. the EURISO-network. This might help to avoid redundancies in the sense that the portal can filter or reject data from the EURISCO-network that are directly delivered by the 'local' data provider or the other way round. (I remember that the BioCASE metadata schema has an element called network.)
WGB: I added the element OtherProviders/OtherProvider to cover this – it's the ID in the UDDI Registry (the UUID) for that provider.

1.17. GHA: The Identifier of a dataset refers to the institution and collection holding a specimen (or a unit dataset). However, databases exist (e.g. within the GLOPP project) that unite first digitizations of specimens from several collections. Moreover, these data may be highly processed and quite different from the data originally taken from the labels in the collections.
WGB: SourceInstitutionID and SourceID were moved to the UnitType. They have thus to be repeated for every individual unit, but this makes it possible to have a dataset

containing units from more than one original source. Documentation of further provision, derivation, or audit trail details of data remain on the level of the dataset.

1.18. GHA, BMO, WGB: The rights and statements types should be united and cleaned up. All statements should be derived from a generalized statement type.
WGB: Done.

1.19. NTH: Before you roll out v1.4, you may wish to check out an apparent anomaly with the ContactType. It appears that there are currently two versions being used in the Schema. There is the condensed version used e.g. in Dataset / DatsetDerivations / Supplier and there is the older, extended version that still pops up in e.g. UnitDigitalImages / ImageIPR / LegalOwner.
WGB: Solved.

1.20. MJA: [Using ABCD ..] the resulting file has to carry a lot of higher tags which are meaningless in the [my project's] context, and many of the tags have names which fail to convey the necessary meaning. A good example is <GatheringAgents><GatheringAgent><AgentText>Sacleux</AgentText></GatheringAgent></GatheringAgents> which I can more simply record as <Collectors>Sacleux</Collectors>. I've concocted a simpler schema .. [and] tried to use element names from the ABCD namespace, but found that many of them sounded awkward in this new context so I dropped that idea. I've tried to keep the semantics and format of the fields the same. I'll have a go at building an XSL to transform the output from this to ABCD, though I guess this is more of a theoretical exercise than anything else.
WGB: This is of course the problem of every individual project, sub-discipline, special interest group, etc. using ABCD. The idea is to provide a generalised schema for all biological collections, which can be used (and be cut down) for many purposes. I should think that actually providing ABCD from your database with an XSL transformation to the simpler format to provide for the project is the better way to go. However, it is already a very important step forward if you keep semantics and format of the elements the same. If you go that way, please do identify this in the annotation of the elements in your schema (and please cite the version of ABCD you are referring to - we are working on a concept-tracking mechanism for different versions and standards in collection data).

1.21. MJA: the SourceLastUpdated field is mandatory but redundant

WGB: The last update can indeed be recorded for the entire dataset or for individual units. On the Dataset level, this is now accomplished by the mandatory Dublin Core element Date_Modified. However, the equivalent unit-level element cannot be mandatory, because most databases can't specify it (although, where present it can be extremely useful). These can at least try and set the one on the dataset level. This is useful in any case, because if no change is indicated, indexing mechanisms do not have to search the individual unit records.

1.22. MDO: The distinction between OriginalSource & DatasetDerivation is unclear to nearly everyone I talked to. What about a clearer seperation between the supplier and the derivations (which may still include a supplier): OriginalSource (not repeatable); Supplier (not repeatable); Derivations (optional, repeatable)?
WGB: The entire structure to describe Datasets was changed.

1.23. OriginalSource/SourceWebAddress should be named similar to the rest: OriginalSource/SourceURL.
WGB: All URLs where changed to URIs, and WebAddress does not exist any more.

1.24. GHA, JKE: Common rules for attributes and element names: element names with capitalised word elements, attributes all lowercase.
WGB: accepted, change in progress.

1.25. DHO: I have noticed a couple of minor glitches in the ABCD Schema naming .. There is no consistency overall between using “-ise”/”-isation” (UK spelling, e.g. “Organisation”, “Mineralisation”) and “-ize”/”-ization” (US spelling, e.g. “Organization”, “Atomized”) for element names. The most obvious case is: GatheringAgent/Organization/OrganisationName.
WGB: solved, used British spelling throughout (exception: xs: types).

1.26. WGB: The choices for UnitCollectionDomain and UnitStateDomain have been hidden under a container element. Since there is no need for this, they have been deleted [v. 1.46d].

1.27. PME [commenting on v. 1.48]: For the ContentContact and TechnicalContact, at the Dataset level we noticed that you can only put one contact. This might be a IPR problem are there are rarely only one person involved. May be it should be possible to add several with a "PreferredFlag" ? This can also be useful if people are on vacation or sick leave to put several contacts. In some cases the providers if they work togehter on a dataset will not allow that only one is cited.
For BCCM data we have people the cite at different levels, we have at the datasets level the coordination team of BCCM. then 4 datasets (LMG, MUCL, IHEM, LMBP) which each time the relevant people to cite. To publish the data at GBIF using ABCD I have to cite all those people at the right place.
[further, on v. 1.49]: We identified clearly that we need repeatable "contacttypes" at different levels : Network, whole datasets or one dataset and finally at the Unit level. The addition of a "MicroAgent" adds confusion. How to decide which "actor" is only a MicroAgent (denomition some people will really not appreciate as it give the impression that some actors are just "Micro")?
WGB: We have to clearly distinguish the technical roles (i.e. TechnicalContact and AdministrativeContact) that go into the registry, from IPR and other legal roles. No ownership or rights are assigned to these contacts. For these there is a clear demand by portals (e.g. GBIF – DHO) to keep them as simple as possible. Both are thus now typed as Microagents (which completely denormalises the name part and thus allows you to find a text that fits your specific organisation. I followed your comment in making the content repeatable to allow for contact during vacation etc. and I included a “preferred” attribute. The same was done for the ContentContact on the unit-level [1.49e]. Any more detailed information on rights and to-be-cited people and organisations should go into the Metadata part, where it is fully repeatable as part of, e.g., the rights statements.
It has been agreed (SDD workshop in Berlin) that metadata should be included for each dataset, even if the TechnicalContact is presently completely repetitive for each dataset included in a single transmission. I also don’t see a problem in repeating the data referring to the coordination team four times, if necessary.

1.28. AIS: I am a bit concerned about the use of the term "unit". It seems to be used for a data record. To many readers, "unit" will mean things like metres and grams.
WGB: we settled on the term ‘unit’ after much discussion in the CDEFD group. The term can be applied to observation records, the object observed (“field unit”), physical objects derived from that field unit, and data records based on such “derived units”. Curatorial unit, Collection unit, Accession, Record, Observation record, among others, are other terms that are (in some contexts) synonymous. See the BioCISE model (http://www.bgbm.org/biodivinf/docs/CollectionModel/) for further details.

1.29. AIS: The naming convention used throughout the schema seems to vary. If I look at NamedAreas I see sub elements like NamedAreaClass. This part seems to indicate a naming that is building on the element name higher in the structure. Then, if I look at other elements (e.g., Country) I see a mix where one element is CountryName (this name takes on the name of the element higher in the structure) and ISO2Letter (this name does NOT take on the name of the element higher in the structure). I think whichever way you go with naming it should be consistent. However, be warned that concatenation of the names will result in some extremely long names that (in my opinion) are not necessary. Just start thinking about the XPath expressions you are going to have to reference the data and you will see the problem with concatenation.
WGB: The convention should be that repitition of the name of container elements should be used only for element names that are used in several places of the schema. Furthermore, repetition of the name of the immediate container should also be avoided, if not necessary for the sake of clarity in discussions (e.g. DatasetGUID vs. UnitGUID). Application of this convention has led to numerous changes (see document comparing the current version with version 1.2).

1.30. AIS: I would encourage the development team to look at the ISO 19115 metadata standard. I know it is a daunting task to read and understand ISO 19115, but it may help with some of the metadata tags for people, phone numbers, citations, disclaimers, etc. Also, this may pay dividends in the end. If the metadata are ISO 19115 compliant, they will ultimately (I predict) be discoverable by crawlers and such things. This has great potential for increased exposure and use of the data.
WGB: Postponed for the next version.

1.31. DLE: We have [introduced an] intermediate level between DataSet and Units called SamplingUnit. This element allows to describe various characteristics of the monitoring events. I continue to think that collection records are really a special case of a SamplingUnit. A sampling unit is the collection of all Units gathered over a particular time and area and the gathering information is now attached to the SamplingUnit rather than the Unit itself. Most sampling event for specimens would possibly contain only one sampling unit, but that change would allow to also accommodate monitoring data.
WGB: I considered to follow this approach but after some discussions and looking ad different scenarios finally (once more) decided against this. The idea of having several units per gathering is very appealing. Actually, the very first large DTD elaborated by Charles Copp as a contribution by ENHSIN for our group had such a feature. I have made a draft following your schema by having a SamplingUnit, that contains 1 Gathering and 1-many Units. In fact, the GatheringUnit than represents what we have earlier described as a GatheringEvent + gathering site. Gathering event items describe the who,

when, how, and under what conditions of the gathering, while the gathering site tells us where (at least down to the scale where the "where" becomes a "how" and/or "under what conditions").
Now, this solves the question of somebody making a lot of observations or collecting a lot of specimens in a single event. However, what happens if somebody is out there watching over a period of time and notes the exact time of every observation (=Unit)? And it starts to rain (change of weather conditions)? Worse still, how do we represent continuous observations (monitoring) where the place is fixed, but even the date and observers change?
I am back to my opinion that this entire problem (i.e.: where in the hierarchy to put which of the time-condition related items of the gathering) can only be solved by a relational structure. XML offers the possibility to use such structures, but up to now (with one exception: UnitAssociation) we did not introduced these because of the difficulties in mapping and interpretation (on the portal side) these structures bring about. This will have to be resolved in a future version of the schema (and the protocol). The actual delimitation of the various structures can probably be investigated statistically (optimal normalisation of units in the GBIF system) at some time in the future.

1.32. MWI (translated): [SysTax publishes information about specimens received from several data providing collections]. The information of these “providers“ is mapped to "/DataSets/DataSet/OriginalSource" and the elements within that element in ABCD [1.2]. Unfortunately, there is no possibility for these collections to provide additional information about themselves.
WGB: We have to distinguish two cases here: (i) the data consists of units from many such sources (e.g. information gathered by a monographer in many collections). In this case there is indeed a problem in v. 1.2, which was solved by moving the SourceInstitutionID and SourceID element to the unit-level. Further information can be supported by the owner, contact, and IPRStatement elements on the same level. (ii) In your case, you actually receive datasets from other institutions. Here it would be more practical to represent them as several datasets within the returned document. In this way you can avoid having to repeat the owner, IPRStatements and contact information for every unit.

1.33. JTO (Salzburg ABCD meeting): If an atomized version of an element is provided than the non-atomized version should be also be provided (make it mandatory)
WGB: This is the case for scientific names, the FullName of a person, and the elements IdentifiersText and GatheringAgentsText.

1.34. AUC (translated): [did not find the following data item of the ITF-2 format in ABCD] Page Code Descriptor (“The International Page Code Number that is specified in the computer's disk operating system and which is normally used by the sending garden in the preparation of data.”)
WGB: XML documents always include a statement specifying the character encoding, e.g. “<?xml version="1.0" encoding="UTF-8"?>”. So, in principle that receiving end has to take care of correct display. However, current data provider software producing ABCD documents (as well as Darwin Core documents) in the GBIF network always use the UTF-8 character encoding, so transformation is normally taken care of at the database wrapper’s side already.

2. General unit-level data (incl. associations, images, etc.)

2.01. NTH: <UnitID Numeric> - should be alphanumeric rather than decimal.
WGB: We introduced this element in addition to UnitID to specifically to allow searches on a range of numbers (where possible). Essentially, it should be a duplication of (the numeric part of) the UnitID. This wouldn't make sense if it's alphanumeric.

2.02. NTH: <UnitAssociation> - text says that the association should be with another unit in this dataset, but the implication of the structure and the preference of the TC meeting (plus logic) is that it should be possible to go across datasets.
WGB: Done. I also deleted the <Rule> saying the same.

2.03. MDO: UnitDigitalImages/UnitDigitalImage/ImageSize should mean the pixel size of a picture. I would reuse ImageSize as this: (1600x1200 pixel) and add an element filesize for the amount of data (12,4 MB)

WGB: Agreed and changed.

2.04. MDO: A remark for UnitID should be added not to use primary keys of a table for this, as it should be assured that the unit IDs keep stable when importing/updating databases.
WGB: Done

2.05. MJA: there are no fields for some image metadata which we will probably have to add in (Image Resolution, Capture Equipment, Date Image Created and Image Creator)
WGB: These have been added to the ImageType, as well as some other items that were identified in the recent ENBI/GBIF workshop.

2.06. PME: Accession Number: Use **SourceInstitutionID**, **SourceID** and **UnitID**, for UnitID not just the numerical part but also the Acronym which is a complete part of the record unique identifier, ie LMG 115 and not just 115. This is for example important as for DMZD the letters of the strain ID is not DMZD but only DMZ. Works are ongoing on a Global Unique Identifier for each strain, but not yet in use.
WGB: Use DSMZ – DSMZ – DMZ115 for the three mandatory ID’s.

2.07. PME: Other culture collection numbers. These are not only previous IDs but contemporary strains that are kept in other collections under a different Identifiers. Suggestions would be to put them in **AssociatedUnitID** with mention of strain duplicates in the **AssociationType**.
WGB: I don’t think the AssociationType is the place to put this. If the history (past and present) is used properly, all the data can be stored there (there is a date for PreviousUnit). If the collection wants to record strains derived from this unit, they should store that new unit in their database or access it by way of their unit id found in the recipients database in the SpecimenUnitHistory.

2.08. PME: History of deposit: The order is important, is there a possibility to use a nested structure?
WGB: The history can either be stored as a text under UnitStateDomain\SpecimenUnit\SpecimenUnitHistory\PreviousUnitsText or as a sequence of references to other units (given by the three-partite ID) under ...\previous

units. If this unit is not available from the original holding institution, it can also be stored in the present institution’s database.

2.09. PME: History of deposit: Important is also to mention the Name under which the strain is known during the history and to add the names and contacts of the Depositor and Isolator. Is it possible to add a **ContactType** in **SpecimenUnitHistory** where the role of the contacts can be mentioned as for the supplier in the DatasetDerivation part WGB: If all this is not stored as a single text (see 2.08), the depositor and date is stored with the individual unit under SpecimenUnit/Acquisition/**AcquiredFrom** (a ContactType), where a date can be stored, too. The isolator is stored under SpecimenUnit/UnitPreparation/**PreparationAgent** (also a Contact type) with the preparation date.

2.10. AGU: The element **ImageURI** provides a links for images in ImageType. Here we probably need a second element to accommodate strings including HTML, Javascript and the like where providers accommodate instructions how to access the images.
WGB: Added element **ImageURIString.**

2.11. MDO: I would like to see a globaly unique identifier for an ABCD unit record, which is independent from the existing 3 parted key made of institution/collection/cataloguenumber.
It is meant for globaly referencing exactly this unit record, which might be different from other unit records held elsewhere, but still talking about the same specimen with the same inst./col./catalogue number. The 3 parted key would still be very important to locate the physical specimen. GBIF is currently examining the use of LSID (Life Science Identifiers) for this or other central registry mechanism to guarantee global uniqueness. As this GUID format is not known yet, the attribute would have to be optional and of type xs:string.
WGB: Added an optional element **UnitGUID** typed String.

2.12. MMA (translated): A unit can be specified as to sex [in Unit/ZoologicalUnit] and number of individuals [in UnitMeasurementsAndFacts]. However, especially in zoology a single unit is often comprised of male and female individuals or mixed with larval stages. Mapping this to ABCD would inflate the number units. I suggest to add at least attributes for number of male, number of female, and number of workers (for ants etc.).
WGB: A controlled vocabulary was introduced using the type BasicSexCodeEnum. Since this can be used broadly in zoology and botany (s.str.), it was moved to the unit level. As you point out, the MeasurementsType was conceived to hold counts. I suggest that in the case described, the value for Unit/Sex is set to “mixed” and for the actual counts you use Unit/UnitMeasurmentsAndFacts/UnitMeasurment/MeasurmentAtomised/ParameterMeassured (e.g. “female”) and ../MeasurementLowerValue. This solution can also be used for mixed stages etc.

2.13. JTO: Change the name of the /DataSets/DataSet/Units/Unit/UnitDescription concept to something like Notes if this is really intended to be used for notes on the Unit. If not, create a new Notes concept.
WGB: Changed the name.

2.14. JTO & MDO: The BioCASE metadata scheme has a CollectionClass: we think this provides the vocabulary for the "RecordBasis" element. But it should be renamed

somehow. CollectionClass is not really adequate for a unit attribute. What about "UnitClass" ?
WGB: I maintained the name for compatibility reasons and because it actually describes the content quite well; I added the BioCASE controlled vocabulary (which will perhaps need some additions).

2.15. AHA: ABCD v. 1.49 (and previous): The concept of digital units seems to be restricted to image data (/Units/Unit/UnitDigitalImages). Where would sound archives fit in, e.g. of cricket or bird songs?
MWI: How about some additional multimedia information, e.g. sounds, videos, etc. ?
WGB: The Image type was revised and renamed MultiMediaObjec.

2.16. EVB: How is information on several preparations from the same specimen kept together? Zoological specimens are often split in eg skeleton and tissue preparations.
WGB: This can be done either in a directed way (e.g. a sample taken from a specimen) by means of UnitAssociations or in a indirected way by means of a UnitAssemblage (in the latter case by assigning an arbitrary common identifier - AssemblageID - to the members of the assemblage).

2.17. EVB: ZoologicalUnitType: not clear what ZoologyPhase is - is this for eg polyp and meduse of the Cnidaria? Is this a field for combined info on phase and stage (latter being juvenile, nauplius, copepodite...)?
WGB: It is, and this has been made explicit by renaming the Element PhaseOrStage.

2.18. MDO/JTO: We suggest to rename the "RecordBasis" element to something more descriptive. CollectionClass is not really adequate for a unit attribute. What about "UnitClass" ? And we suggest the following recommended or controlled vocabulary: living / dormant / culture / preserved / observations / photography / fossil / other
WGB: I prefer to stick to RecordBasis, for compatibility reasons (DwC) and because it adequately denotes what the element describes: the underlying object. I included a controlled vocabulary, namely: PreservedSpecimen /LivingSpecimen / FossileSpecimen / HumanObservation / MachineObservation / DrawingOrPhotograph / MultimediaObject.

2.19. MDO/JTO: We use ObjectClass in the BioCASE metadata profile to describe the objects a collection holds. Feedback from collection holders includes the following terms: whole organisms / antlers / artwork / audio recordings / bark / blood samples / bones / bulbs / claws / cocoons / DNA / eggs / extracts / feathers / feeding remains / fruits / galls / heads / hooves / horns / leaves / mixed / nests / pellets / pollen / roots / seeds / shells / skins / spores / teeth / wood / other. We can’t think of any ABCD element representing this right now, but there are people already asking us for help on this - where to put and how to find objects according to this information. We think "ObjectClass" is allready a quite good name refering to the real physical object. The terms cited are quite extensive but definitely not comprehensive. Should we just mention this list as a recommendation?
WGB: Most of these terms should go under SpecimenUnitPart, although the element’s name is obviously not very intuitive. Also, most of the terms could also refer to observations. “Audio recordings” do not belong here but rather under RecordBasis. “Artwork” must go under UnitNotes, I’m afraid. I renamed SpecimenUnitPart to KindOfObject and placed it under Unit.

2.20. MDO/JTO: We use “PreservationMethod” as another class of keywords to describe collections. It is covered by "PreperationType" in ABCD. Should we include the following list as a recommendation? Or even a controlled vocabulary?
no treatment / alcohol / deep-frozen / dried / dried and pressed / formalin / refrigerated / freeze-dried / glycerin / gum arabic / microscopic preparation / mounted / pinned / other.
WGB: I have included these terms - which represent “real world” input - as examples. I think we should continue to gather information before anything more restrictive is used.

2.21. PME: concerning the element SpecimenUnit/UnitPreparation/PreparationAgent, we experienced while mapping microbial information, that the preparation of the strain or culture takes several steps with several "PreparationAgents". But the PreparationAgent is a not repeatable and has no "Role" associated.
WGB: The UnitPreparation element has been made repeatable, so that a sequence of preparations can be indicated, and the PreparationAgent has been assigned back to Contact type, so a role can be assigned.

2.22. PDA: the strain history recorded in ABCD is the description of the linear history from its deposit into the LMG culture collection back to its point of isolation. Does the ABCD schema takes into account the order of the different subitems [= PreviousUnits]? This is vital for the interpretation of the strain history!! The ABCD schema would only need some minor adaptations to describe the whole strain history as described in http://www.wfcc.info/NEWSLETTER/new/newsletter38/a3.pdf.
WGB: added attribute sequence to element PreviousUnit.

2.23. JTO (Salzburg ABCD meeting): Add a new complex type for number of individuals inside a unit. (example of ants)
WGB: measurements and counts of types of individuals within a unit (sex, role, etc.) are covered by UnitMeasurementsOrFacts.

2.24. AUC (translated): [did not find the following data items of the ITF-2 format in ABCD] Material Transfer Receipt Flag, Material Transfer Supply Data Source, Conservation Status (Threat), Accession Uses (“The description of the (economic) uses of this accession. The term 'economic uses' is used in a very wide sense and incorporates medicinal uses. The information transferred refers uniquely to the economic use of the plant accession record being interchanged” .. not of the taxon itself.).
WGB: to my knowledge these fields (the first two refer to obligations of botanical gardens under the Convention on Biological Diversity, the last two are statement probably recorded by the collector referring to the status at the collection site) are presently under discussion in the Botanical Garden Community. Once clarified, the resulting fields will be introduced to the BotanicalGardenUnit subtype.

2.25. AUC (translated): [did not find the following data items of the ITF-2 format in ABCD] Cultivation Information (“Allows for cultivation information to be passed as free text in order to help the receiving garden care for the transferred accession.”), Propagation Information (“Allows for information about any propagation requirements to be passed as free text in order to help the receiving garden propagate the transferred accession.”), Perennation Flag (“A code to indicate the means of perennation, providing a means of noting living plant accessions that require regular curatorial monitoring.”), Breeding System (“A code to indicate the breeding system of the accession.”)
WGB: These were now added to the BotanicalGardenUnit subtype (elements Cultivation,

Propagation, Perennation, BreedingSystem) presently without the controlled vocabulary in some cases specified by ITF-2.

2.26. MWI (translated): No element for the location of a plant in the botanical garden?
WGB: Was added to the BotanicalGardenUnit subtype (element LocationInGarden).

3. Identification/Taxon issues

3.01. DHO: We are still using TaxonIdentified as well as Identification. If the Synecology element is staying it should use Identification.
WGB: I don't think so. The synecology element records observations taken at the time of the gathering event of a specific unit, so the identification event data should here be the same as the gathering event data. Using formal identifications at this place would unnecessarily increase nesting. For direct synecological observations, individual units would be created for each of the observed taxa, so in this case the identification type would be used.

3.02. DHO: I think that FullNameAuthorYear [changed to FullScientificNameString] is now intended to be the basic location for entering a scientific name. It is after all mandatory. Are users supposed to be able to use this field to enter whatever they have as the best available identification ("A-us b-us Author 1972", "A-us b-us Author", "A-us b-us", "A-us sp.", "A-idae sp.", etc.), even if it is not complete?
WGB: Yes; I hope the new naming makes this easier to recognize.

3.03. AGU: The Complex type given under /DataSets/DataSet/Units/Identifications/Identification/ScientificName/ScientificNameAtomized/Zoological/CombinationAuthorTeam does not have children.
WGB: I think this was solved for v. 1.20 already.

3.04. HSA: Identification should have a person (ID) and timestamp.
WGB: Identification has one or more Identifiers (which may be a person, another [legal] body, and/or a reference from which the entire event was taken. With respect to PersonID see comment under 1.03.

3.05. HSA: Maybe [Identification should have] also some attribute giving the certainty % by which the identification was done to that level (example 100% to genus, 50% to species).
WGB: I would like to refer this to the groups handling due process and best practice in collection digitization. I personally doubt that we can introduce anything numerical here. In Botany, there is a tradition of expressing a certain doubt in an identification by using, e.g., the abbreviation “cf.” (which is accommodated in the IdentificationQualifier element).

3.06. HSA: Identification should also have the method by which it was done.
WGB: Up to now, only the IdentificationNotes could be used to indicate, e.g., that molecular or acoustic methods were used for the identification. Especially in observation records, this can be an essential element for data quality assessment. Again, there is room for standardisation, for the time being, I added a simple text field.

3.07. HSA: I don't think the simple Identification History element ... is useful.
WGB: I came across databases, which dump the preceding determinations in a text field

once a new identification was made. As in many cases, we'd prefer the structured version, but we want to make sure to get the data even if it's not properly structured.

3.08. HSA: In Identification InformalName Language there should be a provision for giving something like Taxonomic Serial Number of ITIS. I don't think TSN is just a "language" as the current schema would accept it. If such number or code is given, the namespace should be given as well.
WGB: An identification should always reflect the result of an identification event (i.e. a decision taken at a certain time and by somebody). Linking the result to an external name-based system like ITIS runs the risk of changing the actual results (e.g. by later changes in the taxon circumscription given by those systems). I think this is a good suggestion for the future, when concept based taxonomic services become available on the network.

3.09. HSA: Indeed this need for being more specific on namespace that "NomenclaturalCode=" applies for ScientificName as well. This should point to some Taxonomic Name Service with closed set of names and URIs.
WGB: There are two issues at hand here. (i) The respective subtype of NameAtomized indicates the fact that an atomized name is structurally handled according to a specific code of nomenclature. Keep in mind that issues like the validity of a name, publication, synonymy, homonymy etc. don’t refer to the results of identifications of collection units but are to be treated in the context of taxonomic and nomenclatural systems. (ii) The taxonomic domain for any identification (also those containing only a scientific name string or even an informal name) can be indicated by using the HigherTaxon element. This is recommended to simplify searches.

3.10. NTH: <InformalNameString> - should be repeatable.
WGB: Look at the new name subtype. Different names should ideally be the result of different events.

3.11. MDO: A controlled vocabulary should be provided for higher taxon ranks.
WGB: Done.

3.12. MDO (following the subgroup discussion in Oeiras): Incorporate DHO’s new IdentificationType structure to make the concatenated name mandatory.
WGB: Done.

3.13. DHO: During the WFCC meeting, we discussed mappings between their standard data sets and the ABCD. Most things I understood, but I could find no ABCD element in which to encode the host taxon or substrate from which the microorganism was collected. Is this supposed to be part of the GatheringEvent? How would you expect such taxonomic identifications to be included? Surely not just buried in Synecology?
WGB: Your question clearly shows that we should have had a microbiologist in Singapore, I think we eliminated the two elements needed to cover substrates during that meeting.
In fact, this is a wider question relating to directed relationships between objects (virtual or real) in the dataset. This includes also questions like host/parasite relationships, organisms found in the stomach content of predators etc.
There are several cases here: (1) We have two units which are related to each other. (2) We have a single unit consisting of two organisms that are related to each other. (3) We

have a single unit consisting of one organism and its substrate, which is (part of) the gathering site [e.g. a type of surface of non organismic character, rotten wood of unknown origin, a cellar wall, etc.).

Case 1 is covered by the UnitAssociationType where the first three elements identify the second unit (e.g. the sample of the substrate organism) and the AssociationType defines the relationship (e.g. "grows on" or “isolated from”).

Case 2 may, theoretically, be resolved into case 1 by actually creating two derived units from the single one (taxonomic homogeneity principle in the BioCISE model - the moment you define a new, previously undefined taxon in your mixed sample, you effectively create two new derived units in your database).

However, in practice we will continue to have two qualified identifications of the same unit. I looked all over ABCD, this is something that was there and is now missing. I think it was the attribute "SpecialTargetCategory" of "Identification" which we did away with in Singapore because we couldn't find out what it meant. So this has to go in again, perhaps Role would be a better name? At a later stage we probably need to provide a controlled (but extendible) vocabulary.

As a result, the host / parasite relationship (where there is only 1 unit) would turn into two preferred identifications, one of which has Role="host", the other Role="parasite". Another example would include “substrate” and “isolated strain”.

This is the clean solution, because it allows to involve the identifier of the host (eg. a plant) as well as the identifier of the organism using it as a substrate. I realize that many databases do not have that information, but they should be encouraged to provide it.

Case 3 was originally covered by the GatheringSite, but it isn't any more because we actually restricted it to the geographical/ecological side of things. I first thought to propose to include a new element to UnitCollectionDomain/CultureCollectionUnit/ (e.g. SubstrateMaterial). However, I keep coming back to CDEFD and BioCISE modelling results and I believe Gregor Hagedorn first brought up the thought of treating such matters as non-taxonomic identifications. Sometimes considerable effort is made to provide a good description or identification of the substrate. This would also open up the way for the inclusion of identifications of other collection objects (e.g. minerals) that are housed side by side with biological objects in natural history museums. Consequently, I have simply extended the IdentificationType with an element MaterialIdentified as an alternative for TaxonIdentified.

3.14. JKE: You explained that relationships between different identifications of a single unit (e.g. host/parasite) can be expressed by two identifications with different roles. How is that role defined?

WGB: The attributes and elements of the Element UnitDataType/Identifications/Identification further describe the role of an identification in the unit-context:

Attribute PreferredIdentificationFlag to designate current identification. In cases where more than one name applies to a single unit, several identifications should be formed and marked as preferred. Attribute NonFlag to designate negative identifications. Element Role to designate the role of the identifcation result (e.g. substrate/isolate, host/parasite, etc.). This should not affect the Names&Concepts standard.

3.15. JKE: Why are TaxonIdentied and MaterialIdentified not alternatives?
WGB: They are now (post v. 1.30) - under a new element IdentificationResult. This will allow further extension of the identification type to cover other areas (e.g. minerals).

3.16. JKE: What is the purpose of the HigherTaxa/HigherTaxon element in the TaxonIdentifiedType?
WGB: It represents a classification of the identification result, not an identification result in itself. It is used (and demanded) by collection holders as a means to include their classification into the result. It is also useful for information access as long there are no complete and effective taxonomic thesauri available for query expansion.

3.17. JKE: Are the elements ScientificNameString and AuthorString not just duplications of elements under NameAtomised and FullScientificNameString? Is this third level of representation of names really necessary?
(DHO expresses similar opinion, see under 3.25 below)
WGB: This was maintained only for compatibility reasons (Darwin Core). Both elements have been removed for ABCD v. 1.43.

3.18. JKE: What is the role of the elements NameAddendum and IdentificationQualifier under ScientificName in the TaxonIdentifiedType?
WGB: They represent additions to the scientific name in the strict sense - used to express an insecurity of the identification (IdentificationQualifier) or a specification of the concept (NameAddendum). They are only to be used when a scientific name is the result of the identification. From v. 1.40 on, the scientific name was turned into the ScientificNameIdentifiedType, which in turn is an extension (for ABCD) of the ScientificNameType, which consists only of the name elements adhering strictly to the codes of nomenclature, namely: NameAuthorYearString (now renamed FullScientificNameString), ScientificNameString, AuthorString, and NameAtomized (entire container). The extension ScientificNameIdentifiedType includes the non-code items NameAddendum and IdentificationQualifier.

3.19. JKE: If the result of identification is the name of a higher taxon, where is the scientific name to be placed?
WGB: Originally, we had it covered by HigherTaxon, but this is now only to be used for classification purposes (see above). Neither the description of the FullScientificNameString nor the atomised name structure currently (v. 1.30) accommodates a suprageneric monomial as the result of an identification. I have changed the annotation of the FullScientificNameString and the Genus elements (now: GenusOrMonomial) in Botanical and Zoological structures accordingly (where the Codes define higher taxa - I have to look this up for bacteria, as far as I remember they don't exist in Viruses).

3.20. PME: The possibilities suggested by WGB for the concept of “isolated from” (see OECD minimum dataset .. and ABCD 1.30, 28. Jan. 2004) are interesting but the information is impossible to split in these different categories for the existing databases, where names and materials are usually mentioned in the same ‘substrate’ field.
WGB: A search for the host or substrate organisms together with other records of that organism would of course be very interesting (e.g. “Are there cultures of microorganisms from Citrus?”). For the time being, I suggest to put all substrate records into Identification\IdentificationResult\MaterialIdentified.

3.21. MJA: I can't see an intuitive way to identify the typified name and the name stored under
WGB: The typified name is in ABCD stored together with the information on the verification of the type status etc. under NomenclaturalTypeDesignations. We have long pondered if to treat this as just another identification, but there are important differences: the only possible result is a scientific name in the strict sense (see v. 1.41 complex type) and there are several addtional data items which would extend and confuse the normal unit Identification construct. Furthermore, if treated as a simple identification and flagged, it would become ambiguous once several designations are made for a single specimen. The name should thus be entered in the NomenclaturalTypeDesignations area when properly veryfied; nevertheless, it can also be cited as a "normal" identification. In contrast, the NameStoredUnder is just another type of taxonomic identification and can thus be covered by an attribute or an extension of the identification type. In v. 1.4 now a flag in the IdentificationType is used to store it alongside with other flags.

3.22. MJA: there is no field for the species author [for names of infraspecific rank]
WGB: This is part of the systematic hierarchy and not required by the code. We are open for discussion on this point, but including the systematics of a name in a schema focussing on units is - if at all deemed necessary - a transient measure, since this function should eventually be replaced by a taxonomic system. However, the same argument holds true for the HigherTaxon hierarchy included - it is just there to facilitate searches for the time being.

3.23. YDJ: I found non-atomised types for authorship and year citation for the 'NameZoologicalType'. An explicit choice?
WGB: Yes. After lengthy discussions we decided not to further atomize author citations, neither into name(s) + year in zoology, nor into name(s) and ex-author name(s) in botany. This is a pragmatic decision; in the zoological case, there is good reason to separate the year, but this would mean introducing year and authors without year as two new elements.

3.24. DHO: For consistency can’t Genus become GenusOrMonomial for NameBacterialType and NameViralType? In each case it may be a family or higher.
WGB: For bacteria this is true and the name has been changed. For Viruses the nomenclatural status of monomials is less clear, but common usage includes “family names”, so this was changed, too.

3.25. DHO: Regarding 3.17, I am increasingly concerned to avoid unnecessary proliferation of different formats. In the GBIF Portal I have an enormous amount of code dedicated to attempting to parse scientific names from the Darwin Core records we are retrieving. The number of ways that data get mapped from databases to the different fields is a nightmare with even the relatively few, well-explained elements in that schema. Some problems are formatting problems (entirely upper case, entirely lower case). Others involve an inability to atomise data to the level required. The ScientificName element in that schema is intended to hold just the genus and epithets, with the ScientificNameAuthor holding just the author citation. I have had to work with providers providing a combined string with all of these elements in the ScientificName field, others who provide everything in one field and parts in the other, and yet others who put everything in both fields. At the same time I have all of the Darwin Core atomised fields to consider. This means that automated parsing of names is fantastically complex (something which does not show up so much if the data are simply formatted and displayed to users as a set of elements). With even more options to consider with ABCD it could become almost impossible to ensure that different tools interpret the names in a consistent way. From the standpoint of software processing I would prefer to see all names collapsed into a much more minimal set of elements. (What is a provider supposed to do if it currently holds a database with records spanning different nomenclatural codes? It could be very hard to generate appropriate NameXxxType elements for everything.) I would strongly recommend moving towards a more simple model, such as:

TaxonIdentifiedType	HigherTaxa	HigherTaxon*	
	ScientificName	FullScientificName	
		Name Addendum	
		NameAtomized	GenusOrMonomial
			Subgenus
			FirstEpithet
			SecondEpithet
			Rank
			HybridFlag
			AuthorString
	CombinationAuthorString		
	CultivatedPlantNameElements		
			Breed
			NamedIndividual
		IdentificationQualifier	
	InformalNameString		
	NameComments		

WGB: The first level under TaxonIdentifiedType is identical under ABCD, only that ScientificName and InformalNameString are explicit choices. At TDWG in Lissabon, there was some discussion about this point, and it is clear that we always have some garbage in the data, but this should be resolved by either placing all as informal names or to or live with the dirty FullScientificName data (and filter them out at the portal's side in the case of name queries).

For the ScientificName in your schema (=ScientificNameIdentified in ABCD) we have also the same elements, only that the identification-specific items are put together as a type extension of the scientific name itself.

The differences is in how to treat atomised names. Your solution strongly resembles the one in an earlier version of the schema. After some discussion, we thought that it makes sense to separate the names into the different codes, for several reasons, among them:

Leaving the communities in charge or "their" name structures

Data integrity rules differ among the Codes and may change differently in the future

Comparison against standard nomenclators is facilitated (e.g. Bacteria!)

Element names can be named according to community usage (e.g. what's the author string?

All this and some more I don't recall right now led to the present structure. However, we anyway have to come back to this question once we have the Names standard defined.

3.26. MDO: Why do the names of the concepts FullScientificNameString, ScientificNameString, and AuthorString contain the term "String"? This is not the case in other elements. Should this be omitted?
WGB: I stand to be converted, but I think this is the only case in the schema where elements actually represent a concatenated string of other elements (those in the NameAtomised section).

3.27. DHO: Is there any chance/desire for any simplification of the TaxonIdentified and/or ScientificNameAtomized elements during this revision cycle for ABCD? I'm not sure when Walter is back and how that relates to the TDWG dates, but I'd really like to see a simplification for the sake of application development.
If someone is performing a federated search for specimens from the genus A-us (and I do not know in advance what kingdom), I need to build a compound looking for "A-us" or "A-us %" in any of the following:
/DataSets/DataSet/Units/Unit/Identifications/Identification/TaxonIdentified/NameAuthorYearString
/DataSets/DataSet/Units/Unit/Identifications/Identification/TaxonIdentified/ScientificNameString
/DataSets/DataSet/Units/Unit/Identifications/Identification/TaxonIdentified/ScientificNameAtomized/Bacterial/Genus
/DataSets/DataSet/Units/Unit/Identifications/Identification/TaxonIdentified/ScientificNameAtomized/Botanical/Genus
/DataSets/DataSet/Units/Unit/Identifications/Identification/TaxonIdentified/ScientificNameAtomized/Zoological/Genus
/DataSets/DataSet/Units/Unit/Identifications/Identification/TaxonIdentified/ScientificNameAtomized/Viral/Genus
Of course it may be simpler if providers follow recommendations on mandatory elements, but this is the practical reality.
WGB: I certainly hope that we will be able to ask providers to follow recommendations on mandatory fields, and from what I hear of current ABCD implementations, this is (in this case) no problem. Searches should be done on the concatenated field.

3.28. GDI: ABCD Botanical Name, SecondEpithet is described "The infraspecific epithet, i.e. the epithet following the indication of the infraspecific rank in the name string" The name SecondEpithet suggests to me that if there were a third epithet as in a quadrinomial (variety of a subspecies) you are to enter the subspecies name here. Is this what you intended?
Chapt. III, Sect. 5, Art. 24 of the Tokyo Code says: "24.1. The name of an infraspecific taxon is a combination of the name of a species and an infraspecific epithet. A connecting term is used to denote the rank. Ex. 1. Saxifraga aizoon subf. surculosa Engl. & Irmsch. This taxon may also be referred to as Saxifraga aizoon var. aizoon subvar. brevifolia f. multicaulis subf. surculosa Engl. & Irmsch.; in this way a full classification of the subforma within the species is given, not only its name."
The code seems to give more importance to the final epithet. Our database at UNA only has a field for one infraspecific epithet and we have been recording the last one. Tony

Kirchgessner at the NY Botanical Garden said they are putting the last epithet for Subspecies in their Darwin Core table. If you mean for the final epithet to be entered you could say "... after the indication of the _last_ infraspecific rank in the name string". You could add "In the case of a variety of a subspecies, record the variety."
WGB: the annotation was made clearer and examples were included. The atomised section of the schema should use names according to the Code; the text cited implies that a quadrinomial is not a name.

3.29. JTO: -Add a parameter in the ScientificName concept called something like “NomenclatureCode” with a control vocabulary like: Bacterial, Botanical, Zoological, Viral .Not mandatory, but highly recommended!
WGB: Done.

3.30. SGI: Taxon Identified/Scientific Name/IdentificationQualifier [annotation needs clarification].
EVB: IdentificationQualifier: comments are not clear to me.
WGB: Such qualifiers are found in identifications (at least botanical ones) and many botanists try to confer different meanings by the position in the string, e.g.:
”cf. Abies alba”: "I think this is Abies alba, but compare it" [to well identified specimen, to description]
”Abies cf. alba”: "I know this is an Abies, but compare if it is really A. alba"
”Abies alba cf. subsp. alba”: "I know this is Abies alba, but compare if it is really subsp. alba".
It gets more complicated with "aff." designations, because these are really negative identifications:
”Abies aff. alba “ means that this is Abies, but probably not A. alba (but something close to it).
The annotation was made more explicit and the attribute name was changed to “insertionpoint”.

3.31. EVB: How do you deal with incomplete identifications above rank of genus (all too frequent in our parts).
WGB: The mandatory element for names is the FullScientificNameString, which allows the entry of higher ranked taxon names and which should ALWAYS be filled. The NameAtomised section is to give users the chance to ALSO access atomised data items, if the provider has them. This can be very useful, for example to harvest data for specimen duplicates in botany, or for error control against standard lists. In the case of an identification only consisting, e.g., of a family name, the field GenusOrMonomial can be used to store this here, too.

3.32. EVB: NameAtomised Botanical: only one taxon for a hybrid?
WGB: We treat only named hybrids presently, hybrid formulas, where used, are accommodated by the FullScientificNameString

3.33. EVB: NameAtomised Botanical: No subgenera/sections? Only two epithets? I thought botanists have regularly more? If FirstEpithet is always Specific epitheton, isn't it better to make this clear from its name?
NameAtomised Zoological: names below subspecies are not regulated by the code, but are often listed.
WGB: (see also 3.28 above). The NameAtomised section explicitly refers to Code-

conformant names. All other constructs can be put into the FullScientificNameString. The epithet of a generic subdivision (section, subgenus) is accommodated by the FirstEpithet (hence the name).

3.34. EVB: NameAtomised Zoological: CombinationAuthorteam and year are normally not known; only known by the hardcore taxonomists - would make sense in a taxonomic structure, which, I assume, this does not pretend to be.
WGB: open to discussion, but the zoologists in our group seemed to prefer to include it.

3.35. EVB: NameAtomised in general: have you considered a hierarchical structure (with links to a parent taxon for every taxon) rather than the flat one proposed here?
WGB: The NameAtomised refers to a name construct as defined by the respective Code of Nomenclature. The name itself does not include relationships to other names (e.g. inclusion in taxa of higher rank), although a classification relationship is clearly implied (by the re-use of generic and specific names at lower ranks). Building such a hierarchy (and the many alternative representations, concepts, etc.) is clearly not the purpose of the schema for collection data. The TDWG NamesAndConcepts standard group tackles this problem.

3.36. PKI: Change name for botanical SecondEpithet to InfraspecificEpithet
WGB: Done.

3.37. MDO: Is the sequence of higher taxa important?
WGB: not really, because this element will largely serve as a search and perhaps indexing aid.

3.38. DGL (translated): [NameViral] should contain an element for the Acronym, which is part of the nomenclature defined by ICTV (International Committee on Taxonomy of Viruses). Useful would also be a place for the Isolate name, which is often queried (at least by Phytovirologists). Although the isolate name has no official standing, the ICTV is providing a search option on its website
http://www.ncbi.nlm.nih.gov/ICTVdb/Ictv/index.htm
WGB: The Acronym was introduced into the NameViral type. Isolate name has to be discussed, I suspect this belongs under unit (microbial strain collections) rather than under the taxon name. However, perhaps a separate ViralUnit subtype should be created.

3.39. JTO (Salzburg ABCD meeting): Is it possible to provide Name Ids for names in reference lists such as IPNI, or ITIS)
WGB: I prefer to wait for the outcome of the Unique Identifier discussion here. IPNI would be appropriate, because it is indeed a name index. Unforatunately it is restricted to Botany. ITIS ID’s really stand for concepts of names, which may change over time, so it’s not a good idea to include them for an identification result.

3.40. JTO (Salzburg ABCD meeting): Provide a flat list of all nomenclature codes concepts and then flag the code that is being used. Performance gain for most providers. Better mapping against Darwin Core.
WGB: see 3.25 above. I had hoped for some guidance from the discussion of the Linnean Core, but that schema is focusing on nomenclatural issues like name relationships, so it is perhaps not of much use in this context. I prefer to stick to the present solution presently, and to revisit that question when we have more data to investigate.

3.41. JTO (Salzburg ABCD meeting), JHO, MDO: Currently there is no possibility to have multiple identifiers.
WGB: Solved in v. 2.00.

3.42. YDJ: Since in zoological databases the 'author' and 'year' data of the taxon authorship are often kept in separate columns it can be considered to atomise the zoological authorship element, now kept within both the DarwinCore and ABCD schemes as merged elements, also as separate elements.
WGB: I think this (as atomisation of the author strings) is more of an issue for the name standard. For names as the result of an idendification I think the present schema is sufficient. However, if we want to use ABCD more as an exchange schema, we will have to consider to atomise both, author teams, and years (and literature references).

3.43. AUC (translated): [did not find the following data item of the ITF-2 format in ABCD] Verification Level ("The level to which the identification of the plant has been verified.")
WGB: This was added to the Identification as a new element, but since the indenfication can refer to other results than organism names, the controlled vocabulary provided by ITF-2 is not enforced.

4. Contact / Person / Organisation issues

4.01. DHO: I think we all agreed to remove SortPersonName from PersonName, and therefore to replace PersonNameType with a plain string.
WGB: Well, this was the Singapore editorial meeting's opinion. However, I got other suggestions later that support a separate type for person names (e.g. HSA wants a possibility to include an ID here, and that does make sense if we make progress on a 'Naturalists Directory' as envisioned in the SYNTHESYS project). I agree that we should rename the SortPersonName element, because it is really much more important for queries than for sorting purposes. The other element is necessary to maintain original text where wanted. In version 1.3 I called them PersonName and PersonNameLastFirst.

4.02. HSA: "PersonName" contains a similarly named element. The containing element better be called "Person".
WGB: The type has been changed accordingly (already in v. 1.2)

4.03. HSA: Person should have an element "PersonCode" that would support any person identification schemes and remote namespaces. Using such identifier, all the person data that is repeated could optionally be retrieved from a directory server somewhere. Keeping track of and certying observers is important if there is no specimen to go back to for verification.
WGB: I put this suggestion in the annotation of Person, it should be implemented once such directory servers become operational [we are waiting for a GBIF initiative here].

4.04. HSA: Person should have a public key. In case someone verifies an identification, there must be a possibility to leave a digital signature with the identification.
WGB: Another suggestion that I think is a little ahead of current usage, but I have included it in the editorial comments of the PersonType, too.

4.05. HSA: Related to above "OrganisationCode" should have a qualifier to identify a namespace, for instance a directory server.

WGB: Another suggestion that I think is a little ahead of current usage, but I have included in the editorial comments of the PersonType, too.

4.06. MDO: All agents should have the possibility to provide a LogoURL to be used in interfaces. Currently this is only possible as a supplier.
WGB: Done.

4.07. PME: [commenting on v. 1.48, rephrased] at the unit level different actors (isolator, depositor, and so on ...) can be nicely placed and defined as to their "Role". However, you cannot relate a Role to an organisation. Would it not be a solution to simple have a repeatable standard ContactType with each time a Role associated at each bigger level cited here above. There are so atypic "Roles" and works nowadays that it might be hard to foreseen proper MicroAgents or contactstype everywhere ? Associated to the ContactRole can be a list with all already known or accepted "Roles" but still extensible either freely or moderated.
WGB: The Role element in the contact type refers to the entire information of the element. I can thus refer to the Person (which belongs to the organisation cited) or to the organisation (e.g. if no person is cited). If you do need to assign different roles, you have to cite two contacts, one for the organization, on for the person. The respective elements were now made unbounded, to allow for this and other options (see 1.27).

4.08. JTO (Salzburg ABCD meeting): Contact/Organisation/Units: bad named
WGB: renamed to OrgUnits/OrgUnit

4.09. JTO (Salzburg ABCD meeting): Is usful to have a PersonNameLastFirst concept? (individual name atomized)
WGB: The AtomisedName provides that concept (largely following Gregor Hagedorn’s ideas).

5. Gathering event

5.01. AGU: Collectors are represented in ABCD with an unbounded Element GatheringAgent. Alternative text representation is missing.
WGB: Has now been included. Should we make this obligatory and use it as the search access point? This would than work in analogy to the taxon name, only that a substring search should be used because we don't have a Code for person teams.

5.02. HSA: "DateTime" under Gathering should have a qualifier available to designate whether the element concerns Period="Begin" or Period="End". It is true that this could be inferred from date but it cannot be inferred from time.
WGB: I think this is (now?) fully covered by the current DateTimeType in ABCD. The *Begin elements and *End elements for ISODateTime as well as TimeOfDay (in combination with either a ISODateTime expressing only date or a JulianDay) fully specify the period.

5.03. NTH: There is some duplication in <GatheringEvent> now that the <DateTimeType> includes elements such as <Calendar>
WGB: I suppose that this had been solved already in version 1.20?

5.04. AHA: GatheringAgent (/DataSets/DataSet/Units/Unit/Gathering/GatheringAgents/GatheringAgent): basically the same question as the one about NamedArea (6.12.: problem of documenting the

sequence of element repetitions) applies here as well: the sorting order seems to depend on the order in which elements are delivered by the provider database which is not very reliable, except for the first collector of a team (PrimaryCollectorFlag). Also here I would like to find something equivalent to a sequence number (attribute?). The fast alternative is using concatenated output in GatheringAgentsText instead, but I assume the aim is to provide the atomised data where they exist.
WGB: The same answer as under 6.12 applies: a sequence number was included facilitate the output.

6. Gathering site

6.01. AHA: The element <SiteText> (free-text description of collection site) does not exist any more in ABCD 1.01. Since site information very often just comes as a text field in provider databases, I think we need this element. As a fallback-option the element <LocalityText> could be abused, but this is dangerous (since it is intended to just carry the original label information).
WGB: You are right about not using LocalityText to further specify the collection site, this is for the entire label text as far as it concerns the collection site. I think the element AreaDetail should be used for free text descriptions of the site where these are given in addition to the atomized (higher-level) elements such as country, nearest named place, etc.

6.02. NTH: Many of the separate elements under <GatheringSite> which is by far the most unwieldy part of the schema, could be eliminated in favour of using either <LookupType> or <MeasurementType> which is what many of them actually are anyway. By using these generic structures, the element count could be reduced and the flexibility enhanced. For example, <Altitude> is a just a measurement.
WGB: This was done for version 1.2.

6.03. NTH: For observations, there may need to be a bit more descriptive material. Steve Wilkinson will propose a structure which I will forward on receipt. He felt that what was currently in <Biotope> is not sufficient.
WGB: I am looking forward to this, and we are in direct communication with the group in the course of connecting their observation warehouse system to BioCASE.

6.04. AGU: Change ISO2Letter to ISOCountry and allow entry of the 4-letter ISO3166-3 codes for formerly used country codes.
WGB: Changed Element name to ISO3166Code and deleted ISO3Letter (is also defined in ISO3166-1).

6.05. DWA: you have the examples of the 2 letter and 3-letter country codes mixed up.
WGB: Should be all right now.

6.06. SBL [during Oeiras meeting]: Gathering biotope measurement is time dependent so it should be under event.
WGB: The elements under GatheringSite are all more or less time-dependent. Countries change their borders, named places change their names, etc. It is very difficult if not impossible to actually define the borderline. I have thus finally taken the conclusion to

remove the container element GatheringSite entirely and put all its elements directly under Gathering.

6.07. JWI: .. I was searching in the ABCD schema for elements to capture the paleontological concepts of Period and Epoch and I was unsuccessful. Do they exist? If so, where are they to be found? If they do not exist, should they? They are strange concepts to categorize since they are kin to collecting events, but not in the same way as for non-paleo disciplines.
WGB: Period and Epoch are indeed site descriptors within the context of the collection event. They are covered by the ChronostratigraphicTerm Element in the StratigraphyType (now part of the ExtensionPalaeontological.xsd sub-schema). In ABCD, they belong to the Stratigraphy Element of the Gathering.
ARI: Period and Epoch are "used and abused" chronostratigraphic terms. General acceptance is that the terms "System" and "Period" are interchangeable, and "Series" and "Epoch" are interchangeable. Caution is required when looking at actual usage. The International Commission on Stratigraphy chart is a good starting point: http://www.eas.purdue.edu/chronos/Divisons_GeolTimeUSGS.pdf but you'll find many variations (for instance the chart the British Geological Survey uses).

6.08: SRO: Form my point of view the wording of following elements is a bit confusing: LocalityText, AreaDetail, Site-Coordinate. Following the comments (January 2004) I understand that LocalityText is the original text on the label and AreaDetail is for additional information (not given on the label, like fieldnotes?). I would call the descriptive data SiteText according to Site-coordinate instead of LocalityText. SiteDetail is for descriptive data like field notes.
WGB: LocalityText is the original label (or field notes / original data entry) text. The following elements are atomised elements within that text, namely areas and further descriptive text (given on the label, i.e. not like fieldnotes). Indirectly you touch upon an important issue – how to distinguish derived data from original data. Up to now, this is done only in the country type, where there is a single element for a derived country name (e.g. the English translation). We will further discuss this issue.

6.09. SRO: LongitudeDecimal, LatitudeDecimal. Are the elements Degree/Minutes/Seconds removed? Of course the decimal degrees are easier to handle, but many providers do only have DMS and are not willing / able to calculate decimal degrees. (In addition errors are more obvious and easier to detect looking at DMS).
WGB: yes, the elements are removed and here we would like to try and have only two element to search in. Our experience from implementing ABCD wrappers is that with a little assistance all providers having clean Degree/Minutes/Seconds data easyly adapt to conversion. An alternative that allows to accommodate Degree/Min/Sec would be using part of the ISO standard described under 6.10. but this would also deviate from usage in Darwin Core.
AHA: if the alternative is to convert DMS (database-side) into either ISO standard or decimal degrees, I am in favour of decimal as the easiest format for display in maps (UI level) - this would also help the error detection mentioned. Regarding the 'unwillingness' to convert: it is true that many users have DMS; however, also this is not uniformly done (sometimes just one text field, sometimes three attributes, sometimes with symbols °,',", sometimes...), which means that in most cases conversion would be necessary anyway.

Your suggestion of supplying database-side help to generate decimal degrees from DMS would help both causes (if it can be introduced as a service function), since mapping could be done via the new decimal number, while the database itself could still maintain DMS in addition.

6.10. AKR: Lat/Lon can be represented using an ISO Standard (see http://www.ftp.uni-erlangen.de/pub/doc/ISO/english/ISO-6709-summary):
ISO 6709:1983 "Standard representation of latitude, longitude and altitude for geographic point locations" a format designed for usage in human readable compact file formats, protocols, etc. The standard allows both a minute/second representation as well as a decimal fraction representation.
Latitude can be represented as

DD.DD	degrees and decimal degrees
DDMM.MMM	degrees, minutes and decimal minutes
DDMMSS.SS	degrees, minutes, seconds and decimal seconds;

prefix with + north of and on equator, and with - south of equator.
Likewise, longitude can be represented as

DDD.DD	degrees and decimal degrees
DDDMM.MMM	degrees, minutes, and decimal minutes
DDDMMSS.SS	degrees, minutes, seconds, and decimal seconds;

prefix with + east of and on prime meridian (Greenwich), and with -west of Greenwich up to the 180th meridian.
Leading zeros are required for latitude and longitude.
Optionally, append altitude in meters (prefix with + above and on the geodetic reference datum and with - below it).
If a termination character is needed in the format, / is recommended.
Examples:

```
+40-075/
+401213.1-0750015.1+2.79/
+40.20361-075.00417/
```

WGB: See comment under 6.09. If implemented, this would not really make things easier for providers, and neither it would be more standardised, because latitude and longitude should definitely remain separated into two elements and altitude kept excluded.
AGU: In any case, here is the regular expression for the standard (without altitude):

```
<xs:pattern value="[\+\-]([0-9]{2}.[0-9]{2}|[0-9]{4}.[0-9]{3}|[0-9]{6}.[0-9]{2})[\+\-]([0-9]{3}.[0-9]{2}|[0-9]{4}.[0-9]{3}|[0-9]{6}.[0-9]{2})"/>
```

6.11. BRI: I'm particularly interested in how I might go about transferring numeric geographic data, such as a geocode taken using a GPS device, and the datum that GPS was set to when the value was collected?
WGB: I think this is covered by elements under SiteCoordinates in the GatheringType, as long as your “geocode” represents coordinates. The usage of the GPS itself is covered by CoordinateMethod, the SpatialDatum and the coordinates themselves unter CoordinatesLatLon.

6.12. AHA: NamedArea (/DataSets/DataSet/Units/Unit/Gathering/NamedAreas/NamedArea): I cannot quite reproduce how hierarchical structuring/sorting of this repeatable group of elements works. If I understand correctly, NamedAreaClass is to carry categories (like county, region, city), and its sibling NamedAreaName the place name. Unless there is some convention for the value of NamedAreaClass, however, from user interface perspective I cannot see how to determine the sequence in which to display these elements: this will only depend on the sequence they are delivered by the provider database, but no sorting is possible to determine that "county" always comes before "region" (especially if there is no controlled vocabulary).

WGB: Controlled vocabularies would not work here, because area categories vary widely and their hierarchy may even be reversed in a single country (I think I remember that this may be the case in Russia?). A sequence attribute was included to allow to include the correct sequence explicitly. (Normally, in the XML document itself the sequence is given by the sequence of named area tags. The query/wrapper would have to make sure that, if a sequence is given in the database or by a convention on the provider's side, this is correctly translated into the sequence in the XML document).

6.13. AHA: SiteFeature/Domain (Unit/Gathering/SiteFeatures/SiteFeature/Domain): Is there any standard list of terms for this (planned)?

CCO [Prompted by WGB]: The site feature domain serves to distinguish items such as earth science features (e.g. fossiliferous horizons) from biological features (e.g. a veteran tree or population of bats) etc. This is useful for sorting and searching and delivering appropriate term lists. I have a provisional list of domains in the thesaurus but it is likely to change with use (The thesaurus is arranged by Subject and domain as a means of distinguishing groups of term lists). It is possible that a field record or a specimen be directly linked to a site feature and this is the reason for its presence in the schema (In the larger data model features can have data of their own including management actions, threats, damage etc.). As far as immediate use of the schema goes, I think it unlikely that many existing data sets are arranged in such a way that this element would be used - however, the extended version of Recorder software (currently in beta test) does use domains and location features. The scope of the schema has been an issue from the start of the project, clearly, working towards a highly modular form that has a limited core to meet 90% of normal needs with extensions that meet more specialised requirements is the best answer - In answer to your question - it probably won't be missed from the core but shouldn't be forgotten completely.

WGB: I removed the entire domain for the time being because it is currently not used and could be confused with other elements.

6.14. AKR: The marine XML schema documentation provides another example for co-ordinate data that may be re-used for ABCD. [See Appendix 1].

WGB: The elements and attributes for coordinates given there are all covered by ABCD, with exeption of the explicit specification of a ground datum for depth and altitude. The ground datum for altitude is implicit in ABCD (above sea level), but for depth and height a corresponding attribute was added. I don't understand the choice between altitude and depth; I think that altitude can serve to further specify the ground datum, e.g. if a depth of 3m below water level is given for something collected in the Dead Sea or a mountain lake.

6.15. TC: The format of geographic coordinates in ABCD schema was examined. The latitude and longitude attributes are defined as String type in the Schema and it's

probably an error that should be checked and numeric type applied instead of String.
WGB: Done.

6.16. AHA: Under SiteCoordinateSets (ABCD v.1.48), I am missing an element that can hold unstructured coordinates. Especially with Deg./Min./Sec. data, these often turn up as one or two string entries in a database (12°15'N, 43°03'E), and thus cannot easily be converted to decimal degrees. The closest fit right now would be the AreaDetail, but an element within SiteCoordinates would seem appropriate.
WGB: I added an element CoordinatesText; hoping that this will not discurage people to convert their well-structured Degree/Min./Sec. data to decimal values.

6.17. PME: In UnitStateDomain under ObservationUnit is an ObservationMethod. I have several databases where the “bait” and traps are present to explain more the collecting or capture method, thus It might be logically to find in SpecimenUnit maybe a CollectingMethod as counterpart the the ObservationMethod? Another place where I looked to find a place for my "traps" and "lure" (for Fruitflies and Butterflies) information was in Gathering (for a GatheringMethod maybe), but could not clearly find any. A suggestion might be to have a general GatheringMethod in Gathering an observation method could be seen as a gathering of information not necessarily of a specimen.
EVB: Most marine scientists would want to keep a certain degree of information on how specimens were collected - trawls, grab, core... Maybe an extra element in GatheringType?
WGB: Added Gathering/GatheringMethod to cover this point and deleted ObservationMethod [v. 1.49]. We need to consider a more detailed Method-Type, but I would like to consolidate the existing first.

6.18. EVB: The equivalent of terrestrial 'slope' and 'aspect' are the sediment characteristics - hard or soft; if soft: granulometry (median grain size, percentage silt, clay...). Also temperature and salinity are virtually always measured.
WGB: Sediment characteristics as well as salinity etc. are covered by the BiotopeData/BiotopeMeasurements construct. However, you are right in that the Slope belongs here, too. Consequently, it was deleted in the schema [v.1.49d].

6.19. EVB: Depth is not unambiguously defined. Is it depth in the water column, or depth in the bottom, or depth of the bottom? If in the water column, is it against reference level (which one?) or measured level (latter including the effect of tides).
WGB: This has already been address by providing an attribute Datum with both Depth and Height elements [v. 1.49b].

6.20. PDA: in the OECD/ABCD field mapping proposed by W. Berendsohn, I feel the 'isolated from'(13) field should be mapped in the same way as the 'geographic origin'(14) field, under the node 'Gathering' of the ABCD schema.
WGB: Although at first sight this would simplify the mapping, I disagree. Treating the organism (or material) the strain was isolated from as an identification allows us to search for all information we have about the organism in only one place, which provides an important link between different collection types (and their contained information). In the case of unspecific materials (soil samples etc.), which are not normally included in searches, further details can always cited under Gathering/AreaDetail.

6.21. AIS: I would also question the use of Altitude, Depth and Height elements. My quick review seemed to indicate that all the sub elements were the same across the three (Altitude, Depth and Height). If that is the case, why not lump them together into a single element, with an attribute that has allowable content of "Altitude" or "Depth" or "Height". UnitMeasurement and Slope looks very similar to Altitude, Depth and Height.
WGB: All these could indeed be covered by a single MeasurementOrFact structure, with the parameter set to the respective content. Where such content is less generally used (e.g. in the case of water temperature, defined in the OBIS DwC extension, or for slope), this is the procedure already in place (element SiteMeasurementsAndFacts). However, to facilitate general readability of the schema we are keeping the Altitude etc. elements separate for the time being.

6.22. AIS: Just beware of mixing the format of the data within the XML tags. Consider the use of the country code elements ISO2Letter and ISO3Letter. These elements are really storing a code that represents the country; one just happens to be 2 characters and the other 3 characters. In this case, the element name is actually describing the format of the data. If you need the format for legacy purposes, why not place the number of characters in an attribute. This solution also scales, because if ISO comes out with a 4 character code, your schema must be changed to create a new tag. The attribute solution just requires the value "4".
WGB: This has been solved by using a single element ISO3166Code, which indeed can be 2-, 3-, or 4-letters long.

6.23. EVB: One more minor issue: if we understand the same under 'Datum' - reference to a ellipsoid - your depth field still does not give enough information. For benthic samples, we need depth within the bottom. Then there is also the depth of the bottom itself, or the depth of an observation - which can be measured either as water under the keel, measured depth in the water column, or as depth against a standard level. The standard level requires a reference level (usually defined as some level against the tidal cycle like average water height over tidal cycle, mean low low water... Not an academic matter: Belgium and Netherlands, eg, use different systems); at this point the reference elipsoid or datum (eg ED50, WGS84...) would not be important - only comes in when you specify coordinates, and can cause a deviation of a couple 100 m at worst. Are 'terrestrial' altitudes measured against an ellipsoid?
WGB: Actually, my definition of “Datum” is rather loose, for me it is the distance to a defined surface. ABCD provides the possibility to fix a vertical coordinate or range of coordinates in three ways: Altitude, which is defined as the distance of a ground (or floor, if you wish) surface from mean sea level (positive above). Height and depth are in fact similar, but traditionally both expressed as positive values for above and below a surface. Here the surface datum can be specified, as a geodetic datum, a level (e.g. a standard sea level), as the ground or floor level (further defined by altitude) or a combination of those. For further specifications the SiteMeasurementsOrFacts element can be used. As with horizontal coordinates, I really hope that these issues can be sorted out using GML soon, once data providers get the tools to translate their multitude of data items into the common standard.

6.24. EVB: A final complication for depth: sometimes the units are pressure rather than length; to convert you need extra knowledge about density of the sea water - so about

salinity and temperature. I'm a poor marine biologist, I don't know how accurate these conversions are, and whether it's worth bothering to keep track of how a depth measurement was arrived at. Anthony will be able to tell us how far off we could be by making simplifying assumptions. Let's hope these things are only important if you drive a submarine, not if you study whales.
WGB: For the time being I would leave that conversion to the data providers.

6.25. JTO (Salzburg ABCD meeting): We had the suggestion .. of including a new concept in ABCD for pictures of the GatheringSite where a unit was gathered. We would like to see a title for the picture and a description too.
WGB: Gathering/SiteImages/SiteImage was added (multimedia type) to v. 2.0.

6.26. JTO (Salzburg ABCD meeting): SpatialDatum should be mandatory when coordinates are provided.
WGB: correct, with the accent on “should be”. For the time being we will have to live with the fact (and the ensuing unreliablities) that the spatial datum is not known for many datasets / label data etc. out there.

6.27. AUC (translated): [did not find the following data item of the ITF-2 format in ABCD] Accuracy of Maximum Altitude
WGB: This was +/- solved in versions >1.2 by using the Measurement type, which includes an accuracy statement. We are not convinced that separate accuracies for the upper and lower end of a range are justified.

7. DateTime

7.01. MDO: this is the new DateTimeType for ABCD. The regular expression checks days etc. but allows 31 days for all months, because it would otherwise become unwieldy.
WGB: Incorporated.

7.02: [Oeiras meeting]: We need to add an Explicit field for ranges of date or time to indicate that the event actually took place for the time of the range given, instead of at some point in time during that period.
WGB: Added element PeriodExplicit to DateTimeType.

7.03. AKI: the following rule should read '....between 01 and 24'

<Rule xml:lang="en">The hour is expressed as a 2-digit value, left zero padded if necessary, ranging between 01 and 31.</Rule>

WGB: Corrected.

7.04. MDO: the ABCD DateTime type should include examples and a detailed explanation of the ISO format:

Year: YYYY (eg 1997)
Year and month: YYYY-MM (eg 1997-07)
Complete date: YYYY-MM-DD (eg 1997-07-16)
Month and day only: --MM-DD (eg --07-16)
Day only: ---DD (eg ---16)
Complete date plus hours and minutes:
YYYY-MM-DDThh:mmTZD (eg 1997-07-16T19:20+01:00)

Complete date plus hours, minutes and seconds:
YYYY-MM-DDThh:mm:ssTZD (eg 1997-07-16T19:20:30+01:00)
Complete date plus hours, minutes, seconds and a decimal fraction of a second
YYYY-MM-DDThh:mm:ss.sTZD (eg 1997-07-16T19:20:30.45+01:00)
where:
YYYY = four-digit year
MM = two-digit month (01=January, etc.)
DD = two-digit day of month (01 through 31)
hh = two digits of hour (00 through 23) (am/pm NOT allowed)
mm = two digits of minute (00 through 59)
ss = two digits of second (00 through 59)
s = one or more digits representing a decimal fraction of a second
TZD = time zone designator (Z or +hh:mm or -hh:mm)
WGB: examples and comments were included in the AppInfo.

7.05 GHA: posits that the DateTimeISOType is problematic, especially for processing purposes. A simple DC compatible type for dates would be more useful. See
http://efgblade.cs.umb.edu/twiki/bin/view/UBIF/CompositeDate
for details.
WGB: we can further discuss this, but in practical use we have up to now had no difficulties with the type as it is.

8. Measurements and Facts

8.01. AHA: under MeasurementAtomized/ParameterMeasured, an information is requested that, imo, is already unambiguously defined by the position of the type within the document (e.g., altitude). Therefore, it seems a bit redundant, unless the type shall also be regarded out of context.
WGB: I agree that it normally can be ignored, but this type could be generally useful.

8.02. HSA: Measurement should support discrete values. How does one represent, for instance, that an observation concerns Gender=Male, Generation=3, or Stage=Larva?
WGB: It does: To take your example:
ParameterMeasured = Generation; MeasurementLowerValue = 3.
ParameterMeasured = Gender; MeasurementLowerValue = Male.
ParameterMeasured = Stage; MeasurementLowerValue = Larva.
I have tried to make that clearer in the annotation given for the respective elements.
[N.B.: Recognizing their importance, both gender and stage are covered by separate elements, namely Sex under Unit and Stage under the ZoologicalUnit subtype of the UnitCollectionDomain section in the schema.]

8.03. NTH: The values in <MeasurementType> should have the ability to accept text. It is unlikely that this element will be used for sorting purposes, but some values may be descriptive.
Before version 1.4x, UnitMeasurement used to be for numeric results, UnitFact for textual results (character states or free text). However, following several suggestions (e.g. CCO) the two things have been united now in the MeasurementOrFact type.

8.04. HSA: For brevity, measurement should support in addition to lower and upper value, just "value".

WGB: I disagree. Searches are made easier with a single element for the lower OR the only value.

8.05. MDO: A controlled vocabulary should be provided for Measurement units
WGB: A controlled vocabulary can be recommended but for facts this element has to remain empty (e.g. counts). Presently I prefer to stick to avoiding controlled vocabularies, where the actual values are either difficult to predict or, like this one, reprent a rather large set (metric and non-metric units of length, area, volume, density, percentages, ppm, etc.).

8.06 GHA: MeasurementType should be better documented. It has significant overlap with SDD but I don't understand a number of aspects. Especially Is Scale intended to include measurement unit? In statistics "measurement scale" is nominal, ordinal, interval and ratio scale. In other areas like images scale is dimensionless (1000 x).
WGB: renamed MeasurementScale to MeasurementUnit. Element not mandatory, so that the type supports dimensionless measurements.

8.07 GHA: UnitDataType/UnitMeasurements: documentation is misleading: "Character/character state combinations such as counts and other measurements with numerical results." Character states are by definition categorical, so they cannot include numerical results.
WGB: corrected

8.08 GHA: It is unclear why UnitFact has attribution (who, reference, date) and UnitMeasurement not.
WGB: Added MeasuredBy, MeasurementDateTime and MeasurementReference to MeasurementType (UnitFact doesn’t exist any more).

8.09 PME: His [Peter Dawyndt’s] suggestion when we first analysed [a previous versions of ABCD] was to put the sequence information in Measurements. He is not sure that a specific SequenceType is interesting, as they will and already have all kind of other information related to their strains (units), as for example Fatty acid profiles all other kind of lab expiriences on it, they also want to share. His suggestion was to rather adapt UNITmeasurement (and eventually UnitFacts)to make it more genreric by adding some concepts as an ID, an URL, an reference, an Agent with a ContactType ... (basically what is now in the Sequencetype) In that way the information of nucleotic and protein sequences would fit nicely in measurment or facts and they could easely add any other type of information on their work there. I had just a small comment I noticed. If you keep the SequenceType, I have seen that in all other parts the concept "Reference" is used for Publications but in Sequence we have used "Publication" to be consistent with the rest, may be also use "Reference" ?
WGB: I have renamed Publication to Reference as you suggested. For the time being I would like to keep the sequence type because of its special importance for linkages to important public databases. However, the facts were united with measurements and are open for further discussion.

8.10 AHA: I am missing the Element-group "UnitFact" under /DataSets/DataSet/Units/Unit/UnitMeasurementsOrFacts in ABCD 1.49d. The element name "UnitMeasurementsOrFacts" suggests that they should be present as sibbling to UnitMeasurement. They have also been the only structured way to represent descriptive

data in v.1.2, which I would not want to miss.
WGB: the UnitMeasurement type now covers both numerical and non-numerical descriptors. Character/State -type facts are covered by ParameterMeasured and LowerValue.

9. References

9.01. PDA: the literature references [in the example output] have been translated into an hybrid format based on the Pubmed/MedLine XML schemes. I think this needs some revision in the ABCD schema, as was proposed in your comments, but at first sight, the BiblioML schema does not seem the best solution to me.
WGB: this item has to be discussed and resolved together with the other TDWG/GBIF standards. For the time being I would like to keep this very simple, i.e. as it is. This could be the ideal candidate to introduce GUIs to ABCD.

10. Nomenclatural types

10.1. JTO (Salzburg ABCD meeting): Create a controlled vocabulary for type designations
WGB: we will follow suit if a vocabulary is defined that covers the entire range of categories of “original material”, not only the terms prescribed by the Codes, and their translations.