A proposal for incorporating NMR relaxation data in the NMR Exchange Format (NEF)

Eliza Płoskoń, Victoria A. Higman, Gary Thompson, Geerten W. Vuister

on behalf of CCPN, University of Leicester, UK

Introduction

A growing number of scientific publications are focusing on NMR to provide insights into dynamics via NMR relaxation measurements. Currently, the submission guidelines of many journals unfortunately do not explicitly mandate the deposition of measured data from these articles in public databases, such as the BMRB, although many research funders have subscribed to the open-data principle. Recent scientific advances in machine learning (ML) and artificial intelligence (AI) have also demonstrated the power of (curated, large) libraries of data. Together, this emphasizes the need for software developers to facilitate the export of relaxation data in deposition-friendly format. For NMR, this leads to an urgent need to expand the NMR Exchange Format (NEF) to encompass NMR data from relaxation experiments. This document outlines a proposal for incorporation of NMR relaxation data into the NEF standard and is accompanied by example files illustrating its workings.

A core principle of the NEF is that it should be a lightweight format, carrying the minimal required essential information for interoperability with other software packages. NEF uses a STAR syntax, which uses tags to organise both data and metadata. Hence, NEF-encoded relaxation data should provide a minimal required number of tags for describing the most common data and scenarios, while simultaneously providing additional options to allow for an extensive coverage the majority of conceivable data sets that could be deposited (a form of the Pareto principle).

This NEF format extension proposal has two overarching aims:

1. Defining a format for information currently required for deposition of data obtained from NMR relaxation experiments into the BMRB.
2. Extending the format to facilitate reproducibility of exchange and processing of relaxation data.

In the NEF format, different types of data, e.g. chemical shifts, peaks, restraints of various types, etc., all reside in designated separate saveframes (a collection of tables and metadata). This structure allows specific attributes for each data type to be included in the saveframe's metadata. Consequently, NMR relaxation data should also be stored in its own saveframe or set of saveframes.

The aim of this NEF proposed standard is to support transfer of measured data, data series and data which has undergone simple fitting such as fitting of exponential relaxation curves and $\{^{1}H\}$-^{15}N heteronuclear relaxation NOE ratio calculation. The current proposal does not seek to address more complex fitting issues such as model free analysis and model selection or chemical exchange analysis and model fitting for CPMG or CEST. These are left for a

further proposal at a later juncture, but this proposal is designed to provide the basis for such a proposal and be compatible with such efforts.

We propose a unified approach by defining all relaxation data types in a single new NEF saveframe the**: `nef_relaxation_list`**.

To satisfy the first aim, the proposed **`nef_relaxation_list`** saveframe must facilitate the easy conversion of the new NEF format into the NMR-STAR formats already available for deposition. Section 3 addresses this issue.

To fulfil the second aim, we envision the inclusion of additional parameters to enhance data exchange between data analysis programs and to ensure the reproducibility of data analysis. Just as the NEF currently establishes a connection between peaks and the resulting restraints, we propose establishing an analogous linkage between peaks and the resulting rates that have been calculated. This connection will be established via a ***non-mandatory*** **`_nef_series_list`** saveframe (*vide infra*, section 5).

It is worth noting that some of the newly proposed tags currently lack their equivalents in the BMRB's NMR-STAR format dictionary. Consequently, the NEF-dynamics format will be a super set of the current BMRB definitions and changes will be required in the BMRB's NMRStar dictionaries if they wish to harvest this extra data into the BMRB database.

It is also important to understand that what underpins storage of the data in the BMRB (or wwPDB) databases is a dictionary (https://mmcif.wwpdb.org/dictionaries/mmcif_nmr-star.dic/Index/). This dictionary defines the tag, loops and metadata used for storing information, e.g. their meaning, the format of the expected values (whether they are float point numbers, integers, words, strings, etc.) and their mandatory or non-mandatory status. Hence, the dictionary document specifies the structure of files which will be provided during the deposition and allows faithful and unambiguous inclusion of the data into the databases. The NEF file format, is **not however** a data storage format, but rather an exchange format; as such, relative to the BMRB NMR-STAR definitions we propose to "condense" some of the aspects for the BMRB definitions (*vide infra* sections 3-5) in order to reduce the cognitive load and the amount of coding required for computer programmes processing or generating the NEF-dynamics files. The current NEF-dynamics proposal, however, still gives credence to both exchange and database-storage aspects and straightforward mappings can be made from NEF to BMRB definitions.

This document outlines below the proposed additions to the NEF format for storing relaxation measurements and the resulting implication for the NMR-STAR dictionary.

1. Concepts

Relaxation/dynamics/kinetic data is often derived from a series of spectra recorded as conditions change; for example as a function of time or concentration. We propose that the data analysis process can document and link the original data, the analysis types and calculation methods used, as well as the final results and errors, so as to assist reproducibility, re-computation and data harvesting. Analysis of such series data involves several key steps:

Preparation of Raw Data: This initial phase includes grouping relevant spectra, peak picking, identifying signals for atoms of interest (applying assignments), and linking signals between spectra to provide series of values.

Selecting Data Analysis: Choosing the appropriate data analysis for the experiment, whether it involves R_1 (T_1), R_2 (T_2), etc. and other relevant parameters.

Selecting Calculation Models: Deciding on the calculation models (equations) that best suit the specific experiment and data characteristics.

Outputting Results: The final step involves generating and outputting the results of the analysis. This provides a clear representation of the derived results. These results are the values currently required for deposition in the BMRB.

Each step plays an important role in ensuring accurate and meaningful insights from the NMR data are extracted and therefore requires NEF to adopt an extensive but flexible representation of the data.

2. Peak grouping

The specification of peaks used for the data analysis does not require any additions to the NEF format; it solely relies on outputting and depositing the peak lists from all the relevant spectra recorded as part of the analysis.

2.1. Grouping within spectra

In some cases, peaks originating from different atoms may be partially overlapped and in other cases signal splitting may be observed. When fitting peaks to determine the correct peak positions, intensities, volumes, etc., such multiple signals should be fitted as a single group to ensure that the analysis appropriately accounts for the influence of overlapping or split signals.

Proposal:
Introduction of ***non-mandatory*** attribute `cluster_id` to a `_nef_peak` loop within a `nef_nmr_spectrum` saveframe:

```
_nef_peak.cluster_id
```

Format: Int

Peaks with the same `cluster_id` will need to be analysed together to ensure data fidelity. This approach has previous provenance and has been used successfully in the NMRPipe program for data analysis, especially using NMRPipe's non-linear peak fitting routine NLINLS.

2.2 Linking peaks across spectra

In the calculation of rates and parameters, the initial step involves grouping of peaks originating from the same atoms across different spectra (or different planes in the case of a pseudo-nD) in the series, which were recorded with different parameters. These groups of peaks and their associated peak parameters (such as volume or height) are linked to a derived relaxation data point, e.g. an R1 rate. We propose making this connection explicit by specifying a link between each peak (or more generally, an experimental data point) and the calculated result.

To achieve linking of data points (peaks) across spectra, we propose a new ***non-mandatory*** saveframe `_nef_series_list`. These kind of data links are already defined in the NMR-STAR BMRB dictionary (specifically the `_Experimental_data_link_list` saveframe). The description of this saveframe is provided in section 5.

3. Experiment types

The most common relaxation/dynamics data types in the BMRB database are T_1, T_2, and heteronuclear NOEs. The BMRB dictionary effectively describes multiple data categories, providing future-proofing with 17 different categories in the kinetic data section and additional ones in the thermodynamic section. It is noteworthy that thus far many of these categories have not been used in any BMRB entry.

In NMR-STAR, each BMRB relaxation/dynamics category has its own designated saveframe with its own set of tags. For example, the heteronuclear NOE saveframe includes a tag for the NOE reference value (`NOE_ref_val`), a parameter not required for other relaxation data types like T_1 or T_2.

As all types of relaxation measurements share many common features, we propose a single or 'one size fits all' approach and hence a single new NEF saveframe:

`save_nef_relaxation_list_`

A ***mandatory*** saveframe level tag **`_nef_relaxation_list.experiment_type`**, using a predefined list, will specify the type of experimental data thus providing a streamlined and consistent structure which alleviates the need for adopters of the NEF format to accommodate multiple new saveframes. The `_nef_relaxation_list.experiment_type` values can be mapped in a 1:1 fashion onto currently defined BMRB kinetic and thermodynamic categories (T1, T2, etc). As an 'escape hatch' the class of experiment '*other'* will be provided for this and other fields when new or novel values are required and a `_nef_relaxation.comment` field would be used to contain further information, which could later be rolled into the standard. Note that in contrast with BMRB definitions, we propose to define rates (and its errors) in this saveframe, rather than decay-time constants, as the former are more universal across the different use cases and are more generally supported by the community.

It is important to emphasise that using one saveframe structure for all data-types would make the value of certain tags meaningless for some of the experiment types. However, their presence would still be required in order to satisfy their mandatory status in the equivalent BMRB dictionary (such as listing the `_nef_relaxation_list.ref_value` for R_1 or R_2). It would be required that in these cases the value of the unused tag would be given the unknown value '.'. In our opinion, this negative is outweighed by the positives of the choice for a single saveframe category (as outlined above).

We also propose a saveframe level tag `_nef_relaxation_list.source` which defines whether the data is theoretical as opposed to experimental, rather than specifying theoretical and experimental data in saveframes with separate categories. This choice also allows for the addition of simulated data (e.g. derived from modelling trajectories) as distinct from theoretical data (based solely on calculations) but does not necessarily require the addition of extra code by software packages to read and accommodate this data.

4. The nef_relaxation_list saveframe for relaxation data

The various sections of the proposed NEF `_nef_relaxation_list` saveframe are outlined below using example data.

4.1 Saveframe metadata (universal for all relaxation data types)

In the proposed `_nef_relaxation_list` saveframe, we envision incorporating the following tags to define attributes for experiment types and which are mandated by the BMRB for the analogous kinetic data. Additionally, we suggest introducing several new tags which may improve the interoperability of the reported data.

Example metadata (cf. Table 1):

```
save_nef_relaxation_list_R1

   _nef_relaxation_list.sf_category              nef_relaxation_list
   _nef_relaxation_list.sf_framecode             nef_relaxation_list_R1
   _nef_relaxation_list.experiment_type          heteronuclear_R1_relaxation
   _nef_relaxation_list.spectrometer_frequency_1H  600.0
   _nef_relaxation_list.value_type               Sz
   _nef_relaxation_list.value_unit               s-1
   _nef_relaxation_list.relaxation_atom_id       1
   _nef_relaxation_list.ref_value                .
   _nef_relaxation_list.source                   experimental
```

4.2 The `_nef_relaxation` loop

A loop, named `_nef_relaxation` contains the actual relaxation data of the various entries in the `_nef_relaxation_list` saveframe. It includes columns for an index, a data ID, a data combination ID, one or more chain codes, sequence codes, residue names, atom names, a relaxation value, and the error of the value, defined in detail in Table 2. Note that the tags defining the atoms involved the relaxation process can be added for each relevant atom (for example a CH_3 group), so the list can be expanded to include as many atoms as are required, in line with the NEF atom naming convention. As the dictionary will require a maximum number of atoms this value should be decided by the community and be chosen to be significantly greater than the maximum number that could reasonably occur in real experiments. Each row in the loop corresponds to a specific relaxation entry.

Example `_nef_relaxation` loop (cf. Table 2)**:**

```
loop_
   _nef_relaxation.index
   _nef_relaxation.data_id
   _nef_relaxation.data_combination_id
   _nef_relaxation.chain_code_1
   _nef_relaxation.sequence_code_1
   _nef_relaxation.residue_name_1
   _nef_relaxation.atom_name_1
   _nef_relaxation.chain_code_2
   _nef_relaxation.sequence_code_2
   _nef_relaxation.residue_name_2
   _nef_relaxation.atom_name_2
   _nef_relaxation.value
   _nef_relaxation.value_error
```

```
     1    1     .    A    18    LYS    N    A    18    LYS    H     2.136     0.054
     2    3     .    A    19    LEU    N    A    18    LYS    H     2.178     0.060
     3    4     .    A    20    ILE    N    A    18    LYS    H     2.170     0.052
stop_
```

Note in this case `_nef_relaxation_list.relaxation_atom_id` in the saveframe metadata is set to 1, so the atom being relaxed is N and the main cause of relaxation is the H atom.

Table 1. `_nef_relaxation_list` saveframe metadata

Tag	Comment	Mandatory	Type
`.sf_category`	Designates the saveframe category (STAR requirement) for kinetic data. "*nef_relaxation_list*"	Yes	
`.sf_framecode`	Specifies the frame code for the saveframe.	Yes	
`.experiment_type`	Enumerated list of possibilities that defines the relaxation experiment data content (see section 3)	Yes	Enumerated list *: auto_relaxation | dipole_CSA_cross_correlations | dipole_dipole_cross_correlations | dipole_dipole_relaxation | heteronuclear_NOEs | heteronuclear_R1_relaxation | heteronuclear_R1rho_relaxation | heteronuclear_R2_relaxation | H_exchange_protection_factors | H_exchange_rates | homonuclear_NOEs CPMG | CEST | other ***
`.spectrometer_frequency_1H`	Absolute spectrometer 1H frequency in MHz (i.e. with sufficient precision)	Yes	Float
`.value_type`	An enumerated list of value_types drawn from BMRB category value types. Required for relaxation and NOE experiments. For heteronuclear NOE it is a keyword that defines the type of value reported (from a non-strict list). For relaxation experiments it describes the nuclei involved in the coherence.	Yes	Enumerated list
`.value_units`	Associated units of the specified value_type. These should be consistent with BMRB types.	Yes	Enumerated list

.relaxation_atom_id	Pointer to the atom and nucleus involved in coherence. We advocate that users place the relaxing atom as atom 1. Any other atoms involved become atoms 2, 3,… For a ^{1}H-^{15}N heteronuclear NOE experiment this would be the ^{15}N atom.	Yes	Int
.ref_value	Value used to calibrate other NOE measurements	Yes	Float, null (‘.’) value allowed.
.source	Enumerated list of possibilities in the nef_dictionary: _nef_relaxation_list.source	Yes	Enumerated list: experimental | simulated | theoretical
.fitting_function **	Function used to fit model	No	Enumerated list: one-phase-decay | exponential-decay | inversion-recovery | other
.minimizer **	The minimizer used in the fitting	No	Enumerated list: leastsq | emcee | differential evolution | brute | basin hopping | ampgo | nelder | lbfgsb | powell | cg | newton | cobyla | bfgs | tnc | trust-ncg| other
.error_method **	The error method used in the fitting	No	Word: Recommendations in comments
.comment ***	Other information about the data set. As well as this any enumerated value defined as other should be resolved in a comment here e.g. fitting function: hammock based estimation	No	Word

* Derived from BMRB’s ‘Saveframe categories in group Kinetic data’ apart from the ones indicated as ‘Theoretical’
** **Proposal**: These tags could provide additional information for interoperability with other software packages; however, their usage would depend on consistent and accessible definitions. Ideally each of these tags would be described with an appropriate equation in the NEF specifications (dictionary).
*** The ‘other’ value is provided to accommodate any currently undefined type of experiment. It is recommended that additional context is provided as a comment.

Table 2. `_nef_relaxation_list` saveframe `_nef_relaxation.` Loop

Tag	Comment	Mandatory	Type
.index	Index of the relaxation data	Yes	Int
.data_id	ID of the relaxation data	Yes	Int
.data_combination_id	ID of the data combination. This can be used to indicate a sequence block or a set of related values (e.g. a number of entries subject to the same dynamic regime in a CPMG experiment; cf. 6.1).	No	Int
.chain_code_1	Chain code for the first atom	Yes	Word
.sequence_code_1	Sequence code for the first atom	Yes	Word
.residue_name_1	Residue name for the first atom	Yes	Word
.atom_name_1	Atom name for the first atom	Yes	Word
.chain_code_2	Chain code for the second atom. Only applicable for certain data_types.	No	Word
.sequence_code_2	Sequence code for the second atom. Only applicable for certain data_types.	No	Word
.residue_name_2	Residue name for the second atom. Only applicable for certain data_types.	No	Word
.atom_name_2	Atom name for the second atom. Only applicable for certain data_types.	No	Word
.value	Value of the calculated data. For chemical_rate it is a value for a simple two state transition A <--> B	Yes	Float
.value_error	Value error	No	Float

5. The _nef_series_data saveframe specifying a series of experimental data points.

Each entry in the _nef_relaxation_list saveframe is derived from a series of spectra (or planes in a pseudo nD spectrum). These spectra or planes differ by a specific value associated with their measurement. For example, in R_1 experiments this series variable is the delay time, for heteronuclear NOE it is the saturated/unsaturated state, for titration experiments it is (relative) ligand concentration etc. We propose a series variable that defines the value of the measurement for each spectrum or plane.
We believe it is important to preserve the link between recorded data (the peaks in a particular spectrum or pseudo spectrum plane) and their interpretation, which are the calculated values reported for each entry in the _nef_relaxation_list saveframe. The latter are derived from a group of peaks and their associated values (intensity, position, etc). We propose defining these links as a separate ***non-mandatory*** _nef_series_list saveframe. A single _nef_series_list saveframe would relate to a single _nef_relaxation_list saveframe. The _nef_series_list saveframe will contain two loops specifying the relationship between the experimentally recorded data and the entries in the _nef_relaxation_list saveframe. The two loops are:

The ***mandatory*** _nef_series_data loop (cf section 5.3 for an example) that serves as a link between various data within the experiment. It connects peaks, or more generally data points, to the series variable, modulated across the spectra/spectrum planes, the resulting data type values, and provides a direct link to the resulting _nef_relaxation_list saveframe entries. While some of this information may already be present in other saveframes within the file, including it in a single loop simplifies its usage by eliminating the need to code for linking various data across multiple saveframes.

The ***non-mandatory*** _nef_series_experiment loop (cf. section 5.2 for an example) is designed to capture information related to the NMR spectra present in the series experiments. Each entry in this loop represents a spectrum (or a plane of a pseudo nD spectrum) within a series experiment and includes details such as the spectrum ID, combination ID, series variable and series variable type.

The variable values in the _nef_series_data loop are generally derived from peaks, either as direct observables (e.g. intensities or volumes), or as functions of these observables (as is the case for CEST and CPMG). Most often, they will correspond to the _nef_series_experiment loop values, but there might be occurrences where this is not the case, e.g. when they represent an effective field, calculated from the series value associated with a spectrum (or plane).

The various elements of the _nef_series_data saveframe are detailed below.

5.1 Saveframe metadata

The saveframe metadata (cf. Table 3) contains details regarding the experiment type (as also defined for the _nef_relaxation_list saveframe), the type of series variables (e.g. time, field-strength, offset, gradient-strength) and the types of the (derived) values of the data points, including whether they are absolute, relative or represent intensity, etc. Note that for "simple" experiments, e.g. R1, R1, the series_variable_type and data_variable_type will essentially be identical, e.g. time in sec. For more complex data, such as CPMG, CEST, these could differ where the series_variable_type describes the parameter varied in the experimental setup, and the data_variable_type can represent a derived value (cf. Section 6).

Example metadata

```
save_nef_series_list_R1
    _nef_series_list.sf_category                nef_series_list
    _nef_series_list.sf_framecode               nef_series_list_R1
     nef_series_list.experiment_type            heteronuclear_R1_relaxation
    _nef_series_list.series_variable_type       time
    _nef_series_list.series_variable_unit       s
    _nef_series_list.data_variable_type         time
    _nef_series_list.data_variable_unit         s
    _nef_series_list.data_value_type            intensity
    _nef_series_list.data_value_unit            .
```

5.2 The _nef_series_experiment loop

This loop encodes a series of individual nD NMR spectra or a single pseudo-nD spectrum acquired, e.g. during a R1 relaxation experiment as in the example below, or a combination. In each entry of the loop, the `_nef_series_experiment` variables (cf. Table 4) provide a link to the spectra via the spectrum ID, and the associated series variables that define the unique value associated with each spectrum (or plane) in the series. In the example below for the R1 experiment, this value is the time duration of the relaxation process. Note that for a single pseudo-nD spectrum both the pseudo-dimension and the pseudo-dimension plane are to be defined for each `_nef_series_experiment` entry. A spectrum can also be designated as the reference experiment for the series (e.g. for CPMG, see section 6).

Example `_nef_series_experiment` loop (cf. Table 4):
The following example loop is taken from a R1 experiment.

```
loop_
    _nef_series_experiment.nmr_spectrum_id
    _nef_series_experiment.reference_experiment
    _nef_series_experiment.combination_id
    _nef_series_experiment.pseudo_dimension
    _nef_series_experiment.pseudo_dimension_point
    _nef_series_experiment.series_variable
    _nef_series_experiment.series_variable_error

    spectrum_R1_1_8ms       false   .   .   .   0.008   .
    spectrum_R1_3_120ms     false   .   .   .   0.120   .
    spectrum_R1_5_224ms     false   .   .   .   0.224   .
stop_
```

5.3 The _nef_series_data loop

The `_nef_series_data` loop is intended to capture all data points derived from the series NMR experiments. Each entry in this loop corresponds to a specific data point, often obtained from a single peak within the relevant spectrum (or plane), providing details such as the `nmr_spectrum_id, peak_id`, variable value (i.e. “x”), variable error, value (i.e. “y”), value error, relaxation_list_id, and data_id. Note that for CPMG and CEST, these data points may not equate to peaks directly (cf. section 6). Together this information allows for a comprehensive analysis and interpretation of experimental results.
The link to the entry in the `_nef_relaxation_list` saveframe is achieved through the `relaxation_list_id` and `data_id` values. Multiple entries in this loop will link to the identical

relaxation_list_id and data_id values, effectively encoding which data (e.g. peaks) resulted in the derived relaxation rate tabulated in the _nef_relaxation_list saveframe.

Example _nef_series_data loop:
The following example loop is taken from a R1 experiment.

```
loop_
    _nef_series_data.nmr_spectrum_id
    _nef_series_data.peak_id
    _nef_series_data.variable_value
    _nef_series_data.variable_error
    _nef_series_data.value
    _nef_series_data.value_error
    _nef_series_data.relaxation_list_id
    _nef_series_data.data_id

nef_nmr_spectrum_R1_8ms   1   0.008   .   16181764 .   nef_relaxation_list_R1  1
nef_nmr_spectrum_R1_120ms 1   0.120   .   12981652  .  nef_relaxation_list_R1  1
nef_nmr_spectrum_R1_224ms 1   0.224   .   10698450  .  nef_relaxation_list_R1  1
nef_nmr_spectrum_R1_8ms   2   0.008   .   16176964  .  nef_relaxation_list_R1  3
nef_nmr_spectrum_R1_120ms 2   0.120   .   13072074 .   nef_relaxation_list_R1  3
nef_nmr_spectrum_R1_224ms 2   0.224   .   10445134 .   nef_relaxation_list_R1  3
nef_nmr_spectrum_R1_8ms   3   0.008   .   15647703 .   nef_relaxation_list_R1  4
nef_nmr_spectrum_R1_120ms 3   0.120   .   12677407 .   nef_relaxation_list_R1  4
nef_nmr_spectrum_R1_224ms 3   0.224   .   10292315 .   nef_relaxation_list_R1  4
stop_
```

5.4 Using the _nef_series_data saveframe for input data

When we conceived the _nef_series_data saveframe, it was with the idea that it would link peaks with a relaxation parameter, i.e. the result (or output) from fitting peak heights/volumes. However, it also already contains all the data needed to do the fitting. Thus, if there is a workflow in which the user wishes to use program 1, e.g. extract peak heights, and fit these using program 2, there are in effect two ways to get the data from program 1 to program 2.

A) The _nef_series_experiment loop will provide all the series data needed and link to the _nef_nmr_spectrum saveframes whose _nef_peak loops will provide all the peak heights required.
B) A full _nef_series_data saveframe with both a _nef_series_experiment loop and a _nef_series_data loop, but without the _nef_series_data.relaxation_list_id and _nef_series_data.data_id columns.

The community and developers will need to decide whether it essentially wants to disallow workflow B, by making the _nef_series_data.relaxation_list_id mandatory or whether it wants to allow or even recommend workflow B because it will make the coding easier by providing all the data required for fitting within a single saveframe.

Table 3. `_nef_series_list` saveframe metadata

Tag	Comment	Mandatory	Type
`.sf_category`	Saveframe category	Yes	nef_series_list
`.sf_framecode`	Saveframe framecode	Yes	nef_framecode
`.experiment_type`	Type of experiment; identical as defined for `_nef_relaxation_list` equivalent.	Yes	Enumerated list (see Table 1)
`.series_variable_type`	Specifies the type of variable associated with each spectrum in the series loop (e.g. time, saturation, volume, height, shift)	Yes	Word
`.series_variable_unit`	Defines the unit of measurement for the series_variable	Yes	Word
`.data_variable_type`	Specifies the type of data_variable (i.e. the “x”) of the data loop	Yes	Word
`.data_variable_unit`	Defines the unit of measurement for the data_variable	Yes	Word
`.data_value_type`	Specifies the type of data_value (i.e. the “y”) of the data loop	Yes	Word
`.data_value_unit`	Defines the unit of measurement for the data_value	Yes	Word
`.comment`	Comments on the data including any ‘other’ values	No	Word

Table 4. `_nef_series_list` saveframe `_nef_series_experiment.` loop

Tag	Comment	Mandatory	Type
`.nmr_spectrum_id`	Identifies the NMR spectrum within the series with the NEF spectrum saveframe	Yes	Nef_framecode
`.reference_experiment`	Designate spectrum as the reference experiment	No	Bool
`.combination_id`	An optional grouping-id for spectra (e.g. in-phase/anti-phase)	No	Int
`.pseudo_dimension`	The optional pseudo-dimension (1-based)	No	Int
`.pseudo_dimension_point`	The optional pseudo-dimension point (1-based)	No	Int
`.series_variable`	The numerical value of the associated variable (e.g., time in seconds).	Yes	Float
`.series_variable_error`	Error in series variable	No	Float

Table 5. `_nef_series_list` saveframe `_nef_series_data.` loop

Tag	Comment	Mandatory	Type
`.nmr_spectrum_id`	ID of the associated NMR spectrum.	No	nef_framecode
`.peak_id`	ID of the associated Peak	No	nef_framecode
`.variable_value`	Variable value	Yes	Float
`.variable_error`	Error in variable value	No	Float
`.value`	Value	Yes	Float
`.value_error`	Error in value	No	Float
`.relaxation_list_id`	ID of the associated `_nef_relaxation_list`	Yes	nef_framecode
`.data_id`	Data ID	Yes	nef_framecode

6. CPMG and CEST type experiments providing chemical rates and populations of observed states

Since their design and introduction over 20 years ago, the use of CPMG and CEST experiments has increased substantially as they provide a unique opportunity to study slower time scale dynamics and invisible states. This popularity has not, however, been paired with increased deposition numbers to the BMRB database. The current proposal is foremost focussed on heteronuclear T_1 (R_1), T_2 (R_2) and heteronuclear NOE data, as these present the most easily achievable goals. It already has provisions needed for CPMG and CEST, although some additional elements will have to be introduced in order to accommodate all aspects of these experiments.

6.1 The `_nef_relaxation_list` saveframe for CPMG/CEST

The primary goal of both CEST and CPMG experiments is to determine the rates of chemical interconversion between different states of molecules and their respective populations. Note that data can either be fit on a per residue basis or across several residues. The current proposal already introduces the `_nef_relaxation_list` saveframe `_nef_relaxation.combination_id` to indicate which residues are to be evaluated together. Crucially however, we currently have not formulated a proposal for documenting how to incorporate in the `_nef_relaxation_list` saveframe the various models, rates of interconversion and associated chemical shift differences.
One possibility would be to introduce a sufficiently large, but pre-defined number of rates (rate_1, rate_2, ..), chemical shift differences (chemical_shift_difference_1, _chemical_shift_difference_2, …), and populations (population_1, _population_2, …) to accommodate a number of (relatively simple) pre-defined dynamical models. This, however, is not a general solution.
A general definition of a kinetic model for n-states would, in all likelihood, require one or more additional loops in the `_nef_relaxation_list` saveframe, which would introduce a significant new aspect with respect to current BMRB saveframes documenting relaxation data. Hence, we are leaving this issue open for discussion.

Even with a partially completed definition for the CPMG/CEST `_nef_relaxation_list` saveframe, it is possible already to use the `_nef_series_list_`saveframe (detailed below).

6.2 Documenting series for CPMG and CEST

The purpose of the `_nef_series_list` saveframe is to document which set of experiments (either a set of nD spectra, a pseudo-nD spectrum, or combination) and their associated experimental parameters were used, and to document the (derived) experimental data points.

To accommodate CPMG and CEST experiments, the `_nef_series_experiment` loop has the `_nef_series_experiment.reference_experiment` variable to designate a spectrum as the reference experiment.

For the CPMG experiment, we propose the additional metadata variable `_nef_series_list.constant_time`, which will document the constant time period of the CMPG cycles.

For the CEST experiment, we propose the additional metadata variables `_nef_series_list.carrier_frequency` and `_nef_series_list.field_strength` which will document the absolute frequency relative to which the CEST field is applied, and the strength of the CEST field.

Example CPMG series saveframe

This series saveframe documents a CPMG experiment with a 20 ms constant-time period, one reference experiment and a pseudo-3D spectrum with 3 planes encoding 10.0, 40.0, 85.0 CPMG cycles, corresponding to 100.0, 170.0 and 600.0 Hz values in the `_nef_series_data` loop and documenting relative rates for the data_values. Note that a reference to a `_nef_relaxation_list` saveframe can be absent or present. If present, the series_data are linked to a specific atom (i.e. assignment).

```
save_nef_series_list_CPMG_295K
    _nef_series_list.sf_category              nef_series_list
    _nef_series_list.sf_framecode             nef_series_list_CPMG_295K
    _nef_series_list.experiment_type          CPMG
    _nef_series_list.constant_time            0.020
    _nef_series_list.series_variable_type     cycles
    _nef_series_list.series_variable_unit     count
    _nef_series_list.data_variable_type       frequency
    _nef_series_list.data_variable_unit       s-1
    _nef_series_list.data_value_type          relative
    _nef_series_list.data_value_unit          s-1

loop_
    _nef_series_experiment.nmr_spectrum_id
    _nef_series_experiment.reference_experiment
    _nef_series_experiment.combination_id
    _nef_series_experiment.pseudo_dimension
    _nef_series_experiment.pseudo_dimension_point
    _nef_series_experiment.series_variable
    _nef_series_experiment.series_variable_error

    nef_nmr_spectrum_CPMG_ref_295K    true     .    .   .     0.0     .
    nef_nmr_spectrum_CPMG_295K        false    .    3   3     10.0    .
    nef_nmr_spectrum_CPMG_295K        false    .    3   6     40.0    .
    nef_nmr_spectrum_CPMG_295K        false    .    3   12    85.0    .
stop_


loop_
    _nef_series_data.nmr_spectrum_id
    _nef_series_data.peak_id
    _nef_series_data.variable_value
    _nef_series_data.variable_error
    _nef_series_data.value
    _nef_series_data.value_error
    _nef_series_data.relaxation_list_id
    _nef_series_data.data_id

    nef_nmr_spectrum_CPMG_295K   2   100.0  .  27.1  1.2   .              .
    nef_nmr_spectrum_CPMG_295K   8   170.0  .  30.0  1.0   .              .
    nef_nmr_spectrum_CPMG_295K   1   600.0  .  12.1  1.2   .              .
    nef_nmr_spectrum_CPMG_295K   4   100.0  .  15.1  1.2   CPMG_15N_295K  2
    nef_nmr_spectrum_CPMG_295K   22  170.0  .  11.3  1.1   CPMG_15N_295K  2
    nef_nmr_spectrum_CPMG_295K   23  600.0  .   9.1  1.0   CPMG_15N_295K  2
```

```
    nef_nmr_spectrum_CPMG_295K   31  100.0  .  12.1  0.9   CPMG_15N_295K  3
    nef_nmr_spectrum_CPMG_295K   32  170.0  .   6.5  1.1   CPMG_15N_295K  3
    nef_nmr_spectrum_CPMG_295K   33  600.0  .   4.9  1.0   CPMG_15N_295K  3
stop_
save_
```

Example CEST series saveframe:
This series saveframe documents a CEST experiment with a (^{13}C) carrier at 151.345267 MHz, 20 Hz rf field, one reference experiment and a pseudo-3D spectrum with 3 planes recorded with -1500, -1470, -1440 offset frequencies. The derived data points in the `_nef_series_data` loop document the relative intensities (to the reference experiment) at various ppm values. Note that a reference to a `_nef_relaxation_list` saveframe can be absent or present. If present, the series_data are linked to a specific atom (i.e. assignment).

```
save_nef_series_list_CEST_13C
    _nef_series_list.sf_category                nef_series_list
    _nef_series_list.sf_framecode               nef_series_list_CEST_13C
    _nef_series_list.experiment_type                  CEST
    _nef_series_list.carrier_frequency          151.345267
    _nef_series_list.field_strength             20
    _nef_series_list.series_variable_type       frequency
    _nef_series_list.series_variable_unit       s-1
    _nef_series_list.data_variable_type         chemical_shift
    _nef_series_list.data_variable_unit         ppm
    _nef_series_list.data_value_type            relative
    _nef_series_list.data_value_unit            .

loop_
    _nef_series_experiment.nmr_spectrum_id
    _nef_series_experiment.reference_experiment
    _nef_series_experiment.combination_id
    _nef_series_experiment.pseudo_dimension
    _nef_series_experiment.pseudo_dimension_point
    _nef_series_experiment.series_variable
    _nef_series_experiment.series_variable_error

    nef_nmr_spectrum_referenceCEST_13C  true   .    .   .   .          .
    nef_nmr_spectrum_pseudoCEST_13C     false  .    3   1   -1500.0    .
    nef_nmr_spectrum_pseudoCEST_13C     false  .    3   2   -1470.0    .
    nef_nmr_spectrum_pseudoCEST_13C     false  .    3   3   -1440.0    .
stop_

loop_
    _nef_series_data.nmr_spectrum_id
    _nef_series_data.peak_id
    _nef_series_data.variable_value
    _nef_series_data.variable_error
    _nef_series_data.value
    _nef_series_data.value_error
    _nef_series_data.relaxation_list_id
    _nef_series_data.data_id

    nef_nmr_spectrum_pseudoCEST_13C  .   51.0   .  0.493  .  .         .
    nef_nmr_spectrum_pseudoCEST_13C  .   52.0   .  0.404  .  .         .
    nef_nmr_spectrum_pseudoCEST_13C  .   58.0   .  0.512  .  .         .
    nef_nmr_spectrum_pseudoCEST_13C  .   58.0   .  0.527  .  CEST_13C  2
```

```
    nef_nmr_spectrum_pseudoCEST_13C  .   63.0   .  0.397  .  CEST_13C  2
    nef_nmr_spectrum_pseudoCEST_13C  .   68.0   .  0.522  .  CEST_13C  2
stop_
save_
```

7. Conclusions

In conclusion, we hope that these draft proposals present a comprehensive approach to addressing our aims of defining dynamics data within the NEF standard.

In moving forward, we would like to gather additional feedback from the NMR community and refine the proposals based on their input. It is hoped that this feedback will enhance the viability and effectiveness of the proposed initiative.

We are aware that there are moves within the community to propose a controlled vocabulary in relation to NMR dynamics experiments and data. We are in full support of this and we aim for those proposals to be applied to our NEF extension proposed here.